



UMA PROPOSTA DE UM MODELO PERIÓDICO MULTIVARIADO
AUTORREGRESSIVO MULTIPLICATIVO PARA GERAÇÃO DE CENÁRIOS
DE AFLUÊNCIA APLICÁVEL AO MODELO DE PLANEJAMENTO DO
SETOR ELÉTRICO BRASILEIRO

Filipe Goulart Cabral

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia Mecânica, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia Mecânica.

Orientador: José Herskovits Norman

Rio de Janeiro
Março de 2016

UMA PROPOSTA DE UM MODELO PERIÓDICO MULTIVARIADO
AUTORREGRESSIVO MULTIPLICATIVO PARA GERAÇÃO DE CENÁRIOS
DE AFLUÊNCIA APLICÁVEL AO MODELO DE PLANEJAMENTO DO
SETOR ELÉTRICO BRASILEIRO

Filipe Goulart Cabral

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO
ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE
ENGENHARIA (COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE
JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A
OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA
MECÂNICA.

Examinada por:

Prof. José Herskovits Norman, Dr.Ing.

Prof. Getúlio Borges da Silveira Filho, Ph.D.

Prof. Joari Paulo da Costa, Ph.D.

Prof. Fernando Luiz Cyrino Oliveira, Ph.D.

Prof. Thiago Gamboa Ritto, Ph.D.

RIO DE JANEIRO, RJ – BRASIL

MARÇO DE 2016

Cabral, Filipe Goulart

Uma proposta de um modelo periódico multivariado autorregressivo multiplicativo para geração de cenários de afluência aplicável ao modelo de planejamento do setor elétrico brasileiro/Filipe Goulart Cabral. – Rio de Janeiro: UFRJ/COPPE, 2016.

XII, 133 p.: il.; 29, 7cm.

Orientador: José Herskovits Norman

Dissertação (mestrado) – UFRJ/COPPE/Programa de Engenharia Mecânica, 2016.

Referências Bibliográficas: p. 68 – 74.

1. Séries temporais. 2. Modelo periódico. 3. Modelo autorregressivo. 4. Modelo multiplicativo. I. Norman, José Herskovits. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Mecânica. III. Título.

*A todos os brasileiros que ainda
sonham com um país justo e
próspero.*

Agradecimentos

- À minha mãe Rita por todo apoio e incentivo, os quais foram essenciais para a elaboração desta dissertação.
- À minha esposa Pamella por todo seu companheirismo e cumplicidade em todos os momentos, principalmente nos mais difíceis da minha trajetória acadêmica.
- À minha irmã Daniella e família pela torcida e estímulo que me motivaram em todos os momentos.
- Ao meu amigo Claudio por toda sua irmandade e orientação que norteou grande parte da minha formação.
- Ao meu orientador Jose Herkovits pela oportunidade de realização deste trabalho e pelo acolhimento.
- Ao Joari pela sua extraordinária contribuição para esta dissertação e por seus ensinamentos sobre moral e ética, os quais busco praticar em minha conduta.
- Ao professor Getúlio pela disponibilidade e entusiasmo em ensinar e discutir o assunto proposto e pela grande contribuição para esta dissertação.
- Ao Evandro pelas críticas ao trabalho e pelos momentos de descontração em períodos tensos da elaboração da dissertação.
- Aos professores Thiago Ritto e Fernando Cyrino por aceitarem o convite para compor a banca examinadora e pelas observações sobre a dissertação.
- Aos meus professores Fabio Ramos, Bruno Scardua e Marcelo Tavares, cujos ensinamentos me inspiram até hoje.
- Aos amigos Hugo, Guilherme, Ivani, Paloma e Gabriel Castor da Matemática Aplicada da UFRJ, Carolina e Elmer do Programa de Engenharia Mecânica da UFRJ, Diego, Ana Luísa, Alex e José Wilson do Programa de Engenharia de Sistemas da UFRJ, agradeço pelas excelentes recordações da minha vida acadêmica.

- Ao Operador Nacional do Sistema Elétrico (ONS), à equipe DPP/GMC, ao gerente Alberto Kligerman e ao gerente executivo Roberto Fontoura pela oportunidade e apoio a este trabalho.
- Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo suporte financeiro.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

UMA PROPOSTA DE UM MODELO PERIÓDICO MULTIVARIADO
AUTORREGRESSIVO MULTIPLICATIVO PARA GERAÇÃO DE CENÁRIOS
DE AFLUÊNCIA APLICÁVEL AO MODELO DE PLANEJAMENTO DO
SETOR ELÉTRICO BRASILEIRO

Filipe Goulart Cabral

Março/2016

Orientador: José Herskovits Norman

Programa: Engenharia Mecânica

O planejamento da operação energética de longo prazo brasileiro é modelado por um problema de programação estocástica cuja única incerteza representada são as afluições. Um algoritmo que resolve satisfatoriamente esse problema é o Stochastic Dual Dynamic Programming (SDDP) e que faz uso de árvores de cenários para obter a estratégia de operação. Esta dissertação propõe um modelo periódico multivariado linear autorregressivo multiplicativo para descrever as afluições que atende os pressupostos do algoritmo SDDP.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

A PROPOSAL OF A PERIODIC VECTOR AUTOREGRESSIVE
MULTIPLICATIVE MODEL FOR SCENARIO GENERATION OF INFLOWS
APPLICABLE TO THE BRAZILIAN POWER SYSTEM OPERATION
PLANNING

Filipe Goulart Cabral

March/2016

Advisor: José Herskovits Norman

Department: Mechanical Engineering

The Brazilian long-term operation planning is modeled by a stochastic programming problem where the unique uncertainty is represented by the inflows. An algorithm that solves this problem satisfactorily is the Stochastic Dual Dynamic Programming (SDDP) which makes use of scenarios trees to obtain the operation policy. This dissertation proposes a periodic vector autoregressive multiplicative model to describe the inflows that meets the requirements of the SDDP algorithm.

Sumário

Lista de Figuras	xi
Lista de Tabelas	xii
1 Introdução	1
1.1 Objetivo	2
1.2 Organização da dissertação	3
1.3 Revisão bibliográfica	4
1.3.1 Modelos multivariados em hidrologia	4
1.3.2 Modelos periódicos autorregressivos	6
1.3.3 Modelos multiplicativos	6
1.3.4 A origem da proposta do modelo periódico vetorial autorregressivo multiplicativo	6
2 O modelo PVAR	8
2.1 Processos estacionários e ergódicos	8
2.2 Formulação do modelo PVAR	10
2.2.1 Estacionariedade e causalidade do PVAR	10
2.2.2 Propriedades do modelo PVAR <i>causal</i>	18
2.3 Estimação dos parâmetros	19
2.3.1 Método de Yule-Walker	19
2.3.2 Método dos mínimos quadrados	22
2.3.3 Método dos mínimos quadrados com restrições nos coeficientes	27
2.4 Determinação dos resíduos	28
2.5 Estimação da distribuição dos ruídos	28
2.6 Identificação do modelo	31
2.6.1 Critério da autocorrelação parcial periódica	31
2.6.2 Critério de Informação Bayesiano (BIC)	34
2.7 Geração de cenários	36

3	O modelo PVAR_m	37
3.1	Formulação do modelo PVAR _m	37
3.2	Estimação dos parâmetros	40
3.3	Determinação dos resíduos	40
3.4	Estimação da distribuição dos ruídos	41
3.5	Identificação do modelo	41
3.6	Geração de cenários	41
3.6.1	Geração bootstrap	42
3.6.2	Geração PCA-bootstrap	42
3.6.3	Geração PCA-bootstrap suavizado	44
4	Caso exemplo	46
4.1	Cenários gerados	47
4.2	Estatísticas descritivas básicas	50
4.3	Análise das distribuições de ENA	58
4.4	Análise descritiva de “runs”	60
4.4.1	Sequências abaixo da média	60
4.5	Análise descritiva de somas parciais	63
5	Conclusões	66
5.1	Trabalhos futuros	67
	Referências Bibliográficas	68
A	Problema do planejamento de longo prazo da operação	75
A.1	Motivação	76
A.2	Formulação multi-estágio determinística	77
A.3	Formulações multi-estágio estocásticas	80
A.3.1	Formulação por observações	81
A.3.2	Formulação aninhada	83
A.3.3	Formulação por programação dinâmica	85
A.4	Formulação do problema de planejamento de longo prazo da operação	88
A.5	Árvore de cenários	90
A.6	Elementos de análise convexa e algoritmos em otimização estocástica	95
A.6.1	Algoritmo Nested Cutting Plane	100
A.6.2	Aplicação: problema do planejamento de longo prazo da operação	107
A.6.3	Algoritmo SDDP	113
A.6.4	Aplicação: problema do planejamento de longo prazo da operação considerando os modelos PVAR e PVAR _m	119

Lista de Figuras

4.1	PVAR – cenários gerados subsistema NE (3) (aditivo sem restrição)	48
4.2	PVAR _m – cenários gerados subsistema NE (3) (mult. sem restrição)	48
4.3	PVAR – cenários gerados subsistema NE (3) (aditivo com restrição)	49
4.4	PVAR _m – cenários gerados subsistema NE (3) (mult. com restrição)	49
4.5	PVAR – média do subsistema SE (1) (aditivo sem restrição)	51
4.6	PVAR _m – média do subsistema SE (1) (multiplicativo com restrição)	51
4.7	PVAR – desvio padrão do subsistema SE (1) (aditivo sem restrição)	52
4.8	PVAR _m – desvio padrão do subsistema SE (1) (mult. com restrição)	53
4.9	PVAR – assimetria do subsistema SE (1) (aditivo sem restrição)	54
4.10	PVAR _m – assimetria do subsistema SE (1) (mult. com restrição)	54
4.11	PVAR – curtose do subsistema SE (1) (aditivo sem restrição)	55
4.12	PVAR _m – curtose do subsistema SE (1) (mult. com restrição)	56
4.13	PVAR – correl. espacial entre SE (1) e NE (3) (aditivo sem restrição)	57
4.14	PVAR _m – correl. espacial entre SE (1) e NE (3) (mult. com restrição)	57
4.15	PVAR _m – qqplot SE (1) e mês de fevereiro (mult. com restrição)	59
4.16	PVAR _m – qqplot SE (1) e mês de agosto (mult. com restrição)	59
4.17	PVAR _m – Número de seq. negativa - SE (1) (mult. com restrição)	61
4.18	PVAR _m – Número total de seq. negativas (mult. com restrição)	61
4.19	PVAR _m – Duração média de seq. negativas (mult. com restrição)	62
4.20	PVAR _m – Intensidade média de seq. negativas (mult. com restrição)	62
4.21	Comprimento do período crítico - SE (1) (mult. com restrição)	64
4.22	Capacidade de regularização - SE (1) (mult. com restrição)	65
4.23	Energia afluyente média - SE (1) (mult. com restrição)	65
A.1	Árvores de cenários	90
A.2	Funções de custo total	95
A.3	Funções de custo futuro	95
A.4	função poliedral	96
A.5	Plano tangente e vetor subgradiente	98
A.6	Função módulo	99
A.7	Aproximação da função de custo futuro	101

Lista de Tabelas

A.1	Matriz de capacidade instalada de geração de energia elétrica do Brasil.	75
A.2	Exemplo de funções antecessoras	91
A.3	Função antecessora da sub-árvore dependente	92
A.4	Conversão de notação indicial para vetorial.	111

Capítulo 1

Introdução

O planejamento da operação energética do Sistema Interligado Nacional (SIN) é feito com o apoio de uma cadeia de modelos de otimização cujo propósito é definir uma meta de geração para cada usina. Para o longo prazo, considera-se um modelo de otimização estocástica multiestágio discretizado em bases mensais, onde as usinas hidroelétricas são representadas por reservatórios equivalentes de energia (REE), vide ARVANITIDIS e ROSING [2]. Neste contexto, as únicas incertezas consideradas são as energias naturais afluentes (ENA) em cada REE, isto é, a energia elétrica que pode ser gerada a partir da vazão natural em um aproveitamento hidroelétrico. Por definição, a vazão natural é a vazão que ocorreria em uma seção do rio se não houvesse, a montante, ações antrópicas na bacia, como a regularização de reservatórios, as transposições de vazão e as captações para diversos fins.

A extensão de cinco anos do horizonte de planejamento torna impraticável o uso da maioria dos algoritmos de otimização estocástica para a resolução numérica deste problema. Um algoritmo que resolve satisfatoriamente o problema do planejamento de longo prazo e é atualmente empregado no setor é o Stochastic Dual Dynamic Programming (SDDP), vide PEREIRA e PINTO [50]. Entretanto, o algoritmo SDDP possui certos requisitos que atestam o seu bom funcionamento, mas que impõem algumas restrições de modelagem, vide o apêndice A.

O primeiro requisito do SDDP é o processo aleatório possuir componentes independentes (SHAPIRO [63]). Porém, assumir que a afluência é um processo independente desconsidera uma característica importante do fenômeno que é a correlação temporal. Uma maneira de incorporar a informação da correlação temporal sem violar a independência necessária para o SDDP é ajustar às afluências um modelo de séries temporais com termos de *erros* independentes. A equação do modelo ajustado é incluída como uma restrição do problema de otimização estocástica. Assim, reescreve-se o problema de planejamento da operação considerando as afluências energéticas como variáveis e os termos de erro como os *únicos* parâmetros aleatórios.

O segundo requisito do SDDP é a *linearidade* do problema de otimização estocástica. Por isso, para incluir a equação do modelo ajustado que representa a evolução das ENA's como uma restrição é necessário que essa seja expressa por meio de uma fórmula linear. Ou seja, a afluência deve ser uma função *afim* das afluências passadas. Mesmo considerando a generalização do SDDP para problemas de otimização estocástica *convexa* a exigência de linearidade para as restrições de igualdade se mantém, pois uma restrição não linear de igualdade destrói a propriedade de convexidade de um problema de otimização matemática.

Um candidato natural para modelar a série temporal de ENAs é o modelo Periódico Autorregressivo (PAR) descrito em HIPEL e MCLEOD [28]. O PAR consiste em um modelo Autorregressivo (AR) para cada mês. Assim, são levadas em conta tanto as correlações temporais quanto a sazonalidade do processo. Atualmente, o PAR é o modelo utilizado para o planejamento de longo prazo do SIN. Entretanto, por ser um modelo com erro *aditivo*, a principal limitação do PAR é a possibilidade de gerar valores negativos de afluências. Uma tentativa para contornar esse problema (CEPEL [9]) introduziu como efeito colateral a dependência temporal dos ruídos.

Com o intuito de superar essa dificuldade do PAR, propõe-se nesta dissertação o modelo Periódico Vetorial Autorregressivo Multiplicativo ($PVAR_m$). Em sua concepção, o $PVAR_m$ representa a correlação temporal e a sazonalidade de maneira análoga ao PAR *vetorial* (PVAR), porém com erros *multiplicativos*. Esta modificação do conceito de erro aliada a restrições nos coeficientes que garantam previsões sempre positivas possibilita ao $PVAR_m$ superar as limitações apresentadas pelo PAR. Vale ressaltar que as propriedades e limitações apresentadas pelo PAR são as mesmas que a do PVAR. De forma análoga ao PAR, no modelo $PVAR_m$ a afluência também é uma função afim das afluências passadas e portanto sua utilização conjuntamente com o SDDP é legítima.

1.1 Objetivo

O objetivo desta dissertação é apresentar o modelo de séries temporais chamado modelo periódico vetorial autorregressivo com termos de erros multiplicativos ($PVAR_m$).

A vantagem do modelo $PVAR_m$ é gerar cenários positivos de afluência e ser um modelo linear, isto é, a afluência é uma função *afim* com coeficientes estocásticos das afluências passadas. Adicionalmente, o modelo $PVAR_m$ pode ser utilizado nos procedimentos que suportam a elaboração do planejamento da operação do SIN.

1.2 Organização da dissertação

Como a proposta do erro multiplicativo foi inspirada na metodologia com erro aditivo, optou-se por revisar os fundamentos do modelo PVAR antes de se apresentar o PVAR_m. Desta forma, é possível estabelecer comparações entre os modelos e buscar inspiração sobre propostas para o PVAR_m.

A revisão do PVAR se inicia na seção 2 com as hipóteses básicas sobre as observações de ENA. Neste trecho, apresenta-se o conceito de estacionariedade e ergodicidade, isto é, as hipóteses estatísticas fundamentais de séries temporais, e como essas definições se estendem para processos periódicos.

Na seção 2.2, define-se o modelo PVAR. Em seguida, discorre-se sobre as condições nos coeficientes do modelo que garantem sua estacionariedade e causalidade. Um PVAR causal é um PVAR estacionário onde os termos da série podem ser expressos por uma combinação linear infinita dos erros aditivos passados. Ainda nesta seção, avalia-se as propriedades de um PVAR causal, onde constata-se que sua previsão é não viesada e sua variância condicional (volatilidade) é constante sazonalmente.

A seção 2.3 se refere aos estimadores dos coeficientes do modelo. Neste trecho, expõe-se o método de Yule-Walker, o método de mínimos quadrados e o método de mínimos quadrados com restrição nos coeficientes. A seção 2.4 recapitula o conceito de resíduo. A seção 2.5 descreve duas opções de modelagem para a distribuição dos ruídos: a normal multivariada e a log-normal multivariada a três parâmetros. A seção 2.6 compreende a identificação do modelo PVAR com a exposição de dois critérios: o critério de autocorrelação parcial periódica e o Bayesian Information Criterion (BIC). Por último, a seção 2.7 aborda o método usual de geração de cenários.

A proposta do modelo PVAR_m é apresentada na seção 3.1. Conjectura-se que exista uma condição sobre os coeficientes do PVAR_m que permita representar os termos da série em função dos ruídos multiplicativos passados. Esta representação seria o análogo multiplicativo da condição de causalidade do PVAR. Se essa conjectura for verdadeira então as previsões do PVAR_m causal são não viesadas e sua volatilidade depende sazonalmente do produto cruzado da previsão. Isto constituiria uma evidência de que o modelo PVAR_m é estruturalmente diferente do PVAR.

A seção 3.2 apresenta o método de mínimos quadrados com restrição nos coeficientes para a estimação do modelo PVAR_m. Os coeficientes obtidos dessa forma aliados a ruídos multiplicativos positivos garantem cenários de ENA sempre positivos. No excerto seguinte, define-se o conceito de resíduo do PVAR_m. Em seguida, propõe-se a distribuição empírica como um estimador não paramétrico da distribuição dos ruídos multiplicativos e comenta-se o uso do critério BIC para a

identificação do modelo.

A seção 3.6 compreende o método de geração de cenários para o $PVAR_m$. Como o estimador de ruídos sugerido é a distribuição empírica, os cenários gerados pelo método usual são finitos, já que o número total n de ruídos “observados” (resíduos) é a quantidade de observações da série temporal. Para superar essa dificuldade, propõe-se dois métodos adicionais. Um método é baseado em componentes principais que descorrelaciona os ruídos multiplicativos, sorteia um valor de cada componente descorrelacionada e correlaciona-se as componentes do vetor gerado, ampliando para $n^{\#Sis}$ o número de possíveis ruídos, onde $\#Sis$ é a quantidade de REE’s. O outro método é semelhante, porém ajusta-se uma distribuição suave em cada componente descorrelacionada e sorteia-se da distribuição ajustada, tornando infinito o número de possíveis ruídos.

O apêndice A apresenta uma breve introdução à programação linear estocástica. O propósito deste material é ser um guia capaz de fornecer um panorama dos principais resultados de otimização que estão ligados com este trabalho. Neste sentido, este material contém a formulação básica do problema de planejamento de longo prazo bem como as respectivas reformulações que possuem o modelo $PVAR$ e $PVAR_m$ como restrições.

1.3 Revisão bibliográfica

A revisão bibliográfica a seguir foi inspirada no livro de Hipel e McLeod (1994) e na análise de alguns dos principais artigos de cada tópico.

1.3.1 Modelos multivariados em hidrologia

A pesquisa em modelos multivariados voltados à hidrologia remonta ao início dos anos 1960, quando pesquisadores como MAASS *et al.* [39] introduziram princípios estatísticos na área da hidrologia. Nas primeiras investigações, quase sempre a forma exata do modelo era imposta antes de qualquer ajuste aos dados. Por exemplo, alguns pesquisadores sugeriam usar um modelo $VAR(1)$ enquanto outros propunham um $VARMA(1,1)$. Este tipo de procedimento pode resultar no uso de um modelo que não se ajusta bem aos dados, como reportado por FINZI *et al.* [20].

Num artigo pioneiro, FIERING [19] propôs um modelo multivariado para explicar o comportamento de duas séries, X_t e Y_t . O modelo de Fiering foi posteriormente modificado por KAHAN [32] e LAWRENCE [38].

MATALAS e WALLIS [41] sugeriu um modelo $VAR(1)$ espacial (“multisite”, em inglês) para uso na hidrologia. Seu modelo preserva as matrizes de covariância cruzada de lag zero e um. Ele também destacou que o modelo poderia ser

simplificado ao usar uma matriz diagonal nos coeficientes do VAR e descreveu um procedimento de estimação de parâmetros baseado no método dos momentos. KUCZERA [35] esclareceu como obter o estimador de máxima verossimilhança em situações de dados faltantes para os parâmetros de um modelo VAR(1) usando o algoritmo EM de DEMPSTER *et al.* [16]. Com o intuito de considerar a sazonalidade presente em algumas séries temporais, YOUNG e PISANO [76] sugeriram primeiro dessazonalizar a série multivariada antes de ajustar o modelo de Matalas. Além disso, eles construíram melhores procedimentos de estimação e sugeriram transformações para a remoção da assimetria dos dados.

Com relação aos modelos VAR(p), PEGRAM e JAMES [49] propuseram um tipo de método de momentos para estimar os parâmetros. No caso de geração de cenários de vazão eles sugerem a diagonalização das matrizes dos coeficientes do VAR(p). Um modelo VAR(p) com parâmetros variando com a sazonalidade foi proposto por SALAS e PEGRAM [58], os quais sugeriram o método dos momentos e máxima verossimilhança para estimar os parâmetros.

FRANCHINI *et al.* [21] desenvolveram um tipo de modelo VAR que possui a característica de preservar as sequências de longa duração e reproduzir as propriedades estatísticas das vazões sazonais em mais de um posto situado em uma dada bacia. Eles apontaram que o seu modelo é capaz de manter as correlações espaço-temporais sazonais, bem como as propriedades anuais das vazões.

Em 1974, MEJÍA *et al.* [45] consideraram a situação onde as séries hidrológicas sintéticas são obtidas de uma mistura de distribuições. Para reproduzir os momentos históricos nas séries simuladas, eles propuseram uma transformação dos dados históricos para a estimação dos parâmetros do modelo. De acordo com STEDINGER [67], este procedimento parece ter pouco fundamento estatístico. Ele mostra que a estimativa direta a partir dos dados históricos pode resultar em estimativas melhores das correlações cruzadas “verdadeiras”.

STEDINGER [67] comparou diferentes abordagens para a estimação da correlação em modelos de vazão multivariados. Ele concluiu que “parece haver pouca justificativa estatística para selecionar um modelo de vazão que reproduza exatamente a correlação histórica observada ... e talvez a lição mais importante a ser aprendida ... é que as estimativas dos parâmetros de muitos modelos de vazão são imprecisas”. Portanto, “é muito razoável usar estimativas estatisticamente eficientes dos parâmetros, mesmo que não se reproduza exatamente as médias, variâncias e correlações das vazões históricas” - STEDINGER e TAYLOR [68].

SALAS *et al.* [59] apresentaram uma ampla revisão das várias abordagens multivariadas de modelagem em hidrologia.

KELMAN *et al.* [33] criaram uma versão multivariada de um modelo proposto por KELMAN [34] que modela separadamente membros ascendentes e descendentes

de hidrogramas diários. A extensão multivariada do modelo segue a abordagem sugerida por MATALAS [42].

SRIKANTHAN [66] propôs um modelo multivariado para a simulação de dados climáticos diários. A precipitação diária foi simulada usando um modelo Markoviano multi-estado de primeira ordem enquanto que as demais variáveis climáticas foram simuladas usando um outro modelo multi-estado (MATALAS [42]; RICHARDSON [54]). NASSERI [47] usou um modelo AR multivariado de ordem um para gerar precipitação horária para uma rede de pluviômetros.

1.3.2 Modelos periódicos autorregressivos

Desde o início da década de 1960 uma quantidade considerável de pesquisas tem sido realizada na área de modelagem periódica. Esta pesquisa inclui contribuições de autores tais como GLADYSHEV [23], GLADYSHEV [24], JONES e BRELSFORD [31], TAO e DELLEUR [69], CROLEY II e RAO [14], MCLEOD e HIPEL [43], PAGANO [48], TROUTMAN [73], DUNSMUIR [17], TIAO e GRUPE [72], SAKAI [57], SALAS *et al.* [59], CIPRA [11], CIPRA [12], VECCHIA [74] e VECCHIA [75], THOMPSTONE *et al.* [71], CIPRA e TLUSTÝ [13], JIMENEZ *et al.* [30] e MCLEOD [44].

A fundamentação da classe de modelos periódicos tem origem nos trabalhos de GLADYSHEV [23] e GLADYSHEV [24]. Esses artigos definiram o conceito de estacionariedade para processos periódicos. O livro de HIPEL e MCLEOD [28] descreve didaticamente o uso dos modelos PAR e PARMA, bem como sua aplicação na modelagem de sistemas hídricos. O artigo de RASMUSSEN *et al.* [53] discute a estimação e validação de modelos PARMA univariados e multivariados aplicáveis à hidrologia.

1.3.3 Modelos multiplicativos

Até a presente seção, todos os modelos supracitados possuem erros aditivos. Modelos de séries temporais com erros multiplicativos são temas recentes de pesquisa. Algumas propostas são encontradas na literatura internacional: LANNE [37], HAUTSCH [26], CIPOLLINI e GALLO [10], BROWNLEES *et al.* [7] e HAUTSCH *et al.* [27]. No Brasil, BRAGA [5] e FERREIRA [18] propuseram um modelo periódico de distribuição gamma para os erros multiplicativos.

1.3.4 A origem da proposta do modelo periódico vetorial autorregressivo multiplicativo

A proposta do modelo periódico vetorial autorregressivo multiplicativo (PVAR_m) foi motivada pelo acordo de cooperação técnica entre ONS e IE-UFRJ ocorrido em 2012. Getúlio B. da Silveira, Joari P. da Costa, Wilson Calmon e Murilo Soares constataram que o modelo periódico autorregressivo com erros multiplicativos atendia ao requisito de linearidade da metodologia de otimização estocástica e reproduzia satisfatoriamente os fatos estilizados da série de ENA. Neste projeto também foram sugeridos estimadores para os parâmetros e distribuição dos ruídos multiplicativos, bem como técnicas de geração de cenários. Esta dissertação foi elaborada a partir da proposta metodológica do acordo de cooperação técnica aliado a contribuição adicional do método de mínimos quadrados com restrição nos coeficientes. Este estimador unido a ruídos multiplicativos positivos garante cenários de ENA sempre positivos.

Capítulo 2

O modelo PVAR

O modelo Periódico Vetorial Autorregressivo (PVAR) satisfaz aos requisitos do SDDP. Atualmente, a versão do PVAR univariada com ruídos correlacionados é utilizada no planejamento de longo prazo do SIN para representar a afluência energética em cada REE. A seguir, serão revisadas as definições e conceitos básicos ligados ao modelo PVAR.

2.1 Processos estacionários e ergódicos

As hipóteses mais fundamentais na modelagem de séries temporais são a *estacionariedade* e a *ergodicidade*. De modo geral, a ideia de estacionariedade está relacionada às propriedades da série temporal que são invariantes no tempo, como a distribuição conjunta das variáveis ou momentos. Já a ergodicidade corresponde a alguns parâmetros (momentos, tipicamente) da série poderem ser estimados consistentemente a partir de seus análogos amostrais. Seja \mathbf{a}_t o vetor de ENAs no tempo t onde cada componente representa a Energia Natural Afluente a um dado REE. O processo estocástico $\{\mathbf{a}_1, \mathbf{a}_2, \dots\}$ é dito *fracamente estacionário* se

$$\begin{aligned} \mathbb{E}[\mathbf{a}_t] &= \mathbb{E}[\mathbf{a}_{t+1}] \\ \text{Cov}[\mathbf{a}_t, \mathbf{a}_{t+k}] &= \text{Cov}[\mathbf{a}_{t+1}, \mathbf{a}_{t+1+k}] \end{aligned}, \quad (2.1)$$

para todo inteiro positivo t e k , isto é, se a respectiva média é constante e a covariância só depende do lag k . Adicionalmente, um processo estocástico $\{\mathbf{a}_1, \mathbf{a}_2, \dots\}$ fracamente estacionário é dito *ergódico* se

$$\begin{aligned} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{\nu=1}^T \mathbf{a}_\nu \stackrel{\text{w.p.1}}{=} \mathbb{E}[\mathbf{a}_t] \\ \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{\nu=1}^{T-k} \mathbf{a}_\nu \mathbf{a}_{\nu+k}^\top \stackrel{\text{w.p.1}}{=} \mathbb{E}[\mathbf{a}_t \mathbf{a}_{t+k}^\top] \end{aligned}, \quad (2.2)$$

para todo inteiro positivo t e k , ou seja, as médias temporais convergem com probabilidade 1 (w. p. 1) para as respectivas esperanças. Vale ressaltar que essas definições seguem de forma análoga para momentos de ordem maiores do que 2. Maiores detalhes podem ser encontrados no livro de HANNAN [25].

Apesar dessas hipóteses serem os pressupostos *usuais* de estacionaridade e ergodicidade, a análise da série de ENA indica que a esperança de cada mês não é constante, sugerindo que essas hipóteses não são adequadas. Uma hipótese admissível para a série de ENA é a *invariância anual* das propriedades de cada mês, ou seja, as características da série temporal são periódicas de período 12. As definições de estacionaridade e ergodicidade análogas a (2.1) e (2.2) para o caso periódico serão apresentados a seguir. O processo estocástico vetorial $\{\mathbf{a}_1, \mathbf{a}_2, \dots\}$ é dito *periodicamente correlacionado* de período \mathbb{S} se

$$\begin{aligned} \mathbb{E}[\mathbf{a}_t] &= \mathbb{E}[\mathbf{a}_{t+\mathbb{S}}] \\ \text{Cov}[\mathbf{a}_t, \mathbf{a}_{t+k}] &= \text{Cov}[\mathbf{a}_{t+\mathbb{S}}, \mathbf{a}_{t+\mathbb{S}+k}] \end{aligned}, \quad (2.3)$$

para todo inteiro positivo t e k , isto é, se a respectiva média é constante módulo \mathbb{S} e a covariância módulo \mathbb{S} só depende do lag k . Adicionalmente, um processo estocástico $\{\mathbf{a}_1, \mathbf{a}_2, \dots\}$ periodicamente correlacionado de período \mathbb{S} é dito *periodicamente ergódico* se

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{1}{N+1} \sum_{\nu=0}^N \mathbf{a}_{s+\nu\mathbb{S}} &\stackrel{\text{w.p.1}}{=} \mathbb{E}[\mathbf{a}_s] \\ \lim_{N \rightarrow \infty} \frac{1}{N+1} \sum_{\nu=0}^N \mathbf{a}_{s+\nu\mathbb{S}} \mathbf{a}_{u+\nu\mathbb{S}}^\top &\stackrel{\text{w.p.1}}{=} \mathbb{E}[\mathbf{a}_s \mathbf{a}_u^\top] \end{aligned}, \quad (2.4)$$

onde s e u são números entre 1 e \mathbb{S} .

O valor s é chamado de *estação* associada ao tempo t se somando \mathbb{S} uma quantidade suficiente de vezes a s se obtém t . Matematicamente, a estação s associada ao tempo t pode ser expressa por

$$s = (t - 1) \% \mathbb{S} + 1$$

onde $a \% b$ é o resto da divisão de a por b . Com isso, é possível afirmar que num processo periodicamente ergódico como (2.4) a média temporal dos termos associados a estação s converge com probabilidade 1 para a respectiva esperança, em toda estação s .

2.2 Formulação do modelo PVAR

Um processo aleatório $\{\mathbf{a}_t\}_{t=1}^{\infty}$ com \mathbf{a}_t tomando valores em \mathbb{R}^d é dito Periódico Vetorial Autorregressivo de período \mathbb{S} e ordem $\mathbf{p} = (p_1, \dots, p_{\mathbb{S}})$, PVAR(\mathbf{p}), se ¹

$$\mathbf{a}_t = \zeta_t + \sum_{\nu=1}^{p_t} \Phi_{t,\nu} \mathbf{a}_{t-\nu} + \epsilon_t, \quad (2.6)$$

para todo inteiro positivo t , onde os erros *aditivos* $\{\epsilon_1, \epsilon_2, \dots\}$ são independentes. Para cada tempo t , os coeficientes $\Phi_{t,\nu}$ e as ordens do modelo são iguais módulo \mathbb{S} ; os erros são identicamente distribuídos módulo \mathbb{S} com média zero e variância constante módulo \mathbb{S} . Mais formalmente, para cada t :

- $\zeta_t = \zeta_{t+\mathbb{S}}$ e $\Phi_{t,\nu} = \Phi_{t+\mathbb{S},\nu}$;
- $p_t = p_{t+\mathbb{S}}$;
- ϵ_t possui a mesma distribuição de $\epsilon_{t+\mathbb{S}}$; e
- $\mathbb{E}[\epsilon_t] = 0$.

Vale ressaltar que o modelo PVAR apresentado é vetorial, pois $\{\mathbf{a}_1, \mathbf{a}_2, \dots\}$ é uma série vetorial. Por conseguinte, $\Phi_{t,1}, \dots, \Phi_{t,p_t}$ são matrizes e ζ_t, ϵ_t são vetores. Observa-se também que o modelo VAR é o caso particular do PVAR quando os coeficientes e a distribuição do erro não variam no tempo, isto é, quando $\mathbb{S} = 1$. Uma referência sobre modelos PVAR é o livro de FRANSES e PAAP [22].

2.2.1 Estacionariedade e causalidade do PVAR

A seção 2.1 considera que as observações foram obtidas por um processo estocástico $\{\mathbf{a}_1, \mathbf{a}_2, \dots\}$ periodicamente correlacionado. Assim, é instrutivo avaliar se a proposta do modelo PVAR reproduz a propriedade de correlação periódica (2.3). Esta avaliação será conduzida considerando sucessivamente os modelos AR(1), AR(p), VAR(p) e finalmente PVAR(p).

¹A definição do modelo PVAR apresentada na literatura, em geral, é a versão *centrada*:

$$\mathbf{a}_t - \boldsymbol{\mu}_t = \sum_{\nu=1}^{p_t} \Phi_{t,\nu} (\mathbf{a}_{t-\nu} - \boldsymbol{\mu}_{t-\nu}) + \epsilon_t, \quad (2.5)$$

onde $\boldsymbol{\mu}_t := \mathbb{E}[\mathbf{a}_t]$. A relação entre o PVAR (2.6) e sua versão centrada (2.5) é estabelecida ao tomar a esperança da equação (2.6) e observar que ζ_t é igual a $\boldsymbol{\mu}_t - \sum_{\nu=1}^{p_t} \Phi_{t,\nu} \boldsymbol{\mu}_{t-\nu}$. A motivação da definição (2.6) está na conveniência de seu uso na descrição do método de mínimos quadrados com restrição nos coeficientes.

Estacionariedade e causalidade de um AR(1)

Por simplicidade, considera-se inicialmente o caso particular univariado de um AR(1) centrado (i.e. $\mathbb{E}[a_t] = 0$ para todo t):

$$a_t = \phi_1 a_{t-1} + \epsilon_t, \quad (2.7)$$

isto é, um PAR centrado de período e ordem iguais a 1. Com isso, o requisito de correlação periódica (2.3) se reduz à estacionariedade fraca (2.1).

A análise de estacionariedade de um modelo de série temporal é guiada pelo conceito de um *processo linear*. Diz-se que $\{a_1, a_2, \dots\}$ é um *processo linear* se cada termo a_t pode ser representado pela seguinte combinação linear infinita de termos não correlacionados:

$$a_t = \sum_{j=-\infty}^{\infty} \psi_j \tilde{\epsilon}_{t-j}, \quad (2.8)$$

para todo t , onde $\{\tilde{\epsilon}_j\}_{j=-\infty}^{\infty}$ é uma sequência de variáveis não correlacionadas de média 0 e variância σ^2 . Os coeficientes $\{\psi_j\}_{j=-\infty}^{\infty}$ são constantes cuja respectiva série converge absolutamente, ou seja, $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$.

A análise de estacionariedade pode ser reduzida a avaliar se o processo aleatório $\{a_1, a_2, \dots\}$ é um processo linear. O teorema de *decomposição de Wold* garante que todo processo fracamente estacionário é um processo linear ou pode ser transformado em um processo linear ao subtrair sua componente determinística (BROCKWELL e DAVIS [6]). Por outro lado, todo processo linear $\{a_1, a_2, \dots\}$ é um processo fracamente estacionário. Com efeito, como a média de cada $\tilde{\epsilon}_t$ é zero então

$$\mathbb{E}[a_t] = \mathbb{E} \left[\sum_{j=-\infty}^{\infty} \psi_j \tilde{\epsilon}_{t-j} \right] = \sum_{j=-\infty}^{\infty} \psi_j \mathbb{E}[\tilde{\epsilon}_{t-j}] = 0,$$

ou seja, a média de cada a_t também é zero. Dado que as variáveis $\{\tilde{\epsilon}_j\}_{j=-\infty}^{\infty}$ são descorrelacionadas então

$$\begin{aligned} \text{Cov}[a_t, a_{t+h}] &= \mathbb{E}[a_t a_{t+h}] \\ &= \mathbb{E} \left[\left(\sum_{j=-\infty}^{\infty} \psi_j \tilde{\epsilon}_{t-j} \right) \left(\sum_{k=-\infty}^{\infty} \psi_k \tilde{\epsilon}_{t+h-k} \right) \right] \\ &= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \mathbb{E}(\tilde{\epsilon}_{t-j} \tilde{\epsilon}_{t+h-k}) \\ &= \sum_{j=-\infty}^{\infty} \psi_j \psi_{j-h} \sigma^2, \end{aligned}$$

ou seja, a covariância $\text{Cov}[a_t, a_{t+h}]$ só depende do lag h .

Uma ideia para verificar a estacionariedade fraca de um modelo AR(1) é manipular a equação (2.7) com o intuito de produzir a representação (2.8) de um processo linear. Se a série dos correspondentes coeficientes convergir absolutamente, então assegura-se que o processo aleatório $\{a_1, a_2, \dots\}$ é um processo linear. Assim, prova-se que o modelo AR(1) é fracamente estacionário.

No caso do modelo AR(1), uma ideia para se alcançar a representação (2.8) de um processo linear é substituir sucessivamente a definição (2.7) do modelo autorregressivo nos termos antecessores a a_t :

$$\begin{aligned} a_t &= \epsilon_t + \phi_1 a_{t-1} \\ &= \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_1^2 a_{t-2} \\ &= \dots \\ &= \epsilon_t + \phi_1 \epsilon_{t-1} + \dots + \phi_1^k \epsilon_{t-k} + \phi_1^{k+1} a_{t-k-1}. \end{aligned} \quad (2.9)$$

Baseado na equação (2.9), a expressão geral para a_t é

$$a_t = \sum_{j=0}^k \phi_1^j \epsilon_{t-j} + \phi_1^{k+1} a_{t-k-1}, \quad (2.10)$$

para todo k inteiro positivo. Se $|\phi_1| < 1$, tem-se a motivação de representar a_t por meio dos ruídos aditivos passados até t :

$$a_t = \sum_{j=0}^{\infty} \phi_1^j \epsilon_{t-j}. \quad (2.11)$$

A série (2.11) satisfaz exatamente a definição de um processo linear, pois os ruídos aditivos $\{\epsilon_t\}_{t=0}^{\infty}$ são i.i.d.'s com média 0, variância σ^2 e a série dos respectivos coeficientes converge absolutamente:

$$\sum_{j=0}^{\infty} |\phi_1|^j = \frac{1}{1 - |\phi_1|} < \infty.$$

Além disso, a variável aleatória definida como o limite da série (2.11) satisfaz à definição do modelo AR(1). Por outro lado, se $|\phi_1| \geq 1$ então a série dos coeficientes de (2.11) não converge absolutamente, mas isso não significa que o modelo não seja um processo linear (2.8) e sim que essa representação não é válida.

Uma outra maneira de representar os termos do modelo AR é substituindo a definição do modelo autorregressivo nos termos sucessores de a_t . Pois, dada a

definição do modelo AR(1) para o termo a_{t+1} ,

$$a_{t+1} = \phi_1 a_t + \epsilon_{t+1},$$

é possível reescrever a_t em função de a_{t+1} :

$$a_t = \frac{-\epsilon_{t+1} + a_{t+1}}{\phi_1} = -\phi_1^{-1}\epsilon_{t+1} + \phi_1^{-1}a_{t+1}.$$

Repetindo esse procedimento sucessivamente, obtém-se as identidades abaixo:

$$\begin{aligned} a_t &= -\phi_1^{-1}\epsilon_{t+1} + \phi_1^{-1}a_{t+1} \\ &= -\phi_1^{-1}\epsilon_{t+1} - \phi_1^{-2}\epsilon_{t+2} + \phi_1^{-2}a_{t+2} \\ &= \dots \\ &= -\phi_1^{-1}\epsilon_{t+1} - \phi_1^{-2}\epsilon_{t+2} - \dots - \phi_1^{-k}\epsilon_{t+k} + \phi_1^{-k}a_{t+k}. \end{aligned} \quad (2.12)$$

Baseado na equação (2.12), uma outra expressão para a_t é

$$a_t = \sum_{j=1}^k -\phi_1^{-j}\epsilon_{t+j} + \phi_1^{-k}a_{t+k} \quad (2.13)$$

para todo k inteiro positivo. Se $|\phi_1| > 1$, tem-se a motivação de representar a_t por meio dos ruídos aditivos futuros a partir de $t + 1$:

$$a_t = \sum_{j=1}^{\infty} -\phi_1^{-j}\epsilon_{t+j}. \quad (2.14)$$

A série (2.14) satisfaz exatamente a definição de um processo linear, pois os ruídos aditivos $\{\epsilon_t\}_{t=0}^{\infty}$ são i.i.d.'s com média 0, variância σ^2 e a série dos respectivos coeficientes converge absolutamente:

$$\sum_{j=1}^{\infty} |\phi_1|^{-j} = \frac{|\phi_1|^{-1}}{1 - |\phi_1|^{-1}} = \frac{1}{|\phi_1| - 1} < \infty.$$

Adicionalmente, a variável aleatória definida como o limite da série (2.14) satisfaz à definição do modelo AR(1).

Por fim, se $|\phi_1| = 1$ então nenhuma das séries dos coeficientes de (2.13) e (2.14) converge absolutamente. Neste caso, o modelo AR(1) é não estacionário. Com efeito, supondo por absurdo que a série $\{a_1, a_2, \dots\}$ gerada pelo modelo autorregressivo seja

estacionária é possível notar a seguinte identidade:

$$\begin{aligned}\text{Var}[a_t - \phi_1^{n+1} a_{t-n-1}] &= \text{Var}[a_t] - 2\phi_1^{n+1} \text{Cov}[a_t, a_{t-n-1}] + \phi_1^{2(n+1)} \text{Var}[a_{t-n-1}] \\ &= 2\gamma(0) - 2\phi_1^{n+1}\gamma(n+1).\end{aligned}\quad (2.15)$$

Por outro lado, lembrando a identidade (2.10):

$$a_t = \sum_{j=0}^n \phi_1^j \epsilon_{t-j} + \phi_1^{n+1} a_{t-n-1},$$

é possível obter uma outra relação para $\text{Var}[a_t - \phi_1^{n+1} a_{t-n-1}]$:

$$\begin{aligned}\text{Var}[a_t - \phi_1^{n+1} a_{t-n-1}] &= \text{Var}\left[\sum_{j=0}^n \phi_1^j \epsilon_{t-j}\right] \\ &= \text{Cov}\left[\sum_{i=0}^n \phi_1^i \epsilon_{t-i}, \sum_{j=0}^n \phi_1^j \epsilon_{t-j}\right] \\ &= \sum_{i=0}^n \sum_{j=0}^n \phi_1^{i+j} \text{Cov}[\epsilon_{t-i}, \epsilon_{t-j}] \\ &= \sum_{i=0}^n \phi_1^{2i} \sigma^2 = (n+1)\sigma^2.\end{aligned}\quad (2.16)$$

Igualando (2.15) com (2.16) segue:

$$2\gamma(0) - 2\phi_1^{n+1}\gamma(n+1) = (n+1)\sigma^2.\quad (2.17)$$

Dividindo ambos os lados de (2.17) por $\gamma(0)$, tem-se:

$$2 - 2\phi_1^{n+1}\rho(n+1) = (n+1)\frac{\sigma^2}{\gamma(0)},\quad (2.18)$$

onde $\rho(n+1) = \gamma(n+1)/\gamma(0)$ é a função de autocorrelação de lag $n+1$. Como o lado esquerdo de (2.18) é limitado e o lado direito é ilimitado em n , observa-se uma contradição. Portanto, quando $|\phi_1| = 1$ o modelo AR(1) é não estacionário.

Em suma, descreve-se três possibilidades para um AR(1):

1. $|\phi_1| < 1$: O modelo é estacionário, pois é um processo linear. Cada termo a_t pode ser representado como uma combinação linear infinita dos erros *passados* até t :

$$a_t = \sum_{j=0}^{\infty} \phi_1^j \epsilon_{t-j}.\quad (2.19)$$

Um modelo com essa característica é dito *causal*.

2. $|\phi_1| > 1$: O modelo é estacionário, pois também é um processo linear. Cada termo a_t pode ser representado como uma combinação linear infinita dos erros *futuros* a partir de $t + 1$:

$$a_t = \sum_{j=1}^{\infty} -\phi_1^{-j} \epsilon_{t+j}. \quad (2.20)$$

3. $|\phi_1| = 1$: O modelo é *não* estacionário.

A representação (2.20) é frequentemente vista como artificial, pois a observação de a_t depende dos valores *futuros* de erro $\{\epsilon_{t+1}, \epsilon_{t+2}, \dots\}$. É habitual em modelagem de séries temporais estacionárias restringir a atenção a processos que dependam apenas dos erros *passados*, ou seja, a processos *causais*. Nesta dissertação considera-se apenas modelos estacionários causais (o que no caso de um AR(1) significa ϕ_1 em módulo ser menor do que 1).

Estacionariedade e causalidade de um AR(p)

Um outro ponto de vista da argumentação anterior, importante para sua generalização, é obtido observando que o modelo AR(1) pode ser escrito da seguinte forma:

$$\phi(B)a_t = \epsilon_t, \quad (2.21)$$

onde $\phi(B) := 1 - \phi_1 B$ e B é o operador de *defasagem* (ou backshift) que associa a_t a a_{t-1} . Nesta perspectiva, as equações (2.19) e (2.20) são o resultado da aplicação da inversa do operador $\phi(B)$ na equação (2.21):

$$a_t = \phi(B)^{-1} \epsilon_t. \quad (2.22)$$

A partir das expressões resultantes é possível extrair a inversa de $\phi(B)$:

$$\phi(B)^{-1} = \begin{cases} \sum_{j=0}^{\infty} \phi_1^j B^j & , \text{ se } |\phi_1| < 1 \\ \sum_{j=1}^{\infty} -\phi_1^{-j} B^{-j} & , \text{ se } |\phi_1| > 1 \end{cases}, \quad (2.23)$$

onde B^j é o operador de defasagem aplicado j vezes, isto é, $B^j \epsilon_t = \epsilon_{t-j}$, e B^{-j} é a inversa do operador de defasagem aplicada j vezes, ou seja, $B^{-j} \epsilon_t = \epsilon_{t+j}$. Além disso, se $|\phi_1| = 1$ então o operador $\phi(B)$ é não inversível, pois caso contrário o modelo AR com esse coeficiente seria fracamente estacionário.

Assim, um procedimento geral para se analisar a estacionariedade de um modelo AR se baseia em buscar condições sobre os respectivos coeficientes que garantam a existência da inversa do operador $\phi(B)$. Adicionalmente, a inversa de $\phi(B)$ deve ser expressa por uma série de potência cujos coeficientes formam uma série absolutamente convergente. Por exemplo, no caso de um modelo AR(p) *centrado*:

$$a_t = \phi_1 a_{t-1} + \cdots + \phi_p a_{t-p} + \epsilon_t,$$

o operador $\phi(B)$ é definido por

$$\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$$

e a inversa deve ser expressa por

$$\phi(B)^{-1} = \sum_{j=-\infty}^{\infty} \psi_j B^j,$$

onde a série formada pelos ψ_j 's converge absolutamente, isto é, $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$.

De acordo com o livro de BROCKWELL e DAVIS [6] (pag. 85) uma condição necessária e suficiente para a existência de $\phi(B)^{-1}$ é o polinômio induzido por $\phi(B)$ nos complexos

$$p(z) := 1 - \phi_1 z - \cdots - \phi_p z^p$$

não ter nenhuma raiz na circunferência unitária, ou seja,

$$p(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \neq 0, \quad \text{para todo } |z| = 1.$$

Por exemplo, o polinômio complexo induzido por $\phi(B)$ no caso do modelo AR(1)

$$p(z) = 1 - \phi_1 z,$$

possui uma única raiz que é $z^* = 1/\phi_1$. Observa-se que se $|\phi_1| < 1$ ou $|\phi_1| > 1$ então a raiz z^* está *fora* da circunferência unitária complexa e portanto o modelo é estacionário. Porém, se $|\phi_1| = 1$ então a raiz z^* *pertence* à circunferência unitária e por isso o modelo é não estacionário. Pela mesma referência supracitada, uma condição necessária e suficiente para $\phi(B)^{-1}$ produzir um modelo *causal* é o polinômio complexo induzido $p(z)$ não ter nenhuma raiz no disco fechado unitário, isto é,

$$p(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \neq 0, \quad \text{para todo } |z| \leq 1.$$

Considerando o mesmo exemplo do AR(1), a raiz $z^* = 1/\phi_1$ não pertence ao disco

fechado unitário se, e somente se, $|\phi_1| < 1$.

Estacionariedade e causalidade de um VAR(p)

O conceito de causalidade para um processo multivariado é semelhante ao correspondente univariado. Uma série temporal *vetorial* $\{\mathbf{a}_1, \mathbf{a}_2, \dots\}$ é dita um processo linear se admite a seguinte representação:

$$\mathbf{a}_t = \sum_{j=-\infty}^{\infty} \Psi_j \boldsymbol{\epsilon}_{t-j}$$

para todo t , onde $\{\boldsymbol{\epsilon}_j\}_{j=-\infty}^{\infty}$ é uma sequência de variáveis não correlacionadas de média $\mathbf{0}$ e matriz de covariância Σ . Além disso, a série formada pelas componentes da matriz Ψ_j deve convergir absolutamente. Um processo vetorial é dito causal se a variável \mathbf{a}_t depende apenas dos ruídos passados até t , ou seja,

$$\mathbf{a}_t = \sum_{j=0}^{\infty} \Psi_j \boldsymbol{\epsilon}_{t-j}.$$

Um exemplo de um processo multivariado é o modelo *autorregressivo vetorial* (VAR). O VAR centrado de ordem p é definido por

$$\mathbf{a}_t = \Phi_1 \mathbf{a}_{t-1} + \dots + \Phi_p \mathbf{a}_{t-p} + \boldsymbol{\epsilon}_t$$

para todo inteiro positivo t , onde os erros $\{\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2, \dots\}$ são independentes, identicamente distribuídos de média $\mathbf{0}$ e matriz de covariância Σ . Uma condição necessária e suficiente descrita em BROCKWELL e DAVIS [6] (pag. 242) para que o modelo VAR(p) seja causal é

$$\det P(z) \neq 0, \quad \text{para todo } |z| \leq 1, \quad (2.24)$$

onde o polinômio $P(z)$ é definido por

$$P(z) = I - \Phi_1 z - \dots - \Phi_p z^p.$$

Observa-se que $P(\cdot)$ associa um número complexo z a uma matriz complexa.

Estacionariedade e causalidade de um PVAR(p)

Finalmente, a definição de causalidade para um processo multivariado *periódico* é similar ao correspondente aperiódico. De acordo com JAHEL [29] e BOSHPNAKOV [4] uma série temporal $\{\mathbf{a}_1, \mathbf{a}_2, \dots\}$ de média $\mathbf{0}$ é *periodicamente causal* se cada

termo \mathbf{a}_t possui a seguinte representação:

$$\mathbf{a}_t = \sum_{j=0}^{\infty} \Psi_{t,j} \boldsymbol{\epsilon}_{t-j}, \quad (2.25)$$

onde $\Psi_{t,j}$ são matrizes \mathbb{S} -periódicas, ou seja, $\Psi_{t,j} = \Psi_{t+\mathbb{S},j}$, a série formada pelas componentes da matriz $\Psi_{t,j}$ somadas em j converge absolutamente e os ruídos $\{\boldsymbol{\epsilon}_j\}_{j=-\infty}^{\infty}$ são não correlacionados, possuem média zero e matriz de covariância constante módulo \mathbb{S} . Uma condição necessária e suficiente para a causalidade periódica do modelo PVAR(\mathbf{p}) é descrita em JAHEL [29]. A ideia consiste em converter o PVAR em um VAR de dimensão maior e aplicar a caracterização de causalidade (2.24) do VAR.

2.2.2 Propriedades do modelo PVAR *causal*

Considerando o modelo PVAR(\mathbf{p}) (2.6), define-se o *Melhor Preditor Linear* de \mathbf{a}_t dada as observações até $t-1$, $\mathbf{a}_{[t-1]} := (\mathbf{a}_1, \dots, \mathbf{a}_{t-1})$, por

$$\text{Prev}[\mathbf{a}_t \mid \mathbf{a}_{[t-1]}] = \zeta_t + \sum_{\nu=1}^{p_t} \Phi_{t,\nu} \mathbf{a}_{t-\nu}. \quad (2.26)$$

Assume-se que as afluências com índice negativo $\mathbf{a}_0, \mathbf{a}_{-1}, \mathbf{a}_{-2}, \dots$ são condições iniciais da série e que são conhecidas. Tomando a esperança condicional dadas as observações até $t-1$ em ambos os lados da equação (2.6), obtém-se a seguinte relação envolvendo a previsão de \mathbf{a}_t :

$$\begin{aligned} \mathbb{E}[\mathbf{a}_t \mid \mathbf{a}_{[t-1]}] &= \zeta_t + \sum_{\nu=1}^{p_t} \Phi_{t,\nu} \mathbf{a}_{t-\nu} + \mathbb{E}[\boldsymbol{\epsilon}_t \mid \mathbf{a}_{[t-1]}] \\ &= \text{Prev}[\mathbf{a}_t \mid \mathbf{a}_{[t-1]}] + \mathbb{E}[\boldsymbol{\epsilon}_t \mid \mathbf{a}_{[t-1]}]. \end{aligned} \quad (2.27)$$

Analogamente, a variância condicional é dada por:

$$\begin{aligned} \text{Var}[\mathbf{a}_t \mid \mathbf{a}_{[t-1]}] &= \text{Var} \left[\zeta_t + \sum_{\nu=1}^{p_t} \Phi_{t,\nu} \mathbf{a}_{t-\nu} + \boldsymbol{\epsilon}_t \mid \mathbf{a}_{[t-1]} \right] \\ &= \text{Var}[\boldsymbol{\epsilon}_t \mid \mathbf{a}_{[t-1]}]. \end{aligned} \quad (2.28)$$

Como apresentado na seção sobre a correlação periódica do PVAR, cada termo \mathbf{a}_t de um PVAR *causal* (2.25) pode ser representado em função dos erros passados até t . Como por hipótese os erros $\boldsymbol{\epsilon}_t$ são independentes, as variáveis $\mathbf{a}_{[t-1]}$ e $\boldsymbol{\epsilon}_t$ são independentes. Desta forma, as equações (2.27) e (2.28), no caso de um PVAR

causal, podem ser reescritas sem o condicionante $\mathbf{a}_{[t-1]}$ no termo de erro $\boldsymbol{\epsilon}_t$:

$$\mathbb{E}[\mathbf{a}_t | \mathbf{a}_{[t-1]}] = \text{Prev}[\mathbf{a}_t | \mathbf{a}_{[t-1]}] + \mathbb{E}[\boldsymbol{\epsilon}_t], \quad (2.29)$$

$$\text{Var}[\mathbf{a}_t | \mathbf{a}_{[t-1]}] = \text{Var}[\boldsymbol{\epsilon}_t]. \quad (2.30)$$

Por hipótese, a média do erro $\boldsymbol{\epsilon}_t$ é $\mathbf{0}$ e a variância é constante sazonalmente. Essas informações modificam as expressões (2.29) e (2.30) da seguinte forma:

$$\mathbb{E}[\mathbf{a}_t | \mathbf{a}_{[t-1]}] = \text{Prev}[\mathbf{a}_t | \mathbf{a}_{[t-1]}], \quad (2.31)$$

$$\text{Var}[\mathbf{a}_t | \mathbf{a}_{[t-1]}] = \Sigma_t, \quad (2.32)$$

onde $\Sigma_t := \text{Var}[\boldsymbol{\epsilon}_t]$ e $\Sigma_t = \Sigma_{t+\mathbb{S}}$, ou seja, a previsão de \mathbf{a}_t se torna igual a média condicional e a variância condicional (volatilidade) de \mathbf{a}_t se torna constante módulo \mathbb{S} . Uma interpretação simples deste resultado é: as previsões do modelo PVAR são *não-viesadas* e a volatilidade depende apenas da sazonalidade e não das observações passadas.

2.3 Estimação dos parâmetros

Nesta seção serão apresentados três métodos de estimação de parâmetros: o método de Yule-Walker, o método de mínimos quadrados e o método de mínimos quadrados com restrição nos coeficientes.

2.3.1 Método de Yule-Walker

O método de Yule-Walker se baseia em construir um sistema linear para os coeficientes do modelo PVAR a partir de sua definição e propriedades. Esta construção é feita de modo que o número de equações seja igual ao número de incógnitas e com isso seja explícito o cálculo deste estimador. A seguir, deduz-se o sistema linear utilizado para calcular o estimador de Yule-Walker. Uma dedução para o caso univariado pode ser encontrada no livro de HIPEL e MCLEOD [28].

O sistema de Yule-Walker é deduzido com base na versão *centrada* do modelo PVAR:

$$\mathbf{a}_t - \boldsymbol{\mu}_t = \sum_{\nu=1}^{p_t} \Phi_{t,\nu}(\mathbf{a}_{t-\nu} - \boldsymbol{\mu}_{t-\nu}) + \boldsymbol{\epsilon}_t, \quad (2.33)$$

onde $\boldsymbol{\mu}_t := \mathbb{E}[\mathbf{a}_t]$. A relação entre o PVAR (2.6) e sua versão centrada (2.33) é estabelecida ao tomar a esperança da equação (2.6) e observar que ζ_t é igual a $\boldsymbol{\mu}_t - \sum_{\nu=1}^{p_t} \Phi_{t,\nu} \boldsymbol{\mu}_{t-\nu}$. Multiplicando em ambos os lados de (2.33) por $(\mathbf{a}_{t-k} - \boldsymbol{\mu}_{t-k})^\top$

e aplicando a esperança, obtém-se a seguinte equação:

$$\begin{aligned} \mathbb{E} [(\mathbf{a}_t - \boldsymbol{\mu}_t)(\mathbf{a}_{t-k} - \boldsymbol{\mu}_{t-k})^\top] &= \sum_{\nu=1}^{p_t} \Phi_{t,\nu} \mathbb{E} [(\mathbf{a}_{t-\nu} - \boldsymbol{\mu}_{t-\nu})(\mathbf{a}_{t-k} - \boldsymbol{\mu}_{t-k})^\top] + \\ &+ \mathbb{E} [\boldsymbol{\epsilon}_t(\mathbf{a}_{t-k} - \boldsymbol{\mu}_{t-k})^\top]. \end{aligned} \quad (2.34)$$

Observa-se que cada termo de (2.34) pode ser representado como a covariância entre os respectivos vetores aleatórios:

$$\text{Cov} [\mathbf{a}_t, \mathbf{a}_{t-k}] = \sum_{\nu=1}^{p_t} \Phi_{t,\nu} \text{Cov} [\mathbf{a}_{t-\nu}, \mathbf{a}_{t-k}] + \text{Cov} [\boldsymbol{\epsilon}_t, \mathbf{a}_{t-k}]. \quad (2.35)$$

Na equação (2.35), o único termo que não pode ser estimado a partir dos dados é a covariância entre o erro $\boldsymbol{\epsilon}_t$ e a afluência \mathbf{a}_{t-k} , pois não é possível determinar os valores de erro antes de se estimar os coeficientes do modelo. Deste modo, é necessário investigar outras formas de se expressar $\text{Cov} [\boldsymbol{\epsilon}_t, \mathbf{a}_{t-k}]$. Supondo a condição de causalidade do modelo PVAR, obtém-se como consequência a independência entre $\boldsymbol{\epsilon}_t$ e \mathbf{a}_{t-k} para todo k maior do que ou igual a 1. Em particular, a covariância entre $\boldsymbol{\epsilon}_t$ e \mathbf{a}_{t-k} é 0, ou seja, $\text{Cov} [\boldsymbol{\epsilon}_t, \mathbf{a}_{t-k}] = 0$ para todo k maior do que ou igual a 1. Quando k é igual a 0, substitui-se a expressão do modelo PVAR (2.33) em $\text{Cov} [\boldsymbol{\epsilon}_t, \mathbf{a}_t]$:

$$\begin{aligned} \text{Cov} [\boldsymbol{\epsilon}_t, \mathbf{a}_t] &= \text{Cov} \left[\boldsymbol{\epsilon}_t, \left(\zeta_t + \sum_{\nu=1}^{p_t} \Phi_{t,\nu}(\mathbf{a}_{t-\nu} - \boldsymbol{\mu}_{t-\nu}) + \boldsymbol{\epsilon}_t \right) \right] \\ &= \sum_{\nu=1}^{p_t} \text{Cov} [\boldsymbol{\epsilon}_t, \Phi_{t,\nu} \mathbf{a}_{t-\nu}] + \text{Cov} [\boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_t] \\ &= \sum_{\nu=1}^{p_t} \text{Cov} [\boldsymbol{\epsilon}_t, \mathbf{a}_{t-\nu}] \Phi_{t,\nu}^\top + \Sigma_t \\ &= \Sigma_t, \end{aligned}$$

Dessa forma, a equação (2.35) é melhor descrita considerando as relações obtidas para $\text{Cov} [\boldsymbol{\epsilon}_t, \mathbf{a}_t]$:

$$\text{Cov} [\mathbf{a}_t, \mathbf{a}_{t-k}] = \sum_{\nu=1}^{p_t} \Phi_{t,\nu} \text{Cov} [\mathbf{a}_{t-\nu}, \mathbf{a}_{t-k}] + \delta_{k,0} \Sigma_t, \quad \forall k \geq 0, \quad (2.36)$$

onde $\delta_{k,0}$ é a função delta de Kronecker que vale 1 se k é igual a 0 e vale 0 se k é diferente de 0.

Com o intuito de facilitar a exposição das equações, denota-se a autocovariância da série temporal \mathbf{a}_t por $\Gamma_t(k)$, ou seja,

$$\Gamma_t(k) := \text{Cov} [\mathbf{a}_t, \mathbf{a}_{t-k}].$$

Pela simetria de $\text{Cov}[\cdot, \cdot]$, deduz-se a seguinte propriedade para $\Gamma_t(k)$:

$$\Gamma_t(k) = \text{Cov}[\mathbf{a}_t, \mathbf{a}_{t-k}] = \text{Cov}[\mathbf{a}_{t-k}, \mathbf{a}_t]^\top = \Gamma_{t-k}(-k)^\top. \quad (2.37)$$

Por ser mais sucinta, esta notação é usada para se reescrever a expressão (2.36):

$$\Gamma_t(k) = \sum_{\nu=1}^{p_t} \Phi_{t,\nu} \Gamma_{t-k}(\nu - k) + \delta_{k,0} \Sigma_t, \quad \forall k \geq 0. \quad (2.38)$$

Observa-se que a equação (2.38) com k igual a 0 determina o valor de Σ_t ,

$$\Sigma_t = \Gamma_t(0) - \sum_{\nu=1}^{p_t} \Phi_{t,\nu} \Gamma_t(\nu). \quad (2.39)$$

Além disso, transpondo as respectivas matrizes em ambos os lados da equação (2.38), obtém-se:

$$\Gamma_t(k)^\top = \sum_{\nu=1}^{p_t} \Gamma_{t-k}(\nu - k)^\top \Phi_{t,\nu}^\top, \quad \forall k \geq 1. \quad (2.40)$$

Para k entre 1 e p_t , a equação (2.40) determina o seguinte sistema formado por *blocos de matrizes*:

$$\begin{bmatrix} \Gamma_t(1)^\top \\ \Gamma_t(2)^\top \\ \vdots \\ \Gamma_t(p_t)^\top \end{bmatrix} = \begin{bmatrix} \Gamma_{t-1}(0)^\top & \Gamma_{t-1}(1)^\top & \cdots & \Gamma_{t-1}(p_t - 1)^\top \\ \Gamma_{t-2}(-1)^\top & \Gamma_{t-2}(0)^\top & \cdots & \Gamma_{t-2}(p_t - 2)^\top \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{t-p_t}(1 - p_t)^\top & \Gamma_{t-p_t}(2 - p_t)^\top & \cdots & \Gamma_{t-p_t}(0)^\top \end{bmatrix} \begin{bmatrix} \Phi_{t,1}^\top \\ \Phi_{t,2}^\top \\ \vdots \\ \Phi_{t,p_t}^\top \end{bmatrix}. \quad (2.41)$$

Pela propriedade $\Gamma_t(k) = \Gamma_{t-k}(-k)^\top$ e pelo fato da matriz $\Gamma_t(0)$ ser simétrica, a matriz que multiplica os coeficientes de (2.41) é simétrica:

$$\begin{bmatrix} \Gamma_t(1)^\top \\ \Gamma_t(2)^\top \\ \vdots \\ \Gamma_t(p_t)^\top \end{bmatrix} = \begin{bmatrix} \Gamma_{t-1}(0) & \Gamma_{t-1}(1)^\top & \cdots & \Gamma_{t-1}(p_t - 1)^\top \\ \Gamma_{t-1}(1) & \Gamma_{t-2}(0) & \cdots & \Gamma_{t-2}(p_t - 2)^\top \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{t-1}(p_t - 1) & \Gamma_{t-2}(p_t - 2) & \cdots & \Gamma_{t-p_t}(0) \end{bmatrix} \begin{bmatrix} \Phi_{t,1}^\top \\ \Phi_{t,2}^\top \\ \vdots \\ \Phi_{t,p_t}^\top \end{bmatrix} \quad (2.42)$$

Esta consideração reduz o número de valores $\Gamma_t(k)$ que precisam ser estimados. A propriedade de correlação periódica abaixo reduz ainda mais o número de matrizes a serem estimadas:

$$\Gamma_t(k) = \text{Cov}[\mathbf{a}_t, \mathbf{a}_{t-k}] = \text{Cov}[\mathbf{a}_{t+s}, \mathbf{a}_{t+s-k}] = \Gamma_{t+s}(k).$$

Portanto, as matrizes $\hat{\zeta}_t, \hat{\Phi}_{t,1}, \dots, \hat{\Phi}_{t,p_t}$ e $\hat{\Sigma}_t$ são definidas como o estimador de *Yule-Walker* para os coeficientes do PVAR se satisfazem as seguintes equações:

$$\begin{bmatrix} \hat{\Gamma}_t(1)^\top \\ \hat{\Gamma}_t(2)^\top \\ \vdots \\ \hat{\Gamma}_t(p_t)^\top \end{bmatrix} = \begin{bmatrix} \hat{\Gamma}_{t-1}(0) & \hat{\Gamma}_{t-1}(1)^\top & \cdots & \hat{\Gamma}_{t-1}(p_t-1)^\top \\ \hat{\Gamma}_{t-1}(1) & \hat{\Gamma}_{t-2}(0) & \cdots & \hat{\Gamma}_{t-2}(p_t-2)^\top \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\Gamma}_{t-1}(p_t-1) & \hat{\Gamma}_{t-2}(p_t-2) & \cdots & \hat{\Gamma}_{t-p_t}(0) \end{bmatrix} \begin{bmatrix} \hat{\Phi}_{t,1}^\top \\ \hat{\Phi}_{t,2}^\top \\ \vdots \\ \hat{\Phi}_{t,p_t}^\top \end{bmatrix}, \quad (2.43)$$

$$\hat{\Sigma}_t = \hat{\Gamma}_t(0) - \sum_{\nu=1}^{p_t} \hat{\Phi}_{t,\nu} \hat{\Gamma}_t(\nu), \quad (2.44)$$

$$\hat{\zeta}_t = \hat{\mu}_t - \sum_{\nu=1}^{p_t} \hat{\Phi}_{t,\nu} \hat{\mu}_{t-\nu}, \quad (2.45)$$

onde $\hat{\mu}_t$ é definido como $\frac{1}{q_s+1} \sum_{\nu=0}^{q_s} \mathbf{a}_{s+\nu\mathbb{S}}$ em que s é a estação associada a t e q_s é o maior inteiro positivo tal que $s + q_s\mathbb{S} \leq T$. O estimador de $\hat{\Gamma}_t(k)$ é construído com base em:

$$\Gamma_t(k) = \mathbb{E}[\mathbf{a}_t \mathbf{a}_{t-k}^\top] - \mathbb{E}[\mathbf{a}_t] \mathbb{E}[\mathbf{a}_{t-k}]^\top. \quad (2.46)$$

Motivado por (2.46), define-se o estimador $\hat{\Gamma}_t(k)$ como

$$\hat{\Gamma}_t(k) = \frac{1}{q_{s,u} + 1} \sum_{\nu=0}^{q_{s,u}} \mathbf{a}_{s+\nu\mathbb{S}} \mathbf{a}_{u+\nu\mathbb{S}}^\top - \left(\frac{1}{q_s + 1} \sum_{\nu=0}^{q_s} \mathbf{a}_{s+\nu\mathbb{S}} \right) \left(\frac{1}{q_u + 1} \sum_{\nu=0}^{q_u} \mathbf{a}_{u+\nu\mathbb{S}} \right)^\top, \quad (2.47)$$

onde s e u são as estações associadas a t e $t+k$. Adicionalmente, q_s e q_u são os maiores inteiros positivos tais que $s + q_s\mathbb{S} \leq T$ e $u + q_u\mathbb{S} \leq T$, e $q_{s,u}$ é o mínimo entre q_s e q_u . Pela propriedade de ergodicidade (2.4), o estimador $\hat{\Gamma}_t(k)$ é consistente.

2.3.2 Método dos mínimos quadrados

O método dos mínimos quadrados é uma técnica de estimação de parâmetros que busca o “melhor” ajuste do modelo aos dados. Em particular, o critério que mede a qualidade deste ajuste é a soma das diferenças ao quadrado entre o valor realizado e o previsto em cada amostra. Assim, quanto menor o valor do critério mais bem ajustado é o modelo em relação a esta medida. A seguir, constrói-se o problema de otimização que define o método dos mínimos quadrados e a respectiva técnica de solução.

Uma representação do modelo PVAR (2.6) conveniente para as manipulações algébricas envolvendo termos de diferentes instantes de tempo é aquela que permite

identificar coeficientes de mesma sazonalidade (mês). Neste sentido, considera-se a seguinte formulação do modelo PVAR:

$$\mathbf{a}_t = \zeta_s + \sum_{\nu=1}^{p_s} \Phi_{s,\nu} \mathbf{a}_{t-\nu} + \boldsymbol{\epsilon}_t, \quad (2.48)$$

onde s é a estação associada ao tempo t . A versão indicial da equação (2.48) reforça o caráter multivariado do PVAR em questão:

$$\begin{aligned} \begin{bmatrix} a_{t,1} \\ a_{t,2} \\ \vdots \\ a_{t,n} \end{bmatrix} &= \begin{bmatrix} \zeta_{s,1} \\ \zeta_{s,2} \\ \vdots \\ \zeta_{s,n} \end{bmatrix} + \begin{bmatrix} \phi_{s1,11} & \phi_{s1,12} & \cdots & \phi_{s1,1n} \\ \phi_{s1,21} & \phi_{s1,22} & \cdots & \phi_{s1,2n} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{s1,n1} & \phi_{s1,n2} & \cdots & \phi_{s1,nn} \end{bmatrix} \begin{bmatrix} a_{t-1,1} \\ a_{t-1,2} \\ \vdots \\ a_{t-1,n} \end{bmatrix} + \\ &\dots + \begin{bmatrix} \phi_{sp_t,11} & \phi_{sp_t,12} & \cdots & \phi_{sp_t,1n} \\ \phi_{sp_t,21} & \phi_{sp_t,22} & \cdots & \phi_{sp_t,2n} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{sp_t,n1} & \phi_{sp_t,n2} & \cdots & \phi_{sp_t,nn} \end{bmatrix} \begin{bmatrix} a_{t-p_t,1} \\ a_{t-p_t,2} \\ \vdots \\ a_{t-p_t,n} \end{bmatrix} + \begin{bmatrix} \epsilon_{t,1} \\ \epsilon_{t,2} \\ \vdots \\ \epsilon_{t,n} \end{bmatrix}, \end{aligned} \quad (2.49)$$

onde n é a dimensão de \mathbf{a}_t .

A seguir, deduz-se o método dos mínimos quadrados para os coeficientes $\zeta_s, \Phi_{s,1}, \dots, \Phi_{s,p_s}$ de (2.48). Em geral, uma dedução análoga é descrita na literatura para o modelo PVAR centrado:

$$\mathbf{a}_t - \boldsymbol{\mu}_t = \sum_{\nu=1}^{p_s} \Phi_{s,\nu} (\mathbf{a}_{t-\nu} - \boldsymbol{\mu}_{t-\nu}) + \boldsymbol{\epsilon}_t, \quad (2.50)$$

onde $\zeta_s = \boldsymbol{\mu}_t - \sum_{\nu=1}^{p_s} \Phi_{s,\nu} \boldsymbol{\mu}_{t-\nu}$. Porém, como a intenção é produzir previsões sempre positivas através de restrições nos coeficientes, é mais conveniente descrever o estimador de mínimos quadrados para (2.48). As condições necessárias e suficientes para garantir previsões sempre positivas serão apresentadas na seção seguinte.

Com relação a descrição do estimador de mínimos quadrados, a estrutura matricial facilita a exposição dos desvios entre o valor realizado e o previsto. Adiante,

reescreve-se a equação (2.49) em formato matricial agregado:

$$\begin{bmatrix} a_{t,1} \\ a_{t,2} \\ \vdots \\ a_{t,n} \end{bmatrix} = \begin{bmatrix} \zeta_{s,1} & \phi_{s1,11} & \phi_{s1,12} & \cdots & \phi_{s1,1n} & \cdots & \phi_{sp_t,11} & \phi_{sp_t,12} & \cdots & \phi_{sp_t,1n} \\ \zeta_{s,2} & \phi_{s1,21} & \phi_{s1,22} & \cdots & \phi_{s1,2n} & \cdots & \phi_{sp_t,21} & \phi_{sp_t,22} & \cdots & \phi_{sp_t,2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \zeta_{s,n} & \phi_{s1,n1} & \phi_{s1,n2} & \cdots & \phi_{s1,nn} & \cdots & \phi_{sp_t,n1} & \phi_{sp_t,n2} & \cdots & \phi_{sp_t,nn} \end{bmatrix} \begin{bmatrix} 1 \\ a_{t-1,1} \\ a_{t-1,2} \\ \vdots \\ a_{t-1,n} \\ \vdots \\ a_{t-p_t,1} \\ a_{t-p_t,2} \\ \vdots \\ a_{t-p_t,n} \end{bmatrix} \quad (2.51)$$

$$+ \begin{bmatrix} \epsilon_{t,1} \\ \epsilon_{t,2} \\ \vdots \\ \epsilon_{t,n} \end{bmatrix}.$$

Por meio de blocos de matrizes, é possível descrever (2.51) de forma mais condensada:

$$\mathbf{a}_t = \begin{bmatrix} \zeta_s & \Phi_{s,1} & \cdots & \Phi_{s,p_s} \end{bmatrix} \begin{bmatrix} 1 \\ \mathbf{a}_{t-1} \\ \vdots \\ \mathbf{a}_{t-p_s} \end{bmatrix} + \boldsymbol{\epsilon}_t, \quad (2.52)$$

onde $t = s, s + \mathbb{S}, \dots, s + q_s \mathbb{S}$ e q_s é o maior inteiro tal que $s + q_s \mathbb{S} \leq T$. Usando esta notação, as amostras que estão ligadas à estação s podem ser agrupadas:

$$\begin{bmatrix} \mathbf{a}_s & \mathbf{a}_{s+\mathbb{S}} & \cdots & \mathbf{a}_{s+q_s \mathbb{S}} \end{bmatrix} = \begin{bmatrix} \zeta_s & \Phi_{s,1} & \cdots & \Phi_{s,p_s} \end{bmatrix} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \mathbf{a}_{s-1} & \mathbf{a}_{s+\mathbb{S}-1} & \cdots & \mathbf{a}_{s+q_s \mathbb{S}-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{a}_{s-p_s} & \mathbf{a}_{s+\mathbb{S}-p_s} & \cdots & \mathbf{a}_{s+q_s \mathbb{S}-p_s} \end{bmatrix} + \quad (2.53) \\
+ \begin{bmatrix} \boldsymbol{\epsilon}_s & \boldsymbol{\epsilon}_{s+\mathbb{S}} & \cdots & \boldsymbol{\epsilon}_{s+q_s \mathbb{S}} \end{bmatrix}.$$

A representação de (2.53) pode ser escrita de forma sucinta por meio da seguinte

notação:

$$\begin{aligned}
A_s &:= \begin{bmatrix} \mathbf{a}_s & \mathbf{a}_{s+\mathbb{S}} & \cdots & \mathbf{a}_{s+q_s\mathbb{S}} \end{bmatrix}, \\
\Phi_s &:= \begin{bmatrix} \zeta_s & \Phi_{s,1} & \cdots & \Phi_{s,p_s} \end{bmatrix}, \\
\mathbb{A}_s &:= \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \mathbf{a}_{s-1} & \mathbf{a}_{s+\mathbb{S}-1} & \cdots & \mathbf{a}_{s+q_s\mathbb{S}-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{a}_{s-p_s} & \mathbf{a}_{s+\mathbb{S}-p_s} & \cdots & \mathbf{a}_{s+q_s\mathbb{S}-p_s} \end{bmatrix}, \\
E_s &:= \begin{bmatrix} \boldsymbol{\epsilon}_s & \boldsymbol{\epsilon}_{s+\mathbb{S}} & \cdots & \boldsymbol{\epsilon}_{s+q_s\mathbb{S}} \end{bmatrix},
\end{aligned}$$

onde esses parâmetros são substituídos em (2.53):

$$A_s = \Phi_s \mathbb{A}_s + E_s, \quad s = 1, \dots, \mathbb{S}. \quad (2.54)$$

Pelo formato (2.54), a matriz $A_s - \Phi_s \mathbb{A}_s$ contém todos os desvios entre o valor realizado e o previsto associados à estação s . Neste sentido, a soma de cada entrada de $A_s - \Phi_s \mathbb{A}_s$ ao quadrado representa a parcela do critério relacionada com s . O critério do método dos mínimos quadrados é dado por

$$\sum_{s=1}^{\mathbb{S}} \|A_s - \Phi_s \mathbb{A}_s\|^2,$$

onde $\|M\| := \sqrt{\sum_{i=1}^m \sum_{j=1}^n m_{i,j}^2}$ é chamada de norma de *Frobenius* para a matriz $M \in \mathbb{R}^{m \times n}$.

Os valores de $\hat{\zeta}_s, \hat{\Phi}_{s,1}, \dots, \hat{\Phi}_{s,p_s}$ para s desde 1 até \mathbb{S} são definidos como o estimador de *mínimos quadrados* para os coeficientes do PVAR se $\hat{\Phi}_s := [\hat{\zeta}_s \ \hat{\Phi}_{s,1} \ \cdots \ \hat{\Phi}_{s,p_s}]$ com s de 0 a $\mathbb{S} - 1$ é solução ótima do seguinte problema de minimização:

$$\min_{\Phi_1, \dots, \Phi_{\mathbb{S}}} \sum_{s=1}^{\mathbb{S}} \|A_s - \Phi_s \mathbb{A}_s\|^2. \quad (2.55)$$

O cálculo do estimador de mínimos quadrados pode ser feito separadamente para cada estação s , pois a função objetivo e restrições do problema (2.55) são não acopladas em $\Phi_0, \dots, \Phi_{\mathbb{S}-1}$, ou seja,

$$\min_{\Phi_1, \dots, \Phi_{\mathbb{S}}} \sum_{s=1}^{\mathbb{S}} \|A_s - \Phi_s \mathbb{A}_s\|^2 = \sum_{s=1}^{\mathbb{S}} \min_{\Phi_s} \|A_s - \Phi_s \mathbb{A}_s\|^2.$$

Além disso, é possível decompor cada problema

$$\min_{\Phi_s} \|A_s - \Phi_s \mathbb{A}_s\|^2 \quad (2.56)$$

em subproblemas mais simples representando as matrizes A_s e Φ_s por meio dos seus respectivos vetores linha. Sejam $\mathbf{a}_{s,k}^\top = [a_{s,k} \ a_{s+\mathbb{S},k} \ \dots \ a_{s+q_s\mathbb{S},k}]$ e

$$\varphi_{s,k}^\top = [\zeta_{s,k} \ \phi_{s1,k1} \ \phi_{s1,k2} \ \dots \ \phi_{s1,kn} \ \dots \ \phi_{sp_t,k1} \ \phi_{sp_t,k2} \ \dots \ \phi_{p_t,kn}]$$

as k -ésimas linhas das matrizes A_s e Φ_s , respectivamente. Então, vale que:

$$A_s = \begin{bmatrix} \text{---} & \mathbf{a}_{s,1}^\top & \text{---} \\ \text{---} & \mathbf{a}_{s,2}^\top & \text{---} \\ & \vdots & \\ \text{---} & \mathbf{a}_{s,n}^\top & \text{---} \end{bmatrix}, \quad \Phi_s = \begin{bmatrix} \text{---} & \varphi_{s,1}^\top & \text{---} \\ \text{---} & \varphi_{s,2}^\top & \text{---} \\ & \vdots & \\ \text{---} & \varphi_{s,n}^\top & \text{---} \end{bmatrix},$$

$$A_s - \Phi_s \mathbb{A}_s = \begin{bmatrix} \text{---} & (\mathbf{a}_{s,1}^\top - \varphi_{s,1}^\top \mathbb{A}_s) & \text{---} \\ \text{---} & (\mathbf{a}_{s,2}^\top - \varphi_{s,2}^\top \mathbb{A}_s) & \text{---} \\ & \vdots & \\ \text{---} & (\mathbf{a}_{s,n}^\top - \varphi_{s,n}^\top \mathbb{A}_s) & \text{---} \end{bmatrix}. \quad (2.57)$$

Com isso, pode-se escrever (2.56) da seguinte forma:

$$\min_{\Phi_s} \|A_s - \Phi_s \mathbb{A}_s\|^2 = \min_{\varphi_{s,1}, \dots, \varphi_{s,n}} \sum_{k=1}^n \|\mathbf{a}_{s,k}^\top - \varphi_{s,k}^\top \mathbb{A}_s\|^2 \quad (2.58)$$

$$= \min_{\varphi_{s,1}, \dots, \varphi_{s,n}} \sum_{k=1}^n \|\mathbf{a}_{s,k} - \mathbb{A}_s^\top \varphi_{s,k}\|^2 \quad (2.59)$$

$$= \sum_{k=1}^n \min_{\varphi_{s,k}} \|\mathbf{a}_{s,k} - \mathbb{A}_s^\top \varphi_{s,k}\|^2, \quad (2.60)$$

onde (2.60) evidencia que o problema (2.56) é não acoplado em relação a $\varphi_{s,1}, \dots, \varphi_{s,n}$. A justificativa para a identidade (2.58) é o fato da norma ao quadrado de uma matriz ser igual a soma das normas ao quadrado das respectivas linhas e a justificativa para a identidade (2.59) é que a norma é invariante por transposição de matrizes, lembrando que a norma em questão é a de Frobenius.

Por último, omitindo os subíndices s e k , o problema

$$\min_{\varphi} \|\mathbf{a} - \mathbb{A}^\top \varphi\|^2 \quad (2.61)$$

é um exemplo clássico de otimização irrestrita, convexa e diferenciável em que uma

das técnicas de solução se baseia em encontrar um vetor φ^* para o qual a derivada da função objetivo de (2.61) é igual a zero:

$$\nabla_{\varphi}(\|\mathbf{a} - \mathbb{A}^{\top}\varphi\|^2)\Big|_{\varphi=\varphi^*} = (-2\mathbb{A}^{\top})^{\top}(\mathbf{a} - \mathbb{A}^{\top}\varphi^*) = 0,$$

ou seja, o vetor $\varphi^* = (\mathbb{A}\mathbb{A}^{\top})^{-1}\mathbb{A}\mathbf{a}$ é a solução de (2.61).

2.3.3 Método dos mínimos quadrados com restrições nos coeficientes

Nesta seção, apresenta-se uma abordagem capaz de garantir que a previsão seja sempre positiva por meio de restrições nos coeficientes. Supondo que todos os coeficientes de $\zeta_t, \Phi_{t,1}, \dots, \Phi_{t,p_t}$ sejam positivos então a previsão

$$\text{Prev}[\mathbf{a}_t \mid \mathbf{a}_{[t-1]}] = \zeta_t + \sum_{\nu=1}^{p_t} \Phi_{t,\nu} \mathbf{a}_{t-\nu} \quad (2.62)$$

é sempre positiva, pois supõe-se que o vetor \mathbf{a}_t é sempre. Admitindo que *qualquer* valor positivo de afluência possa ocorrer então os coeficientes serem positivos é uma condição necessária e suficiente para produzir previsões sempre positivas. Com efeito, se ζ_t possui alguma componente negativa então a previsão para o caso em que as componentes de $\mathbf{a}_{[t-1]}$ são todas nulas, $\text{Prev}[\mathbf{a}_t \mid \mathbf{0}] = \zeta_t$, é negativa em alguma componente. Se $\Phi_{t,\nu}$ possui coeficiente $\phi_{t\nu,ij}$ negativo então a afluência $\mathbf{a}_{t-\nu}$ definida por

$$[\mathbf{a}_{t-\nu}]_k = \begin{cases} r, & \text{se } k = j \\ 0, & \text{se } k \neq j \end{cases}$$

gera um vetor resultante $\Phi_{t,\nu} \mathbf{a}_{t-\nu}$ igual a $r[\phi_{t\nu,1j} \ \phi_{t\nu,2j} \ \dots \ \phi_{t\nu,nj}]^{\top}$, ou seja, é a j -ésima coluna de $\Phi_{t,\nu}$ multiplicada por $r > 0$ cuja a i -ésima componente é negativa. Aumentando o valor de r é possível fazer com que a i -ésima coordenada de $\Phi_{t,\nu} \mathbf{a}_{t-\nu}$ fique arbitrariamente negativa. Mantendo os valores de afluência fixos e aumentando suficientemente r é possível criar uma previsão (2.62) negativa. Portanto, os coeficientes das matrizes $\zeta_t, \Phi_{t,1}, \dots, \Phi_{t,p_t}$ precisam ser todos positivos para se garantir que as previsões sejam sempre positivas.

Um método simples que considera esta informação é o método dos mínimos quadrados com restrições nos coeficientes. As matrizes $\zeta_s, \hat{\Phi}_{s,1}, \dots, \hat{\Phi}_{s,p_s}$ para s desde 1 até \mathbb{S} são definidas como o estimador de *mínimos quadrados com restrições nos coeficientes* para o modelo PVAR se o conjunto de coeficientes gerado $\hat{\Phi}_i := [\zeta_s \ \hat{\Phi}_{s,1} \ \dots \ \hat{\Phi}_{s,p_s}]$ com s de 1 a \mathbb{S} é solução ótima do seguinte problema de

minimização:

$$\min_{\Phi_1 \geq 0, \dots, \Phi_S \geq 0} \sum_{s=1}^S \|A_s - \Phi_s \Lambda_s\|^2. \quad (2.63)$$

A técnica de solução que divide (2.63) em problemas menores segue de forma análoga ao caso dos mínimos quadrados originais (2.55). Em cada etapa da argumentação basta incluir a restrição de positividade dos coeficientes.

2.4 Determinação dos resíduos

O erro (ou ruído) do modelo PVAR é definido como a diferença entre o valor observado e o previsto:

$$\boldsymbol{\epsilon}_t = \mathbf{a}_t - \left(\zeta_t + \sum_{\nu=1}^{p_t} \Phi_{t,\nu} \mathbf{a}_{t-\nu} \right). \quad (2.64)$$

Como os valores “*verdadeiros*” dos parâmetros (2.64) não são conhecidos a priori o ruído $\boldsymbol{\epsilon}_t$ é não observável. O resíduo, por outro lado, é uma estimativa observável do erro aditivo $\boldsymbol{\epsilon}_t$ e é definido como a diferença entre o valor observado e o previsto *estimado*:

$$\hat{\boldsymbol{\epsilon}}_t = \mathbf{a}_t - \left(\hat{\zeta}_t + \sum_{\nu=1}^{p_t} \hat{\Phi}_{t,\nu} \mathbf{a}_{t-\nu} \right). \quad (2.65)$$

Vale ressaltar que os resíduos $\hat{\boldsymbol{\epsilon}}_t$ para t de 1 até T são em geral dependentes, enquanto que os erros $\boldsymbol{\epsilon}_t$ são, por hipótese, independentes.

Um dos principais usos dos resíduos é na estimação de quantidades que dependem dos ruídos $\boldsymbol{\epsilon}_t$. Na próxima seção, descrevem-se alguns modelos de distribuição do ruído $\boldsymbol{\epsilon}_t$ e como estimá-los.

2.5 Estimação da distribuição dos ruídos

Uma das hipóteses do modelo PVAR é a distribuição do erro aditivo $\boldsymbol{\epsilon}_t$. Nesta seção, serão discutidas duas opções de modelagem para a densidade de probabilidade do erro:

- 1) $\boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \Sigma_t)$, ou seja, o erro $\boldsymbol{\epsilon}_t$ segue uma distribuição multivariada normal com média 0 e matriz de covariância Σ_t . Segue da propriedade $\Sigma_t = \Sigma_{t+S}$ que $\boldsymbol{\epsilon}_t$ têm (módulo \mathbb{S}) a mesma distribuição;

2) $\boldsymbol{\epsilon}_t \sim \ln \mathcal{N}(\boldsymbol{\Delta}_t, \mathbf{u}_t, W_t) = \boldsymbol{\Delta}_t + \ln \mathcal{N}(\mathbf{u}_t, W_t)$, ou seja, o erro $\boldsymbol{\epsilon}_t$ segue uma distribuição multivariada lognormal com parâmetros \mathbf{u}_t e W_t , deslocada por um vetor $\boldsymbol{\Delta}_t$. Segue da propriedade $(\boldsymbol{\Delta}_t, \mathbf{u}_t, W_t) = (\boldsymbol{\Delta}_{t+\mathbb{S}}, \mathbf{u}_{t+\mathbb{S}}, W_{t+\mathbb{S}})$ que $\boldsymbol{\epsilon}_t$ têm (módulo \mathbb{S}) a mesma distribuição;

A primeira opção de modelagem, $\boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \Sigma_t)$, admite que o erro aditivo possa assumir qualquer valor real. Com isso, sempre existe a possibilidade de uma sequência de erros negativos resultar em um valor negativo para \mathbf{a}_t . Nas simulações que consideram esta distribuição do erro, valores negativos são observados.

Ainda neste mesmo caso, destacam-se dois estimadores para a matriz de covariância Σ_t . Um estimador é construído diretamente a partir do método de Yule-Walker:

$$\hat{\Sigma}_t = \hat{\Gamma}_t(0) - \sum_{\nu=1}^{p_t} \hat{\Phi}_{t,\nu} \hat{\Gamma}_t(\nu),$$

e que é adequado se aplicado para a estimação dos coeficientes $\zeta_t, \Phi_{t,1}, \dots, \Phi_{t,p_t}$. O outro é construído a partir dos resíduos:

$$\hat{\Sigma}_t = \frac{1}{q_s + 1} \sum_{\nu=0}^{q_s} \hat{\boldsymbol{\epsilon}}_{s+\nu\mathbb{S}} \hat{\boldsymbol{\epsilon}}_{s+\nu\mathbb{S}}^\top, \quad (2.66)$$

onde s é a estação associada ao tempo t . Além disso, q_s é o maior inteiro positivo tal que $s + q_s\mathbb{S} \leq T$. O estimador (2.66) pode ser usado em conjunto com o método de mínimos quadrados.

A segunda opção de modelagem, $\boldsymbol{\epsilon}_t \sim \boldsymbol{\Delta}_t + \ln \mathcal{N}(\mathbf{u}_t, W_t)$, surgiu como uma forma de produzir valores sempre positivos de \mathbf{a}_t . A proposta de CEPTEL [9] utiliza uma regra para a obtenção de $\boldsymbol{\Delta}_t$ que depende dos valores passados $\mathbf{a}_{[t-1]}$. Entretanto, esta característica viola a hipótese de independência do erro aditivo $\boldsymbol{\epsilon}_t$, que é um requisito para o algoritmo SDDP.

Uma outra sugestão para contornar o efeito de gerar valores negativos para as aflúncias é elaborada a partir da abordagem de previsões positivas. Como visto na seção de mínimos quadrados com restrição nos coeficientes, é possível garantir previsões sempre positivas se todos os coeficientes $\zeta_t, \Phi_{t,1}, \dots, \Phi_{t,p_t}$ forem positivos. Admitindo que qualquer sequência positiva de aflúncias possa ocorrer, o menor valor de previsão é ζ_t :

$$\text{Prev}[\mathbf{a}_t \mid \mathbf{a}_{[t-1]}] = \zeta_t + \sum_{\nu=1}^{p_t} \Phi_{t,\nu} \mathbf{a}_{t-\nu} \geq \zeta_t.$$

Supondo que o erro segue uma log-normal transladada por $\boldsymbol{\Delta}_t$, $\boldsymbol{\epsilon}_t \sim \boldsymbol{\Delta}_t + \ln \mathcal{N}(\mathbf{u}_t, W_t)$, então o menor valor que o erro $\boldsymbol{\epsilon}_t$ pode assumir é $\boldsymbol{\Delta}_t$. Consequen-

temente, é possível obter uma relação entre o menor valor de previsão ζ_t e o menor valor de erro Δ_t de forma a garantir que a afluência \mathbf{a}_t seja sempre positiva:

$$\mathbf{a}_t = \text{Prev}[\mathbf{a}_t \mid \mathbf{a}_{[t-1]}] + \boldsymbol{\epsilon}_t \geq \zeta_t + \Delta_t \geq 0,$$

isto é, a relação é $\Delta_t \geq -\zeta_t$. A possibilidade da afluência \mathbf{a}_t ser zero equivale a Δ_t ser igual a $-\zeta_t$, pois o menor valor que este modelo pode gerar para a afluência é $\zeta_t + \Delta_t$. Portanto, assume-se que Δ_t é igual a $-\zeta_t$. Além disso, os parâmetros \mathbf{u}_t e W_t são escolhidos de forma a produzir um erro $\boldsymbol{\epsilon}_t$ com média zero e covariância igual a Σ_t . Com efeito, se $\boldsymbol{\epsilon}_t$ segue uma distribuição multivariada log-normal a três parâmetros então:

$$\mathbb{E}[\boldsymbol{\epsilon}_t]_i = \Delta_{t,i} + e^{u_{t,i} + W_{t,ii}/2}, \quad i = 1, \dots, n. \quad (2.67)$$

$$\text{Cov}[\boldsymbol{\epsilon}_t]_{ij} = e^{u_{t,i} + u_{t,j} + (W_{t,ii} + W_{t,jj})/2} (e^{W_{t,ij}} - 1), \quad i, j = 1, \dots, n. \quad (2.68)$$

Por meio de manipulações algébricas é possível isolar \mathbf{u}_t e W_t :

$$u_{t,i} = \log \left(\frac{\mathbb{E}[\boldsymbol{\epsilon}_t]_i - \Delta_{t,i}}{\sqrt{R_{t,ii}}} \right),$$

$$W_{t,ij} = \log(R_{t,ij}),$$

$$R_{t,ij} = 1 + \frac{\text{Cov}[\boldsymbol{\epsilon}_t]_{ij}}{(\mathbb{E}[\boldsymbol{\epsilon}_t]_i - \Delta_{t,i})(\mathbb{E}[\boldsymbol{\epsilon}_t]_j - \Delta_{t,j})},$$

onde $i, j = 1, \dots, n$. Como a média do erro é zero, a covariância é Σ_t e o deslocamento é igual a $-\zeta_t$ então

$$u_{t,i} = \log \left(\frac{\zeta_{t,i}}{\sqrt{R_{t,ii}}} \right),$$

$$W_{t,ij} = \log(R_{t,ij}), \quad (2.69)$$

$$R_{t,ij} = 1 + \frac{\Sigma_{t,ij}}{\zeta_{t,i}\zeta_{t,j}}.$$

As limitações deste método são

1. os valores de $R_{t,ij}$ devem ser todos positivos;
2. a matriz formada por $W_{t,ij} = \log(R_{t,ij})$ deve ser semi-definida positiva; e
3. todas as componentes do vetor ζ_t devem ser estritamente positivas.

Apesar dessa abordagem garantir valores positivos de afluência, na prática essas hipóteses são muito restritivas para serem aplicadas diretamente. Em geral, o sistema (2.69) com os requisitos apresentados é *sobredeterminado*, ou seja, não possui

solução. DE OLIVEIRA *et al.* [15] sugerem uma maneira de flexibilizar as condições (2.69) para a determinação dos parâmetros da log-normal a três parâmetros.

2.6 Identificação do modelo

Nas seções anteriores foram expostos métodos para se estimar os coeficientes do modelo PVAR e a distribuição dos ruídos. Porém, nenhuma maneira de se identificar a ordem do modelo em cada estação foi mencionada. Nesta seção, descreve-se dois métodos usuais para a seleção de ordem: o critério de autocorrelação parcial periódica e o critério BIC.

2.6.1 Critério da autocorrelação parcial periódica

O critério da autocorrelação parcial periódica é um método de seleção de modelos aplicável ao caso *univariado*. Apresenta-se a seguir o critério de autocorrelação parcial periódica e o seu emprego na estimação do modelo PAR.

A autocorrelação *parcial* periódica de lag k e estação s para uma série temporal estacionária *univariada* $\{a_1, a_2, \dots\}$ é definida como o k -ésimo coeficiente $\phi_{s,k}$ do ajuste de um PAR centrado e reduzido com ordem k para a estação s :

$$\frac{a_t - \mu_t}{\sigma_t} = \sum_{\nu=1}^k \phi_{s,\nu} \left(\frac{a_{t-\nu} - \mu_{t-\nu}}{\sigma_{t-\nu}} \right) + \epsilon_t, \quad (2.70)$$

onde $t = s, s + \mathbb{S}, \dots, s + q_s \mathbb{S}$, $\mu_t = \mathbb{E}[a_t]$, $\sigma_t = \text{Var}[a_t]^{1/2}$ e ϵ_t é um erro independente, identicamente distribuído módulo \mathbb{S} com média 0 e variância 1. Por uma questão de notação, é conveniente acrescentar a informação da ordem k do PAR como subíndice dos coeficientes de (2.70), ou seja, $\phi_{s,\nu} = \phi_{s,\nu k}$.

Um modo de calcular o valor de $\phi_{s,kk}$ é via sistema de Yule-Walker. Multiplicando por $(a_{t-j} - \mu_{t-j})/\sigma_{t-j}$ em ambos os lados de (2.70) e tomando a esperança se obtém a seguinte identidade:

$$\begin{aligned} \mathbb{E} \left[\left(\frac{a_t - \mu_t}{\sigma_t} \right) \left(\frac{a_{t-j} - \mu_{t-j}}{\sigma_{t-j}} \right) \right] &= \sum_{\nu=1}^k \phi_{s,\nu k} \mathbb{E} \left[\left(\frac{a_{t-\nu} - \mu_{t-\nu}}{\sigma_{t-\nu}} \right) \left(\frac{a_{t-j} - \mu_{t-j}}{\sigma_{t-j}} \right) \right] \\ &\quad + \mathbb{E} \left[\epsilon_t \left(\frac{a_{t-j} - \mu_{t-j}}{\sigma_{t-j}} \right) \right]. \end{aligned} \quad (2.71)$$

Na equação (2.71), cada esperança envolvendo apenas os termos a_t se refere à autocorrelação periódica que é definida por:

$$\rho_t(j) = \frac{\text{Cov}[a_t, a_{t-j}]}{\text{Var}[a_t]^{1/2} \text{Var}[a_{t-j}]^{1/2}} = \frac{\mathbb{E}[(a_t - \mu_t)(a_{t-j} - \mu_{t-j})]}{\sigma_t \sigma_{t-j}}. \quad (2.72)$$

Supondo a condição de causalidade para o modelo PAR, tem-se que o erro ϵ_t é independente de a_{t-j} para todo j maior do que ou igual a 1. Consequentemente, a equação (2.71) pode ser escrita como:

$$\rho_t(j) = \sum_{\nu=1}^k \phi_{s,\nu k} \rho_{t-j}(\nu - j), \quad \forall j \geq 1. \quad (2.73)$$

Variando j entre 1 e k é possível montar o seguinte sistema linear envolvendo os coeficientes do modelo PAR:

$$\begin{bmatrix} \rho_t(1) \\ \rho_t(2) \\ \vdots \\ \rho_t(k) \end{bmatrix} = \begin{bmatrix} \rho_{t-1}(0) & \rho_{t-1}(1) & \cdots & \rho_{t-1}(k-1) \\ \rho_{t-2}(-1) & \rho_{t-2}(0) & \cdots & \rho_{t-2}(k-2) \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{t-k}(1-k) & \rho_{t-k}(2-k) & \cdots & \rho_{t-k}(0) \end{bmatrix} \begin{bmatrix} \phi_{t,1k} \\ \phi_{t,2k} \\ \vdots \\ \phi_{t,kk} \end{bmatrix}. \quad (2.74)$$

Pela propriedade $\rho_t(j) = \rho_{t-j}(-j)$, tem-se que

$$\begin{bmatrix} \rho_t(1) \\ \rho_t(2) \\ \vdots \\ \rho_t(k) \end{bmatrix} = \begin{bmatrix} \rho_{t-1}(0) & \rho_{t-1}(1) & \cdots & \rho_{t-1}(k-1) \\ \rho_{t-1}(1) & \rho_{t-2}(0) & \cdots & \rho_{t-2}(k-2) \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{t-1}(k-1) & \rho_{t-2}(k-2) & \cdots & \rho_{t-k}(0) \end{bmatrix} \begin{bmatrix} \phi_{t,1k} \\ \phi_{t,2k} \\ \vdots \\ \phi_{t,kk} \end{bmatrix}, \quad (2.75)$$

ou seja, a matriz que multiplica os coeficientes de (2.74) é simétrica, o que reduz o número de valores $\rho_t(j)$ que precisam ser estimados. O sistema de equações dado por (2.75) é chamado de sistema de Yule-Walker. Define-se a autocorrelação parcial de lag k para a estação associada a t , $\phi_{t,kk}$, como a última componente da solução do sistema de Yule-Walker (2.75).

Na prática, para estimar o valor da autocorrelação *parcial* $\phi_{t,kk}$ é preciso estimar os valores da autocorrelação $\rho_t(j)$. Abaixo, sugere-se uma forma de estimar $\rho_t(j)$:

$$\hat{\rho}_t(j) = \frac{\widehat{\text{Cov}}[a_t, a_{t-j}]}{\hat{\sigma}_t \hat{\sigma}_{t-j}}, \quad \hat{\sigma}_t = \widehat{\text{Cov}}[a_t, a_t]^{\frac{1}{2}}, \quad \hat{\mu}_t = \frac{1}{q_s + 1} \sum_{\nu=0}^{q_s} a_{s+\nu\mathbb{S}},$$

$$\widehat{\text{Cov}}[a_t, a_{t-j}] = \frac{1}{q_{s,u} + 1} \sum_{\nu=0}^{q_{s,u}+1} (a_{s+\nu\mathbb{S}} - \hat{\mu}_t)(a_{u+\nu\mathbb{S}} - \hat{\mu}_{t-j}),$$

onde s e u são as estações associadas a t e $t-j$, q_s e q_u são os maiores inteiros positivos tal que $s + q_s\mathbb{S} \leq T$ e $u + q_u\mathbb{S} \leq T$, e $q_{s,u}$ é o mínimo entre q_s e q_u . Pela propriedade de ergodicidade (2.4), esses estimadores são consistentes. Com isso, define-se o *estimador* da autocorrelação parcial de lag k para a estação associada ao tempo t , $\hat{\phi}_{t,kk}$, como a última componente da solução do sistema de Yule-Walker

(2.76):

$$\begin{bmatrix} \hat{\rho}_t(1) \\ \hat{\rho}_t(2) \\ \vdots \\ \hat{\rho}_t(k) \end{bmatrix} = \begin{bmatrix} \hat{\rho}_{t-1}(0) & \hat{\rho}_{t-1}(1) & \cdots & \hat{\rho}_{t-1}(k-1) \\ \hat{\rho}_{t-1}(1) & \hat{\rho}_{t-2}(0) & \cdots & \hat{\rho}_{t-2}(k-2) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\rho}_{t-1}(k-1) & \hat{\rho}_{t-2}(k-2) & \cdots & \hat{\rho}_{t-k}(0) \end{bmatrix} \begin{bmatrix} \hat{\phi}_{1k,t} \\ \hat{\phi}_{2k,t} \\ \vdots \\ \hat{\phi}_{kk,t} \end{bmatrix} \quad (2.76)$$

Uma consequência da autocorrelação parcial de uma série temporal que segue um modelo PAR é que se a ordem k do PAR em uma dada estação s é menor do que um lag m , então o valor da correspondente autocorrelação parcial $\phi_{s,mm}$ é zero. Com essa informação, tem-se a motivação para usar a autocorrelação parcial para identificar a ordem de um modelo PAR *univariado*. Adicionalmente, o respectivo estimador $\hat{\phi}_{s,mm}$ dividido pela raiz quadrada do número de amostras associadas a estação s converge para uma normal de média 0 e variância 1 como enunciado pelo teorema abaixo:

Teorema. *Se $\{a_1, a_2, \dots\}$ é um PAR(\mathbf{p}) causal de ordem p_s para a estação s então para todo m maior do que p_s vale que*

$$\frac{1}{\sqrt{q_s + 1}} \hat{\phi}_{s,mm} \xrightarrow{d} \mathcal{N}(0, 1),$$

onde q_s é o maior inteiro positivo tal que $s + q_s \mathbb{S} \leq T$.

Essa propriedade do estimador de autocorrelação parcial permite construir o seguinte teste de hipótese para a identificação da ordem do modelo PAR:

H_0 : O modelo é um PAR de ordem p_s para a estação s .

H_1 : O modelo não é um PAR de ordem p_s para a estação s .

O teste estatístico consiste em rejeitar H_0 se $|\hat{\phi}_{s,mm}/\sqrt{q_s + 1}| > z_{\alpha/2}$, onde α é o tamanho do teste e z_α é o α -quantil superior da normal padrão, ou seja, $z_\alpha = F^{-1}(1 - \alpha)$ tal que $F(\cdot)$ é a função distribuição acumulada da normal com média 0 e variância 1. A justificativa desse procedimento é dada pelo teorema a seguir:

Teorema. *Assintoticamente o teste acima possui tamanho α , ou seja,*

$$\mathbb{P} \left(\frac{1}{\sqrt{q_s + 1}} |\hat{\phi}_{s,mm}| > z_{\alpha/2} \mid H_0 \text{ é válida} \right) \xrightarrow{q_s \rightarrow \infty} \alpha$$

Demonstração. Sob a hipótese nula H_0 , $\hat{\phi}_{s,mm}/\sqrt{q_s + 1} \xrightarrow{d} \mathcal{N}(0, 1)$. Portanto, a probabilidade de se rejeitar a hipótese nula dado que esta é verdadeira converge

para a probabilidade do módulo da normal padrão ser maior do que $z_{\alpha/2}$:

$$\mathbb{P}\left(\frac{1}{\sqrt{q_s + 1}} \left| \hat{\phi}_{s,mm} \right| > z_{\alpha/2} \mid H_0 \text{ é válida}\right) \xrightarrow{q_s \rightarrow \infty} \mathbb{P}(|W| > z_{\alpha/2}),$$

onde $W \sim \mathcal{N}(0, 1)$. Observa-se também que

$$\begin{aligned} \mathbb{P}(|W| > z_{\alpha/2}) &= \mathbb{P}(W < -z_{\alpha/2}) + \mathbb{P}(W > z_{\alpha/2}) \\ &= \mathbb{P}(W \leq -z_{\alpha/2}) + 1 - \mathbb{P}(W \leq z_{\alpha/2}) \\ &= F(-z_{\alpha/2}) + 1 - F(z_{\alpha/2}) \\ &= 1 - F(z_{\alpha/2}) + 1 - F(z_{\alpha/2}) \\ &= 2 - 2F(z_{\alpha/2}) \\ &= 2 - 2F(F^{-1}(1 - \alpha/2)) \\ &= \alpha \end{aligned}$$

□

A ordem p_s do modelo PAR(\mathbf{p}) *univariado* para a estação s , pode ser identificada aplicando o teste acima para cada ordem candidata $p = 1, 2, \dots, p_{\max}$ e admitindo a que não for rejeitada pelo teste estatístico. Costuma-se utilizar $\alpha = 0.05$ para realizar esse teste de hipótese, isto é, $z_{\alpha/2} \approx 1.96$. Neste sentido, existem dois procedimentos evidentes:

1. percorrer as ordens candidatas de maneira *crescente* e sugerir como ordem do modelo a anterior a primeira não rejeição (“aceitação”) da hipótese nula;
2. percorrer as ordens candidatas de maneira *decrecente* e sugerir como ordem do modelo a primeira que rejeita a hipótese nula.

Essas opções podem gerar resultados diferentes.

2.6.2 Critério de Informação Bayesiano (BIC)

O Critério de Informação Bayesiano (BIC) é um critério para a escolha de modelos baseado em um índice que quantifica a qualidade dos respectivos ajustes. Por definição, o índice associado ao BIC considera dois termos: o logaritmo da verossimilhança e uma penalidade no número de parâmetros pressupostos pelo modelo.

A função de verossimilhança é definida como a probabilidade conjunta dos dados observados $\{\mathbf{a}_1, \mathbf{a}_2, \dots\}$, fixado o modelo estatístico em questão. Intuitivamente, a verossimilhança mede o quão provável seria obter as mesmas observações se o modelo sugerido fosse o “verdadeiro”. Em geral, quanto maior o número de parâmetros maior

é o valor da verossimilhança, porém modelos com muitos parâmetros podem produzir um sobreajuste. O sobreajuste ocorre quando o modelo considera informações puramente aleatórias como parte de sua componente determinística. Modelos com esta característica costumam ter uma péssima capacidade preditiva. Isso esclarece a presença de um termo proporcional ao número de parâmetros para contrabalançar o valor do logaritmo da verossimilhança no índice associado ao BIC. A ideia é que este índice classifique como bom o modelo que possua um equilíbrio entre a explicação dos dados observados e o número de parâmetros e como ruim os demais modelos.

Nesta seção será abordado o critério BIC para modelos lineares univariados. Denota-se por M um dado modelo, por φ o vetor de parâmetros associados, por $a_{[T]}$ as respectivas observações até o tempo T e por $f(a_{[T]} | M, \hat{\varphi})$ a função de verossimilhança. Com base nesta notação, o índice BIC pode ser escrito como:

$$\text{BIC} := -2 \log f(a_{[T]} | M, \hat{\varphi}) + n(\hat{\varphi}) \log T, \quad (2.77)$$

onde $\hat{\varphi}$ é um estimador dos parâmetros, $n(\hat{\varphi})$ é o número de parâmetros e T é o total de observações. Por uma convenção de sinal, quanto menor o valor do BIC (2.77) melhor é o modelo em relação a este critério. O uso do critério BIC será exemplificado considerando um modelo autorregressivo de ordem p ,

$$a_t = \varphi_0 + \sum_{\nu=1}^p \varphi_{t-\nu} a_{t-\nu} + \varepsilon_t, \quad (2.78)$$

cujo erro ε_t segue uma normal de média 0 e desvio padrão σ^2 , ou seja, $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$. Observa-se que o número de parâmetros desse modelo é $p + 2$.

O índice BIC para o modelo AR(p) univariado (2.78) com erro normal é:

$$\text{BIC} = T(\log 2\pi + 1) + T \log(\hat{\sigma}^2) + (p + 2) \log T. \quad (2.79)$$

Supondo que os candidatos sejam apenas modelos autorregressivos de diversas ordens e as amostras sejam fixas e iguais em todos os ajustes, o termo $T(\log 2\pi + 1)$ de (2.79) é uma constante que pode ser dispensada:

$$\text{BIC} = T \log(\hat{\sigma}^2) + (p + 2) \log T. \quad (2.80)$$

Vale ressaltar que se o modelo autorregressivo (2.78) considerasse variáveis exógenas então a expressão (2.80) mudaria apenas o número de coeficientes: de $p + 2$ para $p + b + 2$, onde b é o número de variáveis exógenas.

Uma extensão do critério (2.80) para o caso autorregressivo *multivariado* (VAR)

definido por

$$\mathbf{a}_t = \zeta + \sum_{\nu=1}^p \Phi_{\nu} \mathbf{a}_{t-\nu} + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(0, \Sigma), \quad (2.81)$$

é realizada ao observar que o critério univariado pode ser aplicado em cada componente do modelo. Cada componente do VAR com erro normal multivariado é um AR com variáveis exógenas e erro normal univariado. As variáveis exógenas são as outras componentes do VAR em diferentes lags:

$$a_{t,j} = \zeta_j + \sum_{\nu=1}^p \sum_{k=1}^n \phi_{\nu,jk} a_{t-\nu,k} + \varepsilon_{t,j}, \quad \varepsilon_{t,j} \sim \mathcal{N}(0, \Sigma_{jj}) \quad (2.82)$$

Portanto, o índice BIC de (2.82) é

$$\text{BIC}_j = T \ln \hat{\Sigma}_{jj} + n(\hat{\phi}_j) \ln T, \quad (2.83)$$

onde $n(\hat{\phi}_j)$ é o número de coeficientes da j -ésima componente. Com isso, uma estratégia de seleção do VAR consiste em calcular o índice (2.83) em cada componente, para vários modelos especificados. O modelo eleito consiste na agregação dos correspondentes modelos de menor BIC de cada componente.

O mesmo raciocínio se aplica de maneira análoga para cada estação s e componente j no caso do PVAR:

$$\text{BIC}_{s,j} = q_s \ln \hat{\Sigma}_{s,jj} + n(\hat{\phi}_{s,j}) \ln q_s, \quad (2.84)$$

onde q_s é o maior inteiro positivo tal que $s + q_s \mathbb{S} \leq T$.

2.7 Geração de cenários

Um cenário é uma possível realização das variáveis aleatórias $\{\mathbf{a}_1, \mathbf{a}_2, \dots\}$ definidas por um modelo, isto é, uma possível sequência de ocorrências futuras definida pela modelagem da série temporal. Supondo identificada a ordem do PVAR, estimados os respectivos coeficientes e a distribuição dos ruídos, é possível simular o valor de \mathbf{a}_t apenas sorteando ruídos $\boldsymbol{\varepsilon}_t$ e usando a definição (2.6) do PVAR:

$$\mathbf{a}_t = \hat{\zeta}_t + \sum_{\nu=1}^{p_t} \hat{\Phi}_{t,\nu} \mathbf{a}_{t-\nu} + \boldsymbol{\varepsilon}_t. \quad (2.85)$$

Em seguida, sorteia-se o ruído $\boldsymbol{\varepsilon}_{t+1}$ e por meio da mesma fórmula (2.85) se calcula a realização de \mathbf{a}_{t+1} . Repetindo este procedimento, obtém-se uma realização de $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_T\}$.

Capítulo 3

O modelo PVAR_m

A motivação de se elaborar o modelo Periódico Vetorial Autorregressivo Multiplicativo (PVAR_m) surgiu da dificuldade de se garantir cenários *positivos* de afluência para *modelos lineares* tradicionais. Neste sentido, a ideia do PVAR_m se baseia em uma dinâmica semelhante ao PVAR , porém com uma definição de erro diferente: o erro é a *razão* entre a realização e a previsão de uma dada variável aleatória. Assim, se a previsão for sempre positiva então o erro multiplicativo também será, pois a afluência em cada componente é positiva. Portanto, o *único* requisito da distribuição do erro multiplicativo é que o respectivo suporte esteja contido nos reais positivos, ou seja, não é necessário impor nenhum modelo paramétrico específico para a distribuição do erro multiplicativo como a log-normal a três parâmetros. A seguir, descreve-se mais detalhadamente a proposta do PVAR_m .

3.1 Formulação do modelo PVAR_m

Um processo aleatório $\{\mathbf{a}_t\}_{t=1}^{\infty}$ com \mathbf{a}_t tomando valores em \mathbb{R}^d é dito Periódico Vetorial Autorregressivo Multiplicativo de período \mathbb{S} e ordem $\mathbf{p} = (p_1, \dots, p_{\mathbb{S}})$, $\text{PVAR}_m(\mathbf{p})$, se

$$\mathbf{a}_t = \left(\zeta_t + \sum_{\nu=1}^{p_t} \Phi_{t,\nu} \mathbf{a}_{t-\nu} \right) \bullet \boldsymbol{\eta}_t, \quad (3.1)$$

para todo inteiro positivo t , onde os erros *multiplicativos* $\{\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \dots\}$ são independentes. O símbolo \bullet representa o produto de Hadamard que é definido como um produto entrada a entrada entre vetores ou matrizes, ou seja, $(\mathbf{a} \bullet \mathbf{b})_i = a_i b_i$ e $(A \bullet B)_{ij} = A_{ij} B_{ij}$. Para cada tempo t , os coeficientes e as ordens do modelo são iguais módulo \mathbb{S} ; os erros são identicamente distribuídos módulo \mathbb{S} com média unitária e variância constante módulo \mathbb{S} . Em outras palavras:

$$\bullet \zeta_t = \zeta_{t+\mathbb{S}} \text{ e } \Phi_{t,\nu} = \Phi_{t+\mathbb{S},\nu};$$

- $p_t = p_{t+S}$;
- $\boldsymbol{\eta}_t$ possui a mesma distribuição de $\boldsymbol{\eta}_{t+S}$; e
- $\mathbb{E}[\boldsymbol{\eta}_t] = \mathbf{1}$,

onde $\mathbf{1} := (1, \dots, 1)^\top$. Vale ressaltar que o modelo PVAR_m apresentado é vetorial, pois $\{\mathbf{a}_1, \mathbf{a}_2, \dots\}$ é uma série vetorial. Por conseguinte, $\Phi_{t,1}, \dots, \Phi_{t,p_t}$ são matrizes e $\zeta_t, \boldsymbol{\eta}_t$ são vetores.

Considerando o modelo $\text{PVAR}_m(\mathbf{p})$ (3.1), define-se a *previsão* de \mathbf{a}_t dadas as observações até $t-1$, $\mathbf{a}_{[t-1]} = (\mathbf{a}_1, \dots, \mathbf{a}_{t-1})$, por

$$\text{Prev}[\mathbf{a}_t \mid \mathbf{a}_{[t-1]}] = \zeta_t + \sum_{\nu=1}^{p_t} \Phi_{t,\nu} \mathbf{a}_{t-\nu}. \quad (3.2)$$

Assume-se que as afluências $\mathbf{a}_0, \mathbf{a}_{-1}, \mathbf{a}_{-2}, \dots$ são condições iniciais conhecidas da série. Tomando a esperança condicional dadas as observações até $t-1$ em ambos os lados da equação (3.1), obtém-se a seguinte relação envolvendo a previsão de \mathbf{a}_t :

$$\begin{aligned} \mathbb{E}[\mathbf{a}_t \mid \mathbf{a}_{[t-1]}] &= \left(\zeta_t + \sum_{\nu=1}^{p_t} \Phi_{t,\nu} \mathbf{a}_{t-\nu} \right) \bullet \mathbb{E}[\boldsymbol{\eta}_t \mid \mathbf{a}_{[t-1]}] \\ &= \text{Prev}[\mathbf{a}_t \mid \mathbf{a}_{[t-1]}] \bullet \mathbb{E}[\boldsymbol{\eta}_t \mid \mathbf{a}_{[t-1]}]. \end{aligned} \quad (3.3)$$

Para o cálculo da variância condicional, faz-se as seguintes observações:

1. o produto de Hadamard entre vetores pode ser representado como o produto entre uma matriz diagonal criada por um dos vetores e o outro vetor:

$$\mathbf{a} \bullet \mathbf{b} = \text{diag}(\mathbf{a})\mathbf{b} = \text{diag}(\mathbf{b})\mathbf{a}.$$

A matriz $\text{diag}(\mathbf{a})$ é uma matriz diagonal formada pelo vetor \mathbf{a} , isto é, $[\text{diag}(\mathbf{a})]_{ij} = a_i \delta_{ij}$, onde δ_{ij} é a função delta de Kronecker que vale 1 se i é igual a j e 0 caso contrário;

2. a variância do produto de uma matriz M por um vetor aleatório \mathbf{a} é igual a $M \text{Var}(\mathbf{a}) M^\top$:

$$\text{Var}(M\mathbf{a}) = M \text{Var}(\mathbf{a}) M^\top.$$

Denotando por Prev_t a previsão $\text{Prev}[\mathbf{a}_t \mid \mathbf{a}_{[t-1]}]$, a variância condicional das observações até $t - 1$ em ambos os lados da equação (3.1) é obtida a seguir:

$$\begin{aligned}
\text{Var}[\mathbf{a}_t \mid \mathbf{a}_{[t-1]}] &= \text{Var} \left[\left(\zeta_t + \sum_{\nu=1}^{p_t} \Phi_{t,\nu} \mathbf{a}_{t-\nu} \right) \bullet \boldsymbol{\eta}_t \mid \mathbf{a}_{[t-1]} \right] \\
&= \text{Var} [\text{Prev}_t \bullet \boldsymbol{\eta}_t \mid \mathbf{a}_{[t-1]}] \\
&= \text{Var} [\text{diag}(\text{Prev}_t) \boldsymbol{\eta}_t \mid \mathbf{a}_{[t-1]}] \\
&= \text{diag}(\text{Prev}_t) \text{Var} [\boldsymbol{\eta}_t \mid \mathbf{a}_{[t-1]}] \text{diag}(\text{Prev}_t)^\top \\
&= (\text{Prev}_t \text{Prev}_t^\top) \bullet \text{Var} [\boldsymbol{\eta}_t \mid \mathbf{a}_{[t-1]}]. \tag{3.4}
\end{aligned}$$

Em suma, a esperança e a variância condicional da equação (3.1) são descritas por:

$$\mathbb{E}[\mathbf{a}_t \mid \mathbf{a}_{[t-1]}] = \text{Prev}[\mathbf{a}_t \mid \mathbf{a}_{[t-1]}] \bullet \mathbb{E}[\boldsymbol{\eta}_t \mid \mathbf{a}_{[t-1]}], \tag{3.5}$$

$$\text{Var}[\mathbf{a}_t \mid \mathbf{a}_{[t-1]}] = (\text{Prev}_t \text{Prev}_t^\top) \bullet \text{Var}[\boldsymbol{\eta}_t \mid \mathbf{a}_{[t-1]}], \tag{3.6}$$

Uma *conjectura* sobre o PVAR_m inspirada no modelo PVAR de erro aditivo é a existência de coeficientes $\zeta_t, \Phi_{t,1}, \dots, \Phi_{t,p_t}$ que satisfaçam à seguinte condição chamada de causalidade: a \mathbf{a}_t pode ser descrita como uma função somente dos ruídos multiplicativos passados $\boldsymbol{\eta}_{[t]}$, isto é,

$$\mathbf{a}_t = \psi(\boldsymbol{\eta}_{[t]}).$$

Se essa conjectura for válida, assegura-se a independência entre $\boldsymbol{\eta}_t$ e $\mathbf{a}_{[t-1]}$ e com isso as equações (3.5) e (3.6) podem ser reescritas sem o condicionante $\mathbf{a}_{[t-1]}$ no termo de erro $\boldsymbol{\eta}_t$:

$$\mathbb{E}[\mathbf{a}_t \mid \mathbf{a}_{[t-1]}] = \text{Prev}_t \bullet \mathbb{E}[\boldsymbol{\eta}_t], \tag{3.7}$$

$$\text{Var}[\mathbf{a}_t \mid \mathbf{a}_{[t-1]}] = (\text{Prev}_t \text{Prev}_t^\top) \bullet \text{Var}[\boldsymbol{\eta}_t]. \tag{3.8}$$

As hipóteses sobre a média e a variância do erro $\boldsymbol{\eta}_t$ modificam as expressões (3.7) e (3.8) da seguinte forma:

$$\mathbb{E}[\mathbf{a}_t \mid \mathbf{a}_{[t-1]}] = \text{Prev}[\mathbf{a}_t \mid \mathbf{a}_{[t-1]}], \tag{3.9}$$

$$\text{Var}[\mathbf{a}_t \mid \mathbf{a}_{[t-1]}] = (\text{Prev}_t \text{Prev}_t^\top) \bullet \Sigma_t, \tag{3.10}$$

onde $\Sigma_t := \text{Var}[\boldsymbol{\eta}_t]$. Portanto, supondo verdadeira essa conjectura, a média condicional de \mathbf{a}_t é igual a previsão e a variância condicional (volatilidade) de \mathbf{a}_t é proporcional ao produto cruzado das previsões.

O significado da equação (3.10) merece ser enfatizado. Em particular, a variância condicional da i -ésima componente de \mathbf{a}_t pode ser obtida de (3.10):

$$\begin{aligned}\text{Var}[a_{t,i} \mid \mathbf{a}_{[t-1]}] &= \text{Var}[\mathbf{a}_t \mid \mathbf{a}_{[t-1]}]_{ii} \\ &= (\text{Prev}_t \text{Prev}_t^\top)_{ii} \Sigma_{t,ii} \\ &= \text{Prev}_{t,i}^2 \Sigma_{t,ii},\end{aligned}\tag{3.11}$$

ou seja, decorre da equação (3.10) que a volatilidade de uma componente $a_{t,i}$ é proporcional ao quadrado respectiva previsão $\text{Prev}_{t,i}$, onde $\text{Prev}_{t,i} := \text{Prev}[\mathbf{a}_t \mid \mathbf{a}_{[t-1]}]_i$.

Se a conjectura sobre o modelo PVAR_m for verdadeira então a equação (3.10) se torna uma evidência de que o PVAR_m é um modelo diferente do PVAR , pois a variância condicional do PVAR não depende da previsão. Além disso, a propriedade da variância condicional (3.10) pode ser interessante para séries nas quais se observa volatilidade elevada para realizações elevadas e volatilidade baixa para realizações baixas.

3.2 Estimação dos parâmetros

Como o objetivo inicial do modelo PVAR_m é produzir cenários positivos, é necessário garantir previsões sempre positivas. Neste sentido, o método de mínimos quadrados com restrição nos coeficientes, exposto na seção 2.3.3 sobre a estimação do modelo PVAR , atende a esse requisito e se aplica da mesma forma.

O método de mínimos quadrados e o de Yule-Walker são adequados para outras aplicações do PVAR_m que não necessitem de cenários positivos. Porém, para sua dedução, o estimador de Yule-Walker requer a hipótese adicional da esperança condicional do erro multiplicativo condicionado as afluições até $t - 1$ ser um, isto é, $\mathbb{E}[\boldsymbol{\eta} \mid \mathbf{a}_{[t-1]}] = \mathbf{1}$. Admitindo essa hipótese, a construção do estimador de Yule-Walker segue de forma semelhante à feita para o PVAR .

3.3 Determinação dos resíduos

O erro (ou ruído) multiplicativo do modelo PVAR_m é definido como a razão termo a termo entre o valor observado e o previsto:

$$\boldsymbol{\eta}_t = \frac{\mathbf{a}_t}{(\zeta_t + \sum_{\nu=1}^{p_t} \Phi_{t,\nu} \mathbf{a}_{t-\nu})}.\tag{3.12}$$

Como os valores “*verdadeiros*” dos parâmetros (3.12) não são conhecidos a priori, o ruído $\boldsymbol{\eta}_t$ é não observável. O resíduo, por outro lado, é uma estimativa observável

do erro multiplicativo $\boldsymbol{\eta}_t$ e que é definido como a razão entre o valor observado e o previsto *estimado*:

$$\hat{\boldsymbol{\eta}}_t = \frac{\mathbf{a}_t}{\left(\hat{\zeta}_t + \sum_{\nu=1}^{p_t} \hat{\Phi}_{t,\nu} \mathbf{a}_{t-\nu} \right)}. \quad (3.13)$$

Vale ressaltar que os resíduos $\hat{\boldsymbol{\eta}}_t$ para t de 1 até T são em geral dependentes, enquanto que os erros $\boldsymbol{\eta}_t$ são, por hipótese, independentes.

Um dos principais usos dos resíduos é na estimação de quantidades que dependem dos ruídos $\boldsymbol{\epsilon}_t$. Na próxima seção, descrevem-se alguns modelos de distribuição do ruído $\boldsymbol{\epsilon}_t$ e como estimá-los.

3.4 Estimação da distribuição dos ruídos

Como a previsão do PVAR_m é sempre positiva por hipótese, então a única restrição para a estimação da distribuição dos ruídos é atribuir probabilidade *nula* para os erros não contidos nos reais positivos. Com o intuito de reduzir o número de hipóteses do PVAR_m , opta-se pela distribuição empírica como o estimador dos ruídos multiplicativos multivariados:

$$\hat{F}_s(\mathbf{x}) = \frac{1}{q_s + 1} \sum_{\nu=0}^{q_s} \mathbb{I}(\hat{\boldsymbol{\eta}}_{s+\nu\mathbb{S}} \leq \mathbf{x}), \quad (3.14)$$

onde $\hat{F}_s(\cdot)$ é a distribuição acumulada empírica associada a estação s , $\hat{\boldsymbol{\eta}}_{s+\nu\mathbb{S}}$ é um resíduo da estação s e $\mathbb{I}(\cdot)$ é a função indicadora.

3.5 Identificação do modelo

A seleção do PVAR_m consiste em, fixada uma estação s , calcular o índice BIC em cada componente para vários modelos especificados. O modelo eleito se baseia na agregação dos correspondentes modelos de menor BIC de cada componente. O índice usado para a identificação do PVAR_m é o mesmo que o do PAR (univariado) com erro aditivo normal e por isso é um proxy para a identificação da ordem.

3.6 Geração de cenários

O estimador dos ruídos dado pela distribuição empírica possui uma desvantagem: o número de cenários distintos que se pode gerar é finito e limitado pelo número de resíduos $\hat{\boldsymbol{\eta}}_s, \hat{\boldsymbol{\eta}}_{s+\mathbb{S}}, \dots, \hat{\boldsymbol{\eta}}_{s+q_s\mathbb{S}}$. Considerando a distribuição empírica, descreve-se

a maneira tradicional de se gerar os ruídos e dois procedimentos adicionais para aumentar sua variabilidade.

Supondo identificada a ordem do PVAR_m , estimados os respectivos coeficientes e calculados os resíduos é possível simular o valor de \mathbf{a}_t apenas sorteando o ruído $\boldsymbol{\eta}_t$ e usando a definição (3.1) do PVAR_m :

$$\mathbf{a}_t = \left(\hat{\zeta}_t + \sum_{\nu=1}^{p_t} \hat{\Phi}_{t,\nu} \mathbf{a}_{t-\nu} \right) \bullet \boldsymbol{\eta}_t. \quad (3.15)$$

Em seguida, sorteia-se o ruído $\boldsymbol{\eta}_{t+1}$ pelo mesmo procedimento e por meio de (3.15) se calcula a realização de \mathbf{a}_{t+1} . Repetindo este processo, obtém-se uma realização de $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_T\}$.

3.6.1 Geração bootstrap

A geração bootstrap é feita diretamente da distribuição empírica dos ruídos. Por ser uma distribuição discreta com probabilidade igual para cada resíduo, o sorteio da distribuição empírica equivale a uma reamostragem dos resíduos. Assim, existem no máximo $q_s + 1$ sorteios diferentes do erro multiplicativo $\boldsymbol{\eta}_s$ de cada estação s .

3.6.2 Geração PCA-bootstrap

Uma forma de aumentar a variabilidade dos resíduos consiste em utilizar a técnica de análise de componentes principais, ou *Principal Components Analysis* (PCA).

A técnica de componentes principais é baseada na decomposição em valores singulares (SVD) de uma matriz. Esse método de decomposição é válido para qualquer matriz $M \in \mathbb{R}^{m \times n}$:

$$M = UDV^T,$$

onde $U \in \mathbb{R}^{m \times m}$, $V \in \mathbb{R}^{n \times n}$ são matrizes ortogonais e $D \in \mathbb{R}^{m \times n}$ é uma matriz diagonal de entradas não negativas cujos elementos da diagonal são ordenados de forma decrescente. Em particular, se M é uma matriz quadrada de tamanho n , simétrica e semi-definida positiva então é possível mostrar que a decomposição SVD possui a seguinte expressão:

$$M = UDU^T.$$

A ideia de componentes principais consiste em aplicar a fatoração SVD na matriz de covariância Σ do correspondente vetor aleatório \mathbf{X} . Por ser quadrada, simétrica e semi-definida positiva a matriz de covariância Σ possui a seguinte decomposição

SVD:

$$\Sigma = UDU^\top. \quad (3.16)$$

Com isso, o vetor aleatório \mathbf{Y} resultante do produto de U^\top por \mathbf{X} é descorrelacionado:

$$\begin{aligned} \text{Var}[\mathbf{Y}] &= \text{Var}[U^\top \mathbf{X}] \\ &= U^\top \text{Var}[\mathbf{X}](U^\top)^\top \\ &= U^\top UDU^\top U \\ &= D, \end{aligned}$$

pois D é uma matriz diagonal. Para se recuperar o vetor aleatório \mathbf{X} basta calcular $U\mathbf{Y}$, afinal a matriz U é unitária, isto é, $UU^\top = U^\top U = I$.

A ideia do PCA-bootstrap é decompor o vetor de ruídos aleatórios em componentes descorrelacionadas, sortear independentemente um valor de cada coordenada e depois correlacioná-los novamente. Desse modo, obtém-se um número total de $(q_s + 1)^n$ amostras distintas, pois $q_s + 1$ são os possíveis valores de cada componente e n é o número total de componentes.

A desvantagem desse método é que sortear independentemente valores de componentes descorrelacionadas e agrupá-los em um vetor pode implicar em uma perda de informação, já que essas componentes descorrelacionadas ainda podem ser *dependentes*. Porém, se os ruídos possuem uma distribuição normal multivariada então é possível mostrar que a correlação nula equivale a independência das componentes. Com o intuito de mitigar padrões de assimetria e portanto perder menos informação com a técnica de componentes principais ¹, toma-se o logaritmo dos resíduos $\log \hat{\boldsymbol{\eta}}_s, \log \hat{\boldsymbol{\eta}}_{s+\mathcal{S}}, \dots, \log \hat{\boldsymbol{\eta}}_{s+q_s\mathcal{S}}$. A seguir, descreve-se detalhadamente o método PCA-bootstrap.

O logaritmo do ruído $\log \boldsymbol{\eta}_t$ em uma dada estação s possui a seguinte distribuição acumulada:

$$\hat{F}_s^{\log}(\mathbf{y}) = \frac{1}{q_s + 1} \sum_{\nu=0}^{q_s} \mathbb{I}(\log \hat{\boldsymbol{\eta}}_{s+\nu\mathcal{S}} \leq \mathbf{y}), \quad (3.17)$$

onde \mathbf{y} é apenas uma notação para o argumento da distribuição acumulada. Em seguida, aplica-se a decomposição em componentes principais no logaritmo do ruído, isto é, utiliza-se a decomposição SVD na matriz de covariância de $\log \boldsymbol{\eta}_t$ que é

¹A inspiração é a lognormal multivariada, pois toma valores nos reais positivos e ao se aplicar o logaritmo e posteriormente gerar as componentes principais se obtém componentes independentes.

denotada por $\widehat{\Sigma}_s^{\log}$:

$$\widehat{\Sigma}_s^{\log} = UDU^\top. \quad (3.18)$$

Assim, as componentes do vetor aleatório $\boldsymbol{\vartheta}_t$ definido por $U^\top \log \boldsymbol{\eta}_t$ são decorrelacionadas e conjuntamente possuem distribuição acumulada igual a

$$\widehat{G}_s(\mathbf{z}) = \frac{1}{q_s + 1} \sum_{\nu=0}^{q_s} \mathbb{I}(\widehat{\boldsymbol{\vartheta}}_{s+\nu\mathbb{S}} \leq \mathbf{z}), \quad (3.19)$$

onde $\widehat{\boldsymbol{\vartheta}}_{s+\nu\mathbb{S}} := U^\top \log \widehat{\boldsymbol{\eta}}_{s+\nu\mathbb{S}}$.

Para a simulação, assume-se que as coordenadas $(\vartheta_{t,1}, \dots, \vartheta_{t,n})$ de $\boldsymbol{\vartheta}_t$ são independentes e com distribuição acumulada conjunta igual ao produto das marginais:

$$\widehat{G}_s(\mathbf{z}) = \widehat{G}_{s,1}(z_1) \widehat{G}_{s,2}(z_2) \cdots \widehat{G}_{s,n}(z_n), \quad (3.20)$$

onde a função $\widehat{G}_{s,i}(\cdot)$ é a distribuição acumulada da i -ésima coordenada $\vartheta_{i,t}$ enunciada abaixo:

$$\widehat{G}_{s,i}(z_i) = \frac{1}{q_s + 1} \sum_{\nu=0}^{q_s} \mathbb{I}(\widehat{\vartheta}_{s+\nu\mathbb{S},i} \leq z_i), \quad (3.21)$$

na qual $\widehat{\vartheta}_{s+\nu\mathbb{S},i}$ é a i -ésima coordenada do resíduo $\widehat{\boldsymbol{\vartheta}}_{s+\nu\mathbb{S}}$. O procedimento adotado para a geração dos cenários consiste em sortear cada coordenada $\vartheta_{s+\nu\mathbb{S},i}$ formando um vetor $\boldsymbol{\vartheta}_t$, correlacionar o vetor obtido multiplicando por U e exponenciar o resultado recuperando o valor do ruído multiplicativo:

$$\boldsymbol{\eta}_t = e^{U\boldsymbol{\vartheta}_t}.$$

3.6.3 Geração PCA-bootstrap suavizado

O método PCA-bootstrap suavizado tem por objetivo aumentar ainda mais a variabilidade das amostras dos resíduos. O método consiste em decompor o logaritmo do vetor de ruídos em componentes decorrelacionadas assim como o método PCA-bootstrap, porém se propõe representar a distribuição de cada coordenada por um estimador *não paramétrico* contínuo obtido a partir das mesmas amostras consideradas na distribuição acumulada empírica (3.21). Desta forma, é possível gerar um número de amostras distintas tão grande quanto se queira. A partir disso, todo o procedimento é idêntico, ou seja, os valores de cada coordenada são sorteados, agregados em um vetor, correlacionados e exponenciados fornecendo

uma amostra do erro $\boldsymbol{\eta}_t$.

O procedimento utilizado para estimar a distribuição contínua de cada coordenada é o método de Kernel (SILVERMAN [65]) que consiste em aproximar a função densidade de probabilidade do processo aleatório. A expressão do estimador da densidade da i -ésima componente de $\boldsymbol{\vartheta}_t$ associado a estação s é dada abaixo

$$\hat{g}_{s,i}(z_i) = \frac{1}{(q_s + 1)h} \sum_{\nu=0}^{q_s} f\left(\frac{z_i - \hat{\vartheta}_{s+\nu\mathbb{S},i}}{h}\right), \quad (3.22)$$

onde $f(\cdot)$ é a função Kernel e h é um parâmetro suavizador escolhido. A função kernel $f(\cdot)$ utilizada nesta dissertação é a normal padrão. Um método eficiente para simular variáveis aleatórias $\vartheta_{s,i}$ que possuem densidade igual a (3.22) é sortear um índice ν de uma uniforme discreta cujos valores são $\{0, 1, \dots, q_s\}$ e em seguida sortear uma amostra da normal com média $\hat{\vartheta}_{s+(\nu-1)\mathbb{S},i}$ e variância h^2 . A densidade (3.22) e este método de simulação podem ser interpretados como a densidade e a simulação de uma mistura de normais.

Capítulo 4

Caso exemplo

Para ilustrar a metodologia proposta, considera-se uma configuração do SIN constituída por quatro subsistemas interligados (SE, S, NE, N) conforme os dados disponíveis para o PMO de Janeiro de 2015 no site do ONS (ONS [1]). As ENAs correspondentes a esta configuração são usadas no ajuste do modelo proposto e na avaliação dos cenários gerados.

O modelo proposto é o modelo $PVAR_m$ com restrição de positividade nos coeficientes. Para este caso exemplo, os coeficientes do modelo $PVAR_m$ são estimados pelo método de mínimos quadrados com restrição nos coeficientes (seção 3.2), a ordem do modelo é identificada por um proxy do BIC (seção 3.5) e os erros multiplicativos são gerados pelo método PCA-bootstrap (seção 3.6.2).

O desempenho do modelo multiplicativo $PVAR_m$ é comparado com o desempenho do correspondente modelo aditivo PVAR de erro multivariado gaussiano (seção 2). A justificativa para esta comparação é que o modelo aditivo PVAR possui um bom desempenho na reprodução dos fatos estilizados de ENA e sua versão sem restrição nos coeficientes tende a ter um ajuste ainda melhor, embora não esteja imune à ocorrência de valores negativos.

Para cada modelo, a simulação resultante é composta por 5000 cenários que consideram como condição inicial os valores históricos correspondentes ao ano de 1931. Cada cenário é composto de 80 anos totalizando 960 vetores (SE, S, NE, N) de valores mensais de Energia Natural Afluyente (ENA). Assume-se a estacionariedade periódica e onde for necessário a ergodicidade periódica da série histórica. Por exemplo, assume-se que a média, variância, assimetria e curtose de janeiro são iguais ao longo dos anos e os respectivos estimadores convergem. Do mesmo modo para outros meses.

4.1 Cenários gerados

Nesta seção apresentam-se os gráficos dos modelos de erro aditivo e multiplicativo que enfatizam a importância das opções de modelagem da proposta $PVAR_m$ para a geração de cenários positivos. O tipo de gráfico escolhido é um fanplot que representa a concentração dos cenários por meio de uma escala de cores contendo o azul, cinza e branco, onde estas cores representam uma alta, baixa e nula concentração de observações, respectivamente. Destaca-se o quantil 0.5% dos valores de cada estágio por uma linha azul escura de maior espessura. Abaixo desta linha espessa azul existem 25 dos 5000 cenários totais, cujo propósito é ilustrar a existência de valores negativos gerados por alguns modelos. Com o intuito de facilitar a visualização de valores negativos, a escala do eixo x de cada gráfico contém apenas os 40 primeiros meses da simulação.

As figuras 4.1, 4.2, 4.3, e 4.4 apresentam o fanplot dos 5000 cenários gerados para o subsistema NE (subsistema 3) com os modelos PVAR de erro aditivo multivariado normal e $PVAR_m$ de erro multiplicativo multivariado PCA-bootstrap, onde ambos foram estimados por mínimos quadrados *sem* e *com* restrição nos coeficientes. Dos quatro modelos simulados, apenas o modelo multiplicativo $PVAR_m$ com restrição nos coeficientes (figura 4.4) gerou *sempre* cenários positivos de ENA. Por construção, o modelo $PVAR_m$ com restrição nos coeficientes garante a positividade dos cenários.

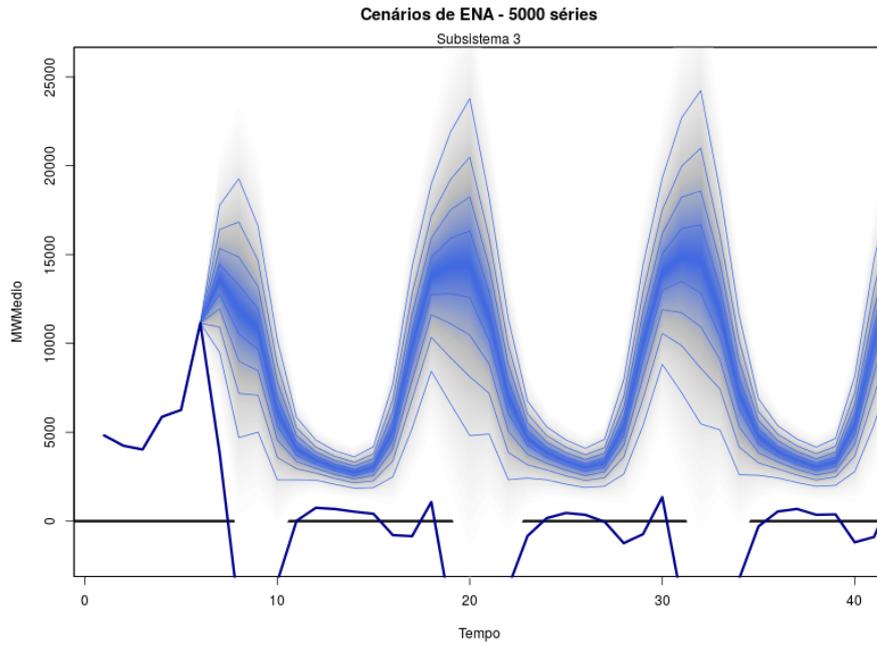


Figura 4.1: PVAR – cenários gerados subsistema NE (3)
(aditivo sem restrição)

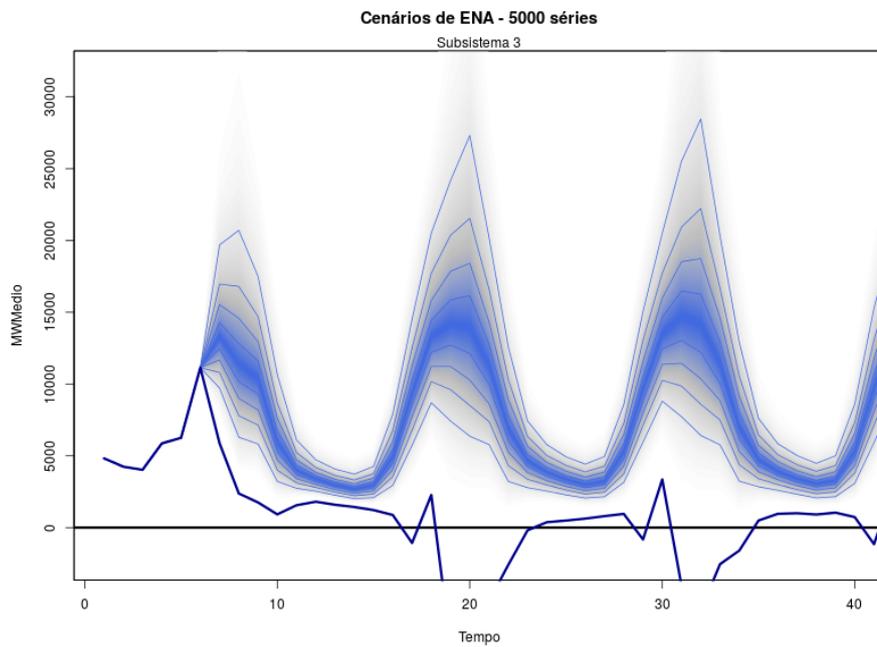


Figura 4.2: PVAR_m – cenários gerados subsistema NE (3)
(mult. sem restrição)

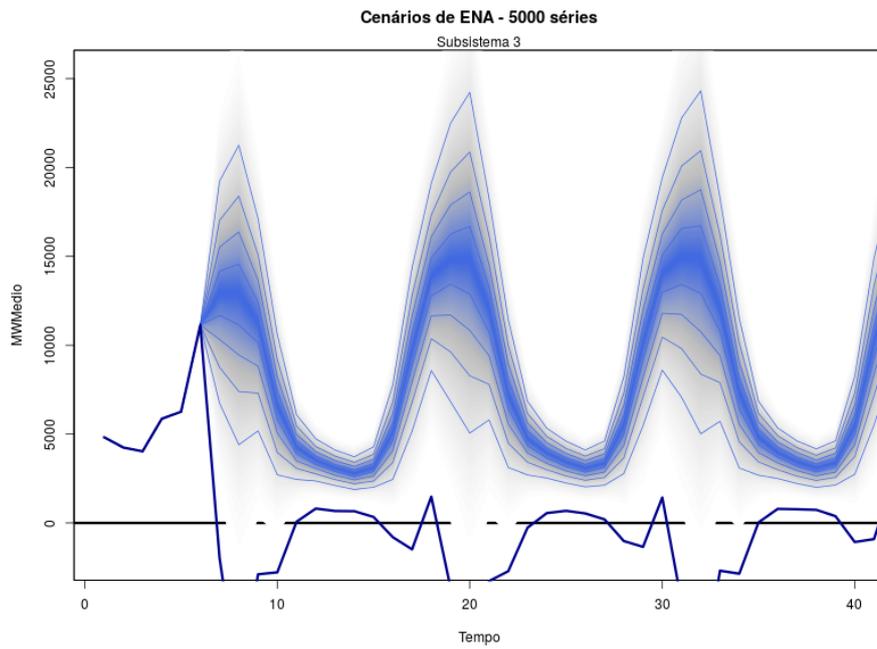


Figura 4.3: PVAR – cenários gerados subsistema NE (3)
(aditivo com restrição)

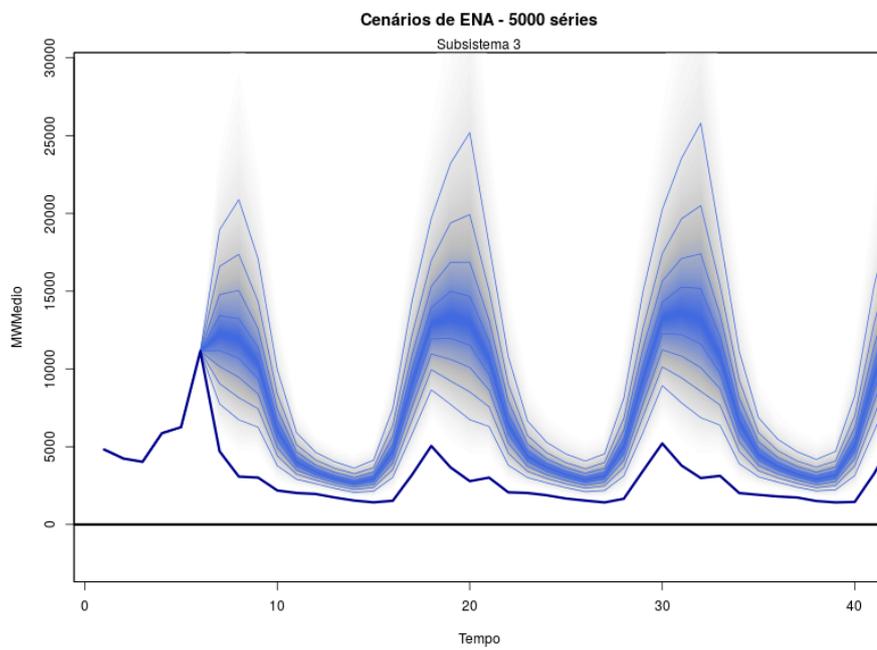


Figura 4.4: PVAR_m – cenários gerados subsistema NE (3)
(mult. com restrição)

4.2 Estatísticas descritivas básicas

As estatísticas descritivas básicas são medidas das características de tendência central e dispersão das amostras. Nas seções subsequentes, apresenta-se uma avaliação gráfica do desempenho dos cenários gerados face aos valores dos registros históricos de ENA disponíveis.

Observa-se que cada modelo de série temporal considera uma *distribuição de probabilidade* para os cenários e, conseqüentemente, para as estatísticas dos cenários. Como a série histórica de ENA é apenas um cenário admissível, optou-se por comparar a *distribuição* das estatísticas dos cenários simulados de cada modelo com as correspondentes *realizações* das estatísticas históricas.

A distribuição estatística dos cenários simulados por cada modelo é representada por um boxplot e a respectiva estatística histórica é representada por uma linha espessa azul. A intuição é que se os fatos estilizados históricos estiverem centrados em torno da distribuição dos fatos estilizados simulados então uma proporção considerável dos cenários simulados apresenta características semelhantes a do histórico.

Os fatos estilizados analisados a seguir são a média, desvio padrão, assimetria, curtose e a correlação espacial entre os subsistemas.

Média

As figuras 4.5 e 4.6 apresentam para o subsistema SE (subsistema 1) os boxplots das correspondentes médias mensais obtidas a partir dos cenários gerados pelo modelo PVAR e PVAR_m, sem e com restrição de positividade nos coeficientes. Relembrando, optou-se por comparar a proposta PVAR_m com restrição com o modelo tradicional PVAR sem restrição pelo bom desempenho que o PVAR apresenta.

Os valores gerados estão indicados em preto e os valores históricos em azul. Nota-se que os valores de média de todos os modelos se ajustam bem aos correspondentes valores históricos. Em termos de média, o desempenho dos modelos PVAR sem restrições e do PVAR_m com restrições nos coeficientes, é indistinguível.

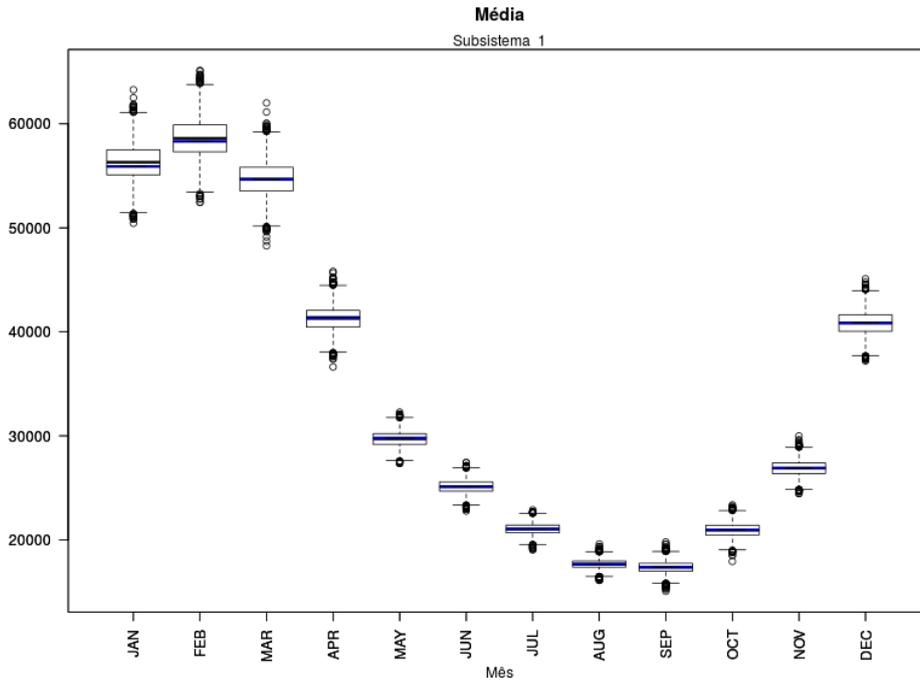


Figura 4.5: PVAR – média do subsistema SE (1)
(aditivo sem restrição)

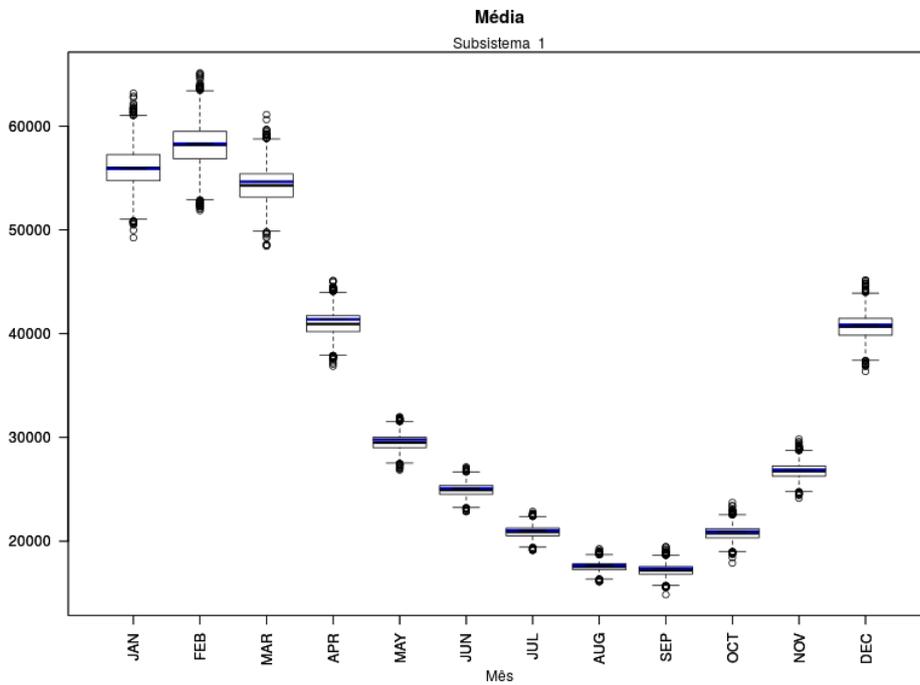


Figura 4.6: PVAR_m – média do subsistema SE (1)
(multiplicativo com restrição)

Desvio padrão

As figuras 4.7 e 4.8 mostram para o subsistema SE (subsistema 1) que o desvio padrão de ambos os modelos PVAR e PVAR_m apresentam um comportamento sazonal semelhante ao histórico. Particularmente, o modelo PVAR_m com restrição nos coeficientes apresenta valores ligeiramente superestimados nos meses de janeiro e fevereiro do sudeste.

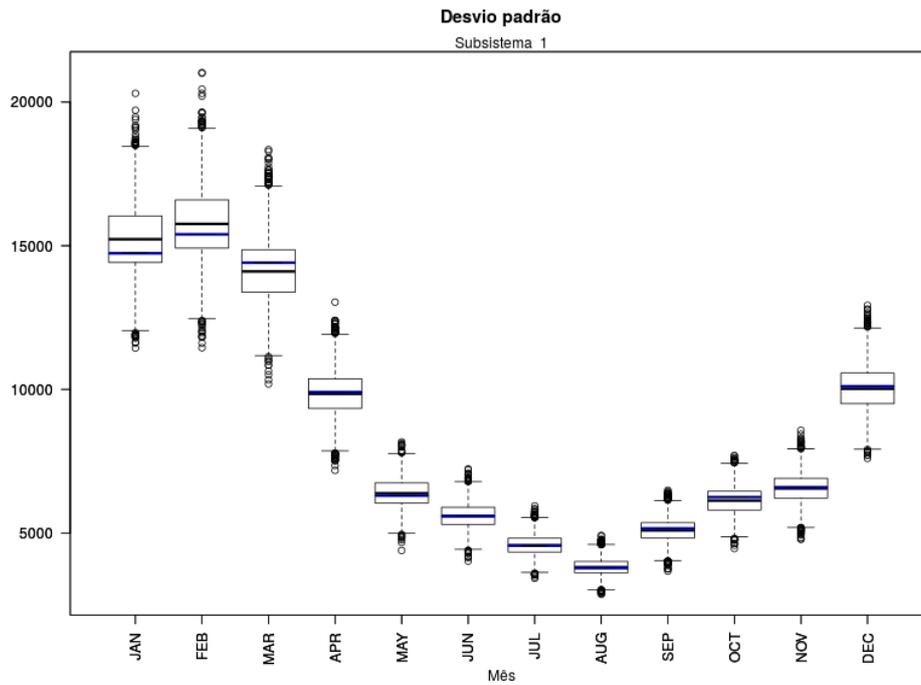


Figura 4.7: PVAR – desvio padrão do subsistema SE (1)
(aditivo sem restrição)

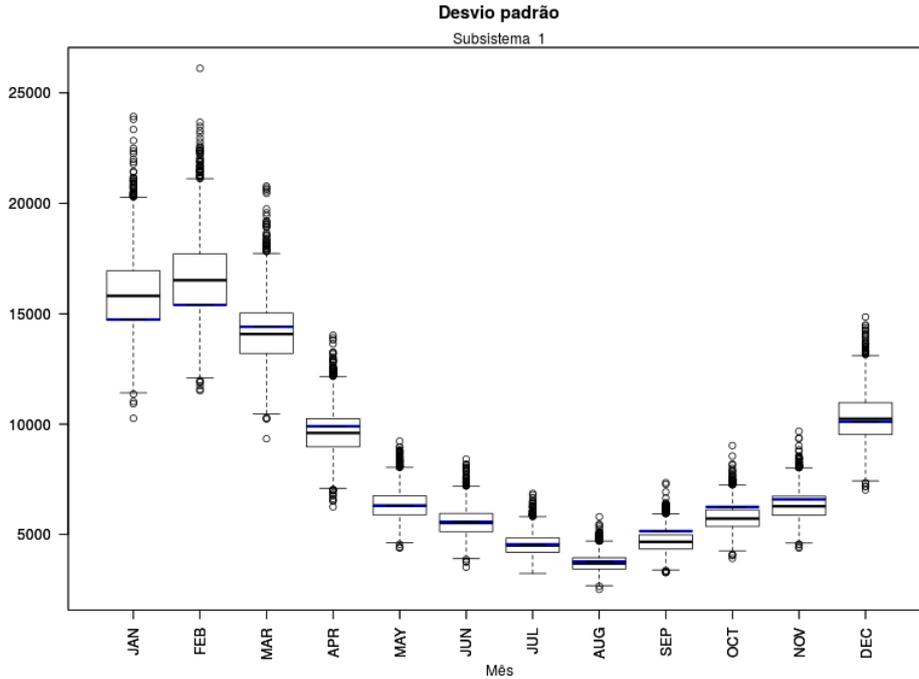


Figura 4.8: $PVAR_m$ – desvio padrão do subsistema SE (1)
(mult. com restrição)

Assimetria

As figuras 4.9 e 4.10 apresentam para o subsistema SE (subsistema 1) os boxplots das correspondentes assimetrias mensais dos modelos PVAR e $PVAR_m$. Por não ser um parâmetro de ajuste dos modelos PVAR e $PVAR_m$, não se espera preservar o coeficiente de assimetria mensal, embora desejável a compatibilidade entre os cenários gerados e históricos com relação a esta estatística.

Destaca-se um elevado número de outliers nos gráficos de assimetria dos cenários gerados pelo modelo multiplicativo, o que significa que o modelo multiplicativo eventualmente gera cenários mais otimistas do que o modelo aditivo. Uma das causas deste otimismo pode ser a restrição de positividade nos coeficientes.

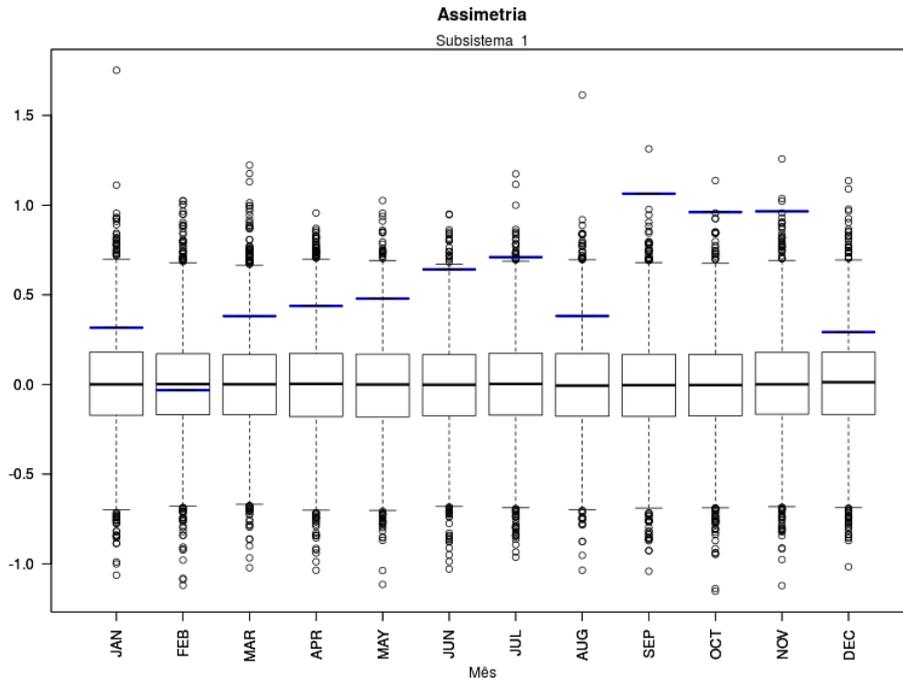


Figura 4.9: PVAR – assimetria do subsistema SE (1)
(aditivo sem restrição)

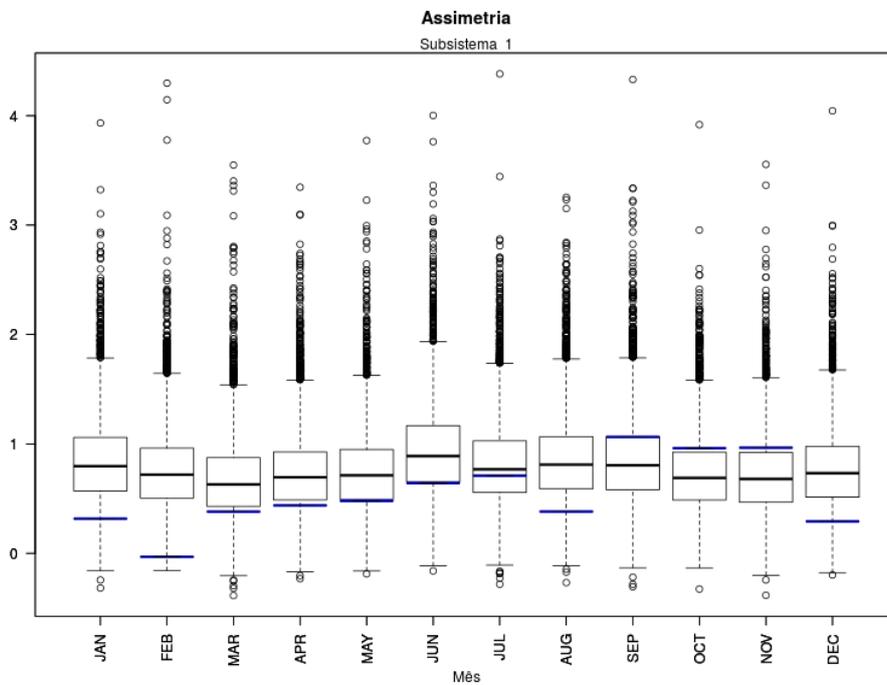


Figura 4.10: PVAR_m – assimetria do subsistema SE (1)
(mult. com restrição)

Curtose

As figuras 4.11 e 4.12 apresentam para o subsistema SE (subsistema 1) os boxplots das correspondentes curtoses mensais obtidas a partir dos cenários gerados pelos modelos PVAR e PVAR_m. Destaca-se a ocorrência de cenários com curtose elevada em ambos os modelos e a diferença entre as escalas. Isto pode ser um indicativo de que ambos os modelos associam probabilidades elevadas á eventos extremos.

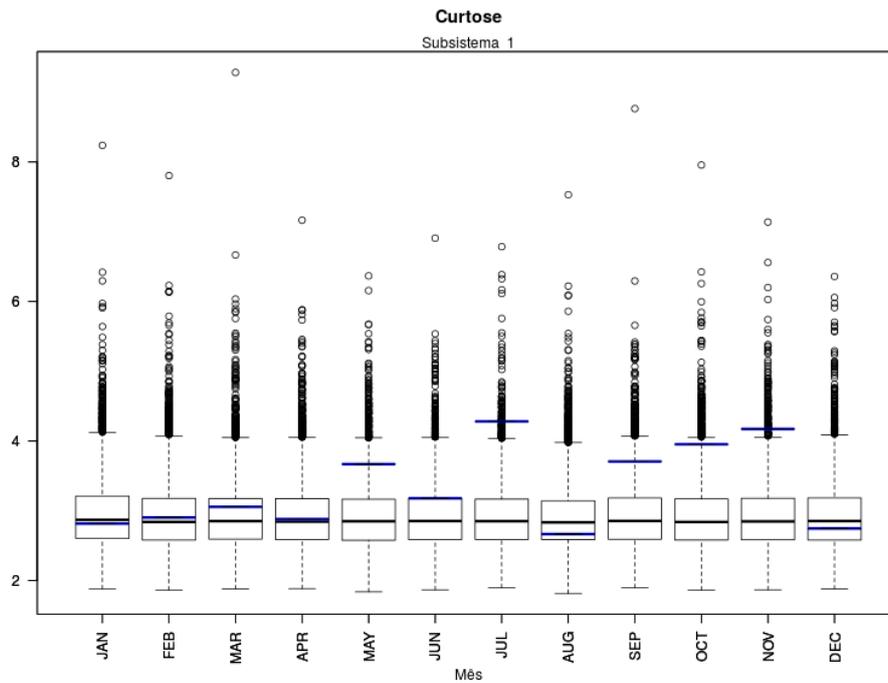


Figura 4.11: PVAR – curtose do subsistema SE (1)
(aditivo sem restrição)

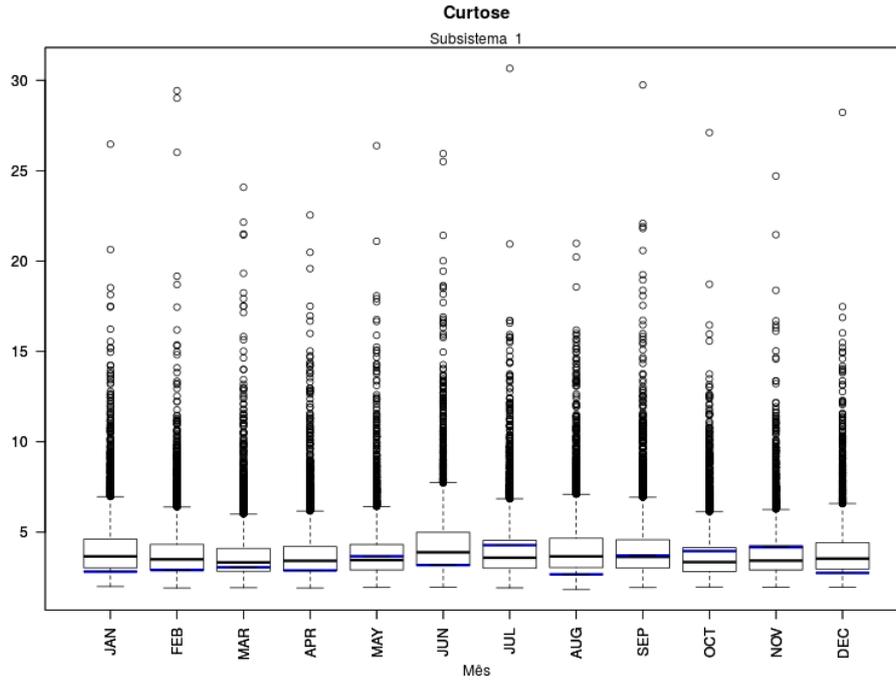


Figura 4.12: $PVAR_m$ – curtose do subsistema SE (1)
(mult. com restrição)

Correlação espacial

As figuras 4.13 e 4.14 apresentam os boxplots das correspondentes correlações espaciais cruzadas mensais contemporâneas entre os subsistemas SE (subsistema 1) e NE (subsistema 3). A correlação espacial contemporânea é a correlação entre dois subsistemas fixado um dado mês. O estimador da correlação espacial associado às séries simuladas, em geral, não desvia demasiadamente dos correspondentes estimadores históricos em ambos os modelos.

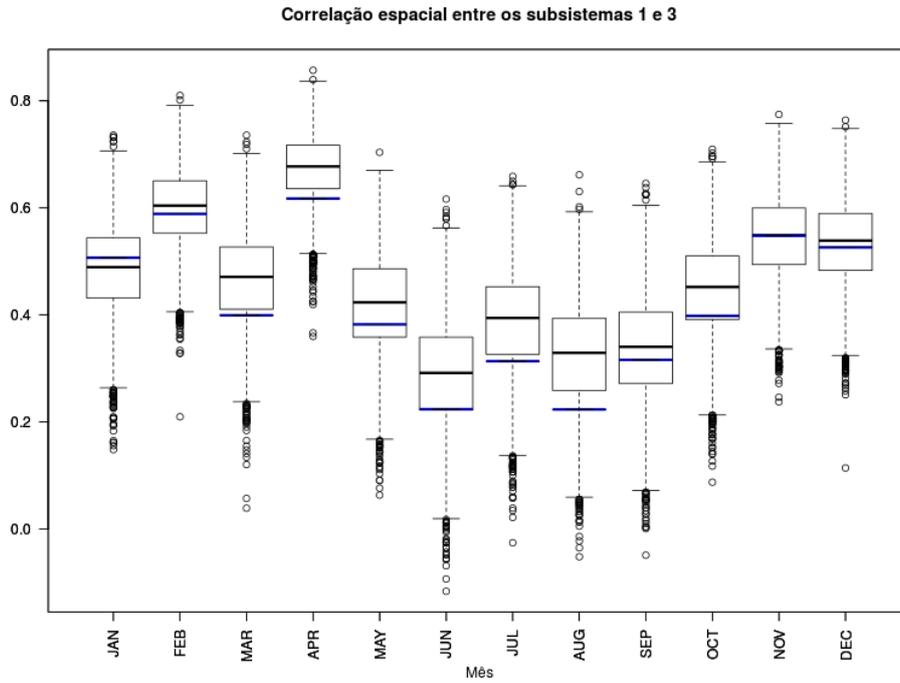


Figura 4.13: PVAR – correl. espacial entre SE (1) e NE (3)
(aditivo sem restrição)

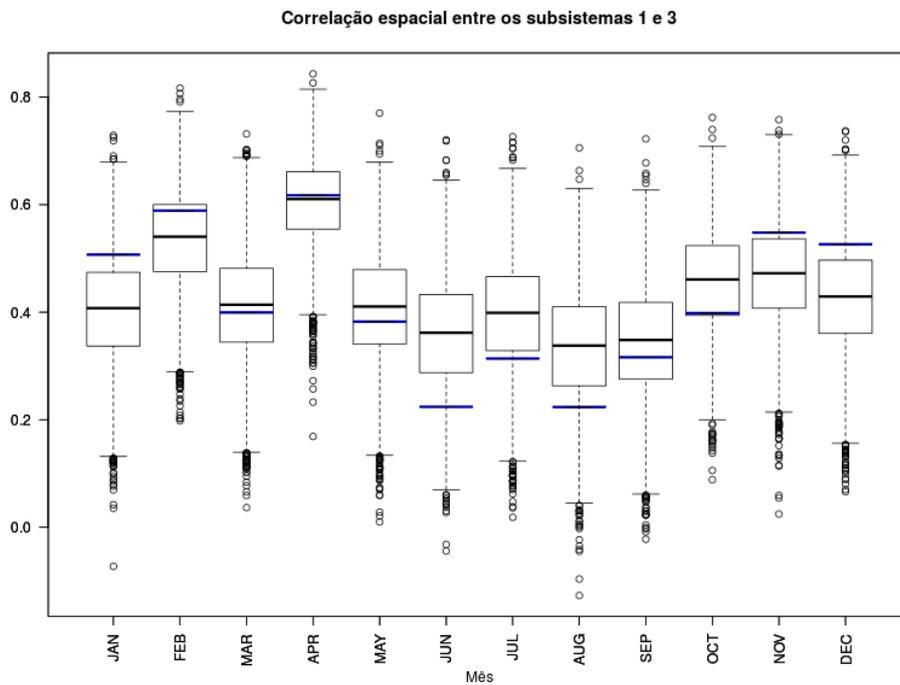


Figura 4.14: $PVAR_m$ – correl. espacial entre SE (1) e NE (3)
(mult. com restrição)

4.3 Análise das distribuições de ENA

Um outro tipo de gráfico escolhido para comparar as distribuições históricas e simuladas de ENA é o qq-plot (gráfico quantil-quantil). O qq-plot compara duas distribuições de probabilidades posicionando os respectivos quantis uns contra os outros. Se duas distribuições comparadas são semelhantes, os respectivos pontos do gráfico estarão próximos da diagonal ($y = x$).

A seguir, apresenta-se o qq-plot das ENAs históricas e sintéticas do modelo $PVAR_m$ com restrição nos coeficientes para os meses de fevereiro (período úmido) e agosto (período seco) em cada subsistema. Os dados utilizados foram centrados e padronizados para facilitar a interpretação dos valores de cada eixo. O valor zero em cada eixo corresponde à média da distribuição e os valores positivos e negativos correspondem às proporções de desvios padrão acima e abaixo da média, respectivamente. A linha vermelha que corta o gráfico é a diagonal ($y = x$) de referência que auxilia a interpretação dos resultados em diferentes escalas.

O eixo x contém o quantil amostral dos 5000×80 dados sintéticos centrados e padronizados de um certo mês, sendo 5000 o total de séries sintéticas geradas e 80 é o número de anos simulados a partir dos dados históricos do ano de 1931. O eixo y contém o quantil amostral dos 83 dados históricos centrados e padronizados de um certo mês, onde 83 é o número de anos do histórico com início em 1931 e término em 2013. Observa-se que os quantis amostrais sintéticos e históricos possuem aproximadamente o mesmo valor em torno da média e valores bem distintos para amostras muito acima da média, em todos os meses e subsistemas. Por simplicidade, apresenta-se apenas os para o subsistema SE (subsistema 1).

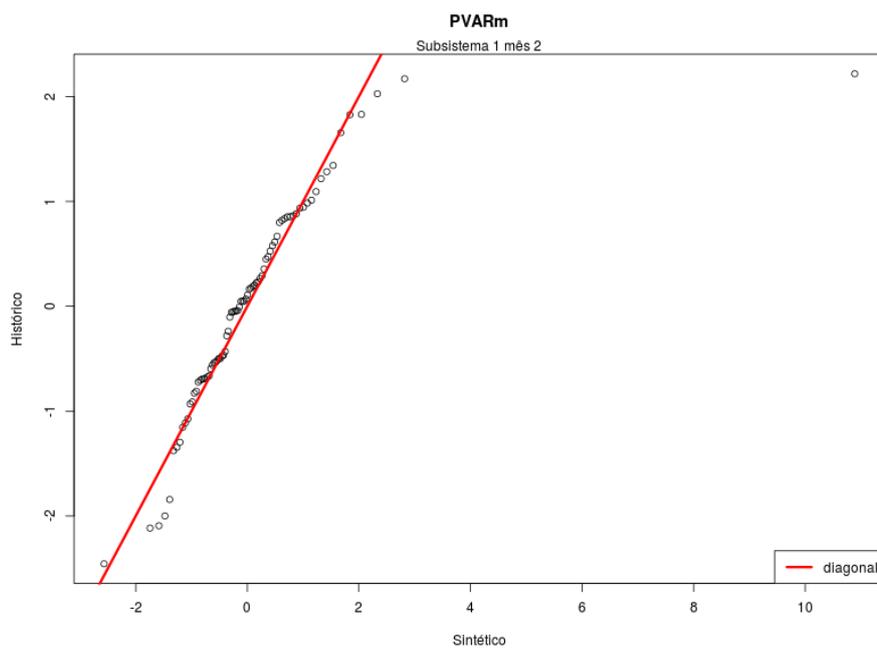


Figura 4.15: $PVAR_m$ – qqplot SE (1) e mês de fevereiro (mult. com restrição)

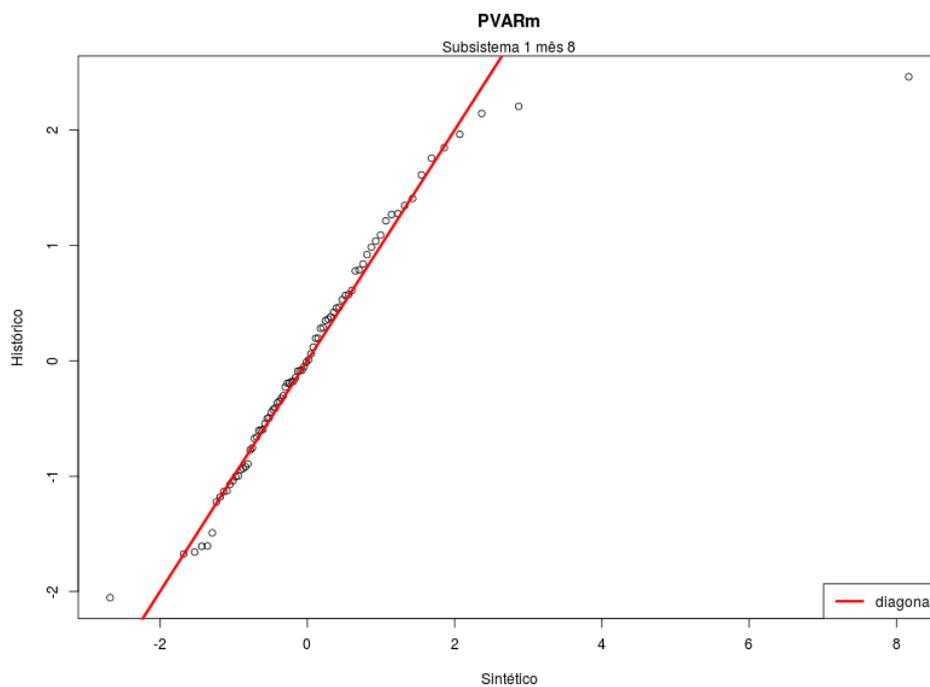


Figura 4.16: $PVAR_m$ – qqplot SE (1) e mês de agosto (mult. com restrição)

4.4 Análise descritiva de “runs”

Entende-se por *run* uma sequência de observações de mesmo tipo precedida e sucedida por uma ou mais observações de tipo diferente. O propósito da *análise de runs* em estudos hidrológicos é avaliar a capacidade de um modelo reproduzir os valores consecutivos históricos de afluência (ENA, por exemplo) abaixo e acima da média.

A análise gráfica exposta a seguir é idêntica à realizada na seção 4.2 para as estatísticas descritivas básicas, isto é, a distribuição estatística dos cenários simulados por cada modelo é representada por um boxplot e a respectiva estatística histórica é representada por uma linha espessa azul. Relembrando, a intuição é que se os fatos estilizados históricos estiverem centrados em torno da distribuição dos fatos estilizados simulados então uma proporção considerável dos cenários simulados apresenta características semelhantes a do histórico.

Com o intuito de não tornar a exposição enfadonha, optou-se por expor apenas as estatísticas de runs de valores abaixo da média.

4.4.1 Sequências abaixo da média

Dada a importância do recurso hidráulico para o Setor Elétrico Brasileiro, é fundamental a representação dos valores de ENA abaixo da média pelos modelos de séries temporais projetados para o uso no planejamento da operação energética. As estatísticas listadas a seguir auxiliam a análise da representatividade dos runs abaixo da média pelo modelo multiplicativo $PVAR_m$:

- **Número de sequências negativas** - número de ocorrências de runs abaixo da média em função da duração;
- **Número total de sequências negativas** - número total de runs abaixo da média em cada subsistema;
- **Duração média de sequências negativas** - duração média dos runs abaixo da média em cada subsistema;
- **Intensidade média de sequências negativas** - razão entre a soma dos valores de ENA associado aos runs abaixo da média e o número total de ocorrências de runs abaixo da média.

Os gráficos 4.17, 4.18, 4.19 e 4.20 expõem os fatos estilizados supracitados e evidenciam um ajuste adequado do modelo ao histórico.

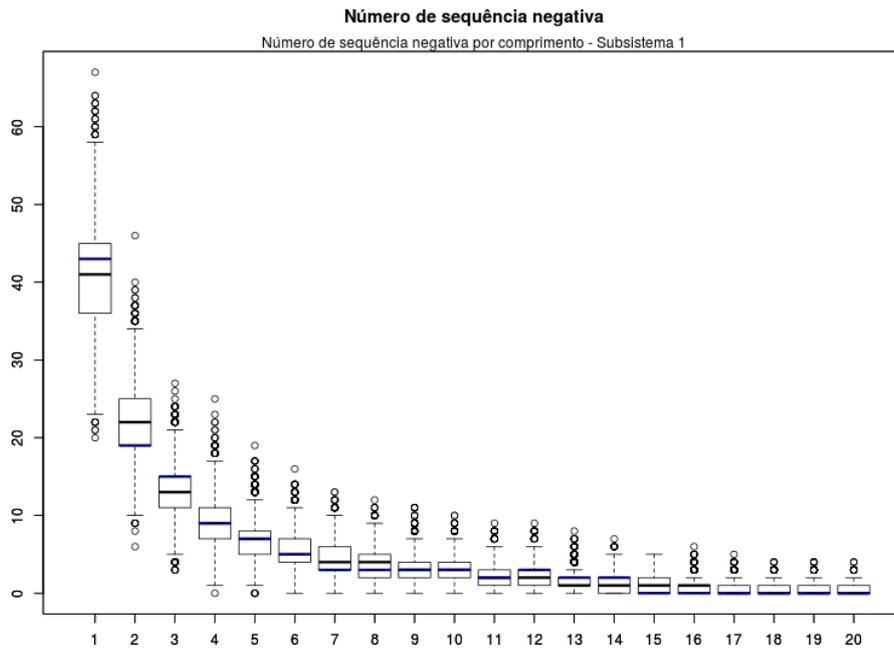


Figura 4.17: $PVAR_m$ – Número de seq. negativa - SE (1)
(mult. com restrição)

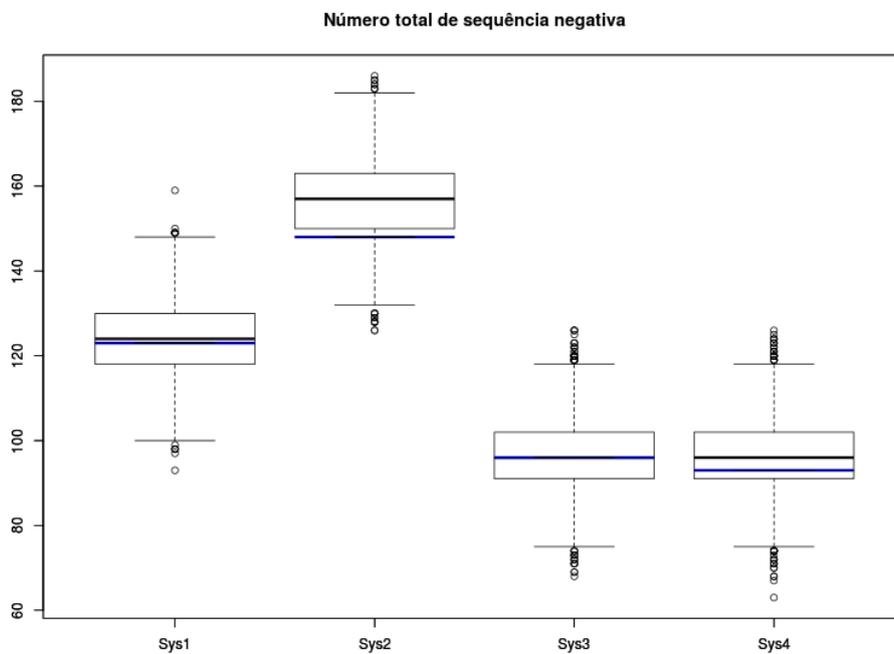


Figura 4.18: $PVAR_m$ – Número total de seq. negativas
(mult. com restrição)

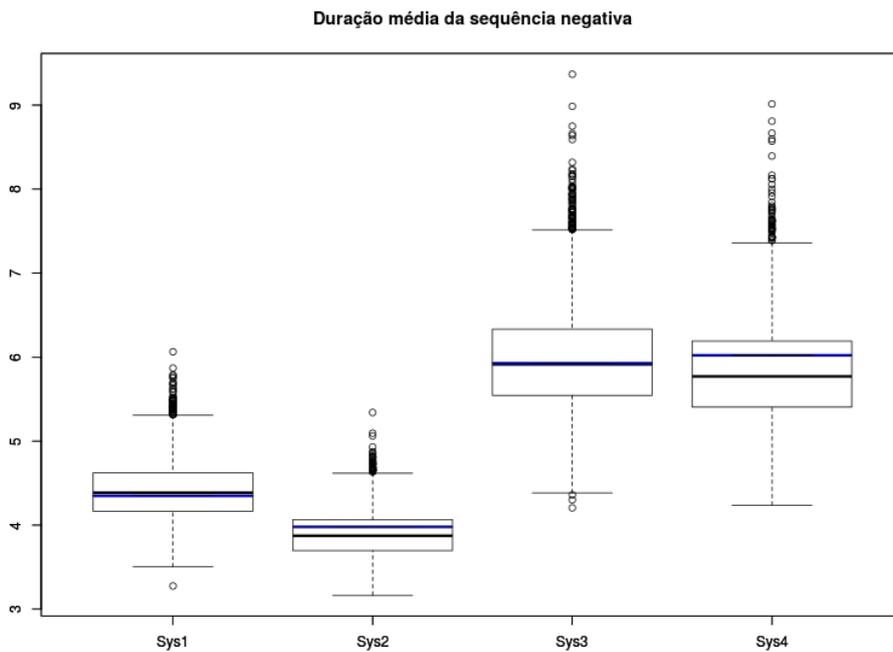


Figura 4.19: $PVAR_m$ – Duração média de seq. negativas (mult. com restrição)

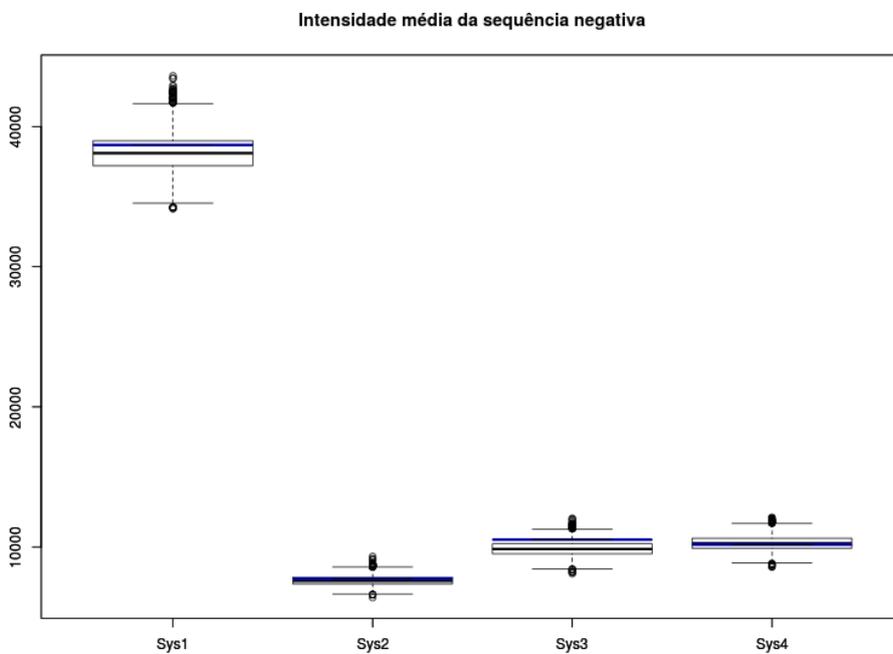


Figura 4.20: $PVAR_m$ – Intensidade média de seq. negativas (mult. com restrição)

4.5 Análise descritiva de somas parciais

A análise de somas parciais corresponde ao estudo de grandezas acumuladas. No caso de um aproveitamento hidrelétrico a grandeza de interesse é a afluência (ENA) descontada um fator associado a uma defluência. Neste contexto, um dos propósitos da análise de somas parciais é avaliar a dimensão de um possível reservatório que garanta uma dada defluência ao longo do tempo.

As definições e conceitos abordados utilizam a seguinte notação. Seja \mathbf{a}_t a ENA associada ao tempo t , $\boldsymbol{\mu}_t := \mathbb{E}[\mathbf{a}_t]$ o respectivo valor esperado e $\beta \in [0, 1]$ a proporção do valor esperado $\boldsymbol{\mu}_t$ que será defluída ao longo do tempo. Se a diferença $\mathbf{a}_t - \beta\boldsymbol{\mu}_t$ for positiva então este valor representa a energia que pode ser armazenada. Se a diferença $\mathbf{a}_t - \beta\boldsymbol{\mu}_t$ for negativa então a defluência energética deve ser garantida pela energia armazenada do reservatório. A *soma parcial* S_t associada ao fator de regularização β é o resultado líquido da operação de defluir $\beta\boldsymbol{\mu}_t$ em cada tempo t :

$$S_t = \sum_{j=1}^t (\mathbf{a}_j - \beta\boldsymbol{\mu}_j).$$

Neste caso, a grandeza S_t corresponde a energia armazenada no tempo t .

A análise gráfica exposta a seguir é análoga a realizada na seção 4.2 e 4.4 para as estatísticas descritivas básicas e análise descritiva de runs. Abaixo são apresentadas as seguintes estatísticas:

- **Comprimento do período crítico** - comprimento do intervalo de tempo que compreende a maior diferença entre um máximo local e um mínimo local posterior da série de somas parciais $\{S_1, S_2, \dots\}$. Este intervalo de tempo é chamado de *período crítico*. O período crítico de uma série $\{\mathbf{a}_1, \mathbf{a}_2, \dots\}$ é interpretado como o período de maior esvaziamento de um dado reservatório;
- **Capacidade associada ao período crítico** - diferença entre o máximo e mínimo valor de soma parcial do período crítico. Esta quantidade é interpretada como a capacidade necessária de um possível reservatório para garantir a defluência $\beta\boldsymbol{\mu}_j$ durante o período crítico de uma dada série de afluências;
- **Energia Natural Afluente média do período crítico** - soma das afluências associadas ao período crítico dividido pelo respectivo comprimento.

Observa-se que para baixos fatores de regularização β o comprimento do período crítico é zero. Neste caso, é tão pouca a energia defluída que os reservatórios sempre armazenam energia. Consequentemente, a capacidade associada ao período crítico e a energia natural afluyente média do período crítico também são nulas. Conforme o

valor do fator de regularização aumenta tende a aumentar também o período crítico e a capacidade associada ao período crítico, pois uma quantidade maior de energia é defluída, o que requer afluências maiores para a restituição desta energia.

As figuras 4.21, 4.22 e 4.23 exibem as distribuições dos valores simulados de comprimento, capacidade e energia afluente média do período crítico, respectivamente. É importante ressaltar que as correspondentes distribuições das estatísticas simuladas estão moderadamente deslocadas para baixo em relação aos respectivos valores históricos. Isto é um outro indicador de que os cenários de ENA simulados são mais otimistas para as estatísticas consideradas do que os cenários históricos de ENA.

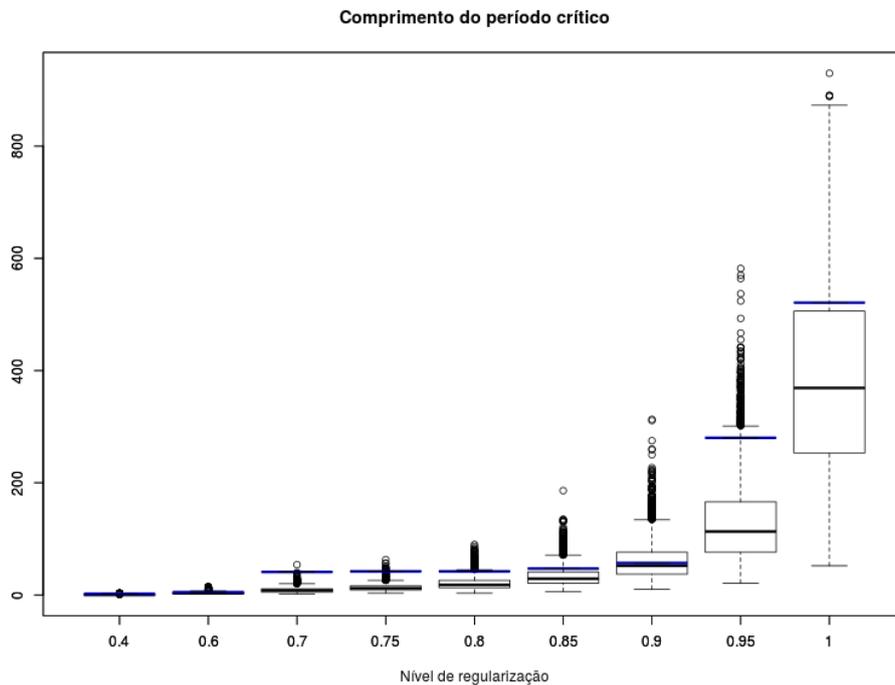


Figura 4.21: Comprimento do período crítico - SE (1)
(mult. com restrição)

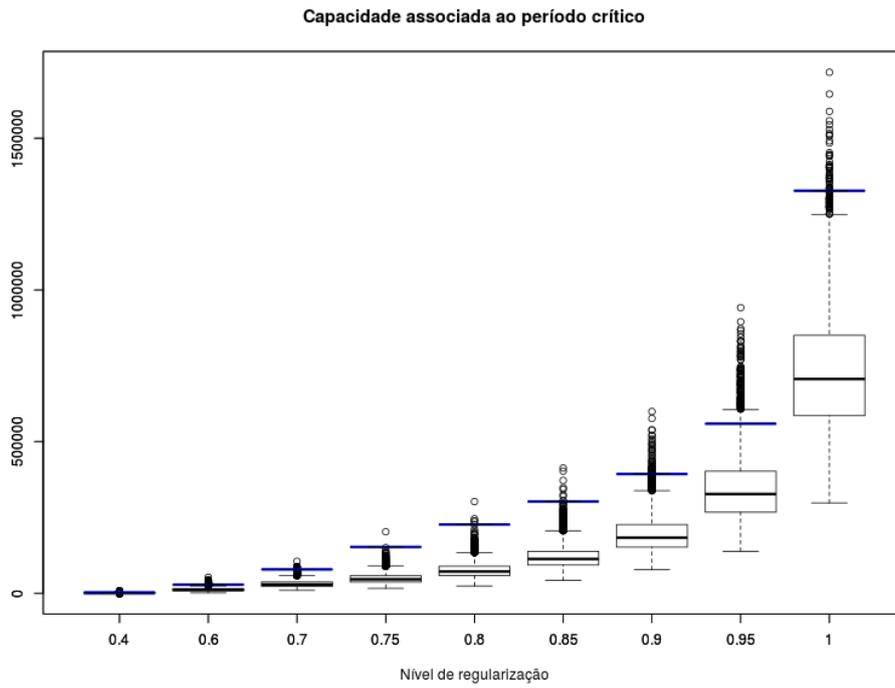


Figura 4.22: Capacidade de regularização - SE (1)
(mult. com restrição)

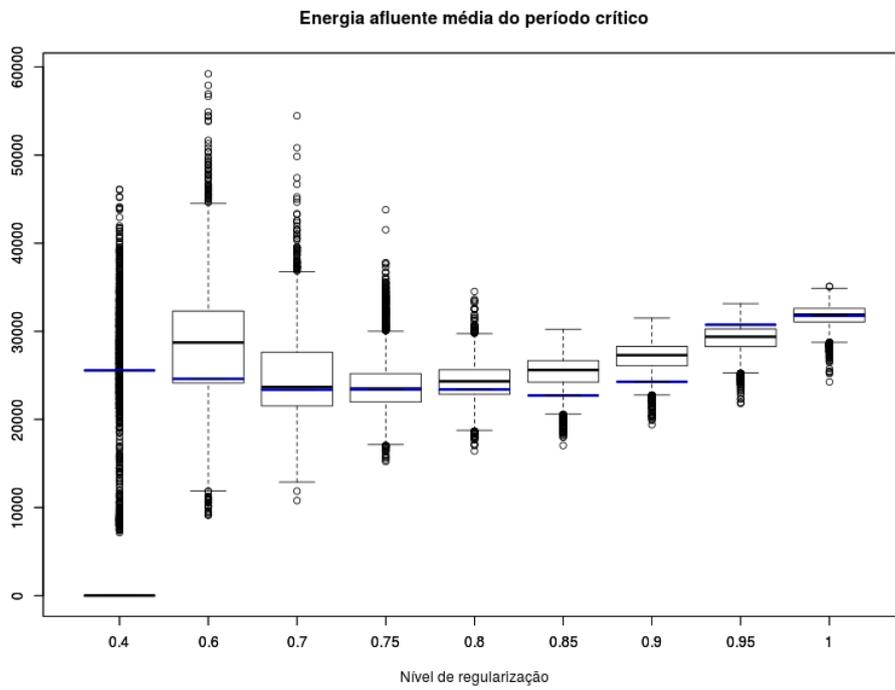


Figura 4.23: Energia afluente média - SE (1)
(mult. com restrição)

Capítulo 5

Conclusões

Como se sabe, cenários gerados a partir de modelos PVAR com ruídos aditivos podem apresentar afluências negativas, o que não é consistente com a realidade. Esta dissertação apresenta a metodologia do modelo Periódico Vetorial Autorregressivo Multiplicativo (PVAR_m) que considera *restrições de positividade* nos coeficientes. Este modelo, por construção, gera apenas valores positivos. Além disso, o modelo PVAR_m não introduz nenhuma artificialidade na distribuição dos ruídos, como a imposição de distribuição específica para evitar a geração de valores negativos, procedimento por vezes utilizado no PVAR.

Como apresentado na seção 4, o modelo PVAR_m proposto reproduz satisfatoriamente os valores de média e desvio padrão históricos assim como o modelo PVAR aditivo tradicional (erro normal e sem restrições nos coeficientes). Apenas a assimetria e a curtose do caso multiplicativo possuem um número considerável de outliers quando comparado com o modelo aditivo.

A seção 4 enfatiza que apenas a imposição de restrição de positividade nos coeficientes do modelo PVAR aditivo não é suficiente para garantir cenários de afluência positivos, o que realça a importância de considerar erros multiplicativos.

Ressalta-se que o modelo PVAR_m é um modelo multivariado no qual a correlação espacial é explicada não apenas pelos termos de erro contemporâneos, mas também pelos termos da própria série defasados no tempo e em todos os subsistemas. Esta maior abrangência pode permitir ao modelo PVAR_m uma capacidade preditiva melhor do que um correspondente modelo univariado.

Do ponto de vista de otimização estocástica, o modelo PVAR_m proposto apresenta duas características fundamentais para uso no problema de otimização associado ao planejamento da operação. A primeira é que a equação do modelo PVAR_m relaciona as afluências de maneira linear, exigência da formulação do problema de otimização. A segunda é a independência temporal dos ruídos multiplicativos (o ruído de um dado tempo é estocasticamente independente dos ruídos passados), exigência do SDDP, algoritmo empregado para resolver o problema

de otimização.

Isto posto, sugere-se a avaliação do uso do modelo $PVAR_m$ no problema de planejamento de longo prazo que suporta as operações do SIN.

5.1 Trabalhos futuros

Os seguintes tópicos são direções importantes de investigação do modelo $PVAR_m$ recomendadas como trabalhos futuros:

- **Condição de causalidade para o caso multiplicativo.** A causalidade do modelo com erros multiplicativo é uma conjectura sobre a possibilidade de, sob certas condições nos coeficientes do modelo, se expressar a afluência de um dado tempo t em função dos erros multiplicativos passados até t , vide seção 3.1. Caso o modelo $PVAR_m$ com erros multiplicativos independentes seja *causal* valem as seguintes propriedades:
 - o valor esperado condicional é igual à previsão;
 - a variância condicional (volatilidade) é proporcional ao produto cruzado da previsão.

Esta última propriedade diferencia o modelo $PVAR_m$ multiplicativo do modelo $PVAR$ aditivo, no qual a variância condicional é constante sazonalmente, isto é, não depende da previsão.

- **Critério BIC para seleção da ordem.** O índice BIC usado para a seleção da ordem do modelo $PVAR_m$ foi um proxy baseado no índice de um correspondente modelo de erro aditivo gaussiano sem restrição nos coeficientes. Neste sentido, os seguintes estudos podem ser desenvolvidos:
 1. o índice BIC para o caso em que o espaço paramétrico é \mathbb{R}_+^n ;
 2. a expressão do logaritmo da verossimilhança.
- **Sensibilidade do ajuste do modelo.** Analisar a variabilidade da ordem dos modelos e dos valores estimados para os parâmetros em função do tamanho da amostra considerada, por exemplo utilizando a técnica de Bootstrap.
- **Métodos de reamostragem dos resíduos:** Investigar outras técnicas de reamostragem que consideram que os resíduos são obtidos de uma série temporal e não de uma sequência de variáveis aleatórias i.i.d.'s (LAHIRI [36]).

Referências Bibliográficas

- [1] “Operador Nacional do Sistema Elétrico – ONS”. Disponível em: <www.ons.org.br>.
- [2] ARVANITIDIS, N. V., ROSING, J., 1970, “Optimal operation of multireservoir systems using a composite representation”, *Power Apparatus and Systems, IEEE Transactions on*, , n. 2, pp. 327–335.
- [3] BIRGE, J. R., LOUVEAUX, F. V., 1997, *Introduction to Stochastic Programming*. Springer Series in Operations Research and Financial Engineering. New York, Springer-Verlag. ISBN: 0-387-98217-5. doi: 10.1007/978-1-4614-0237-4.
- [4] BOSHPNAKOV, G. N., 1997, “Periodically correlated solutions to a class of stochastic difference equations”. In: *Stochastic differential and difference equations*, Springer, pp. 1–9.
- [5] BRAGA, D. B. M., 2011, *Processos Periódicos Autorregressivos Gama: Uma Contribuição para Modelagem da Energia Natural Afluente com Possíveis Usos na Determinação da Função de Custo Futuro*. Tese de Doutorado, IE-UFRJ.
- [6] BROCKWELL, P. J., DAVIS, R. A., 2006, *Introduction to time series and forecasting*. Springer Science & Business Media.
- [7] BROWNLEES, C. T., CIPOLLINI, F., GALLO, G. M., 2011, “Multiplicative error models”, *Available at SSRN 1852285*.
- [8] CEPEL, 2001, *Planejamento da Operação a Médio Prazo – Modelo NEWAVE – Manual de Referência*. Technical Report CEPEL DPP/PEN-107/1995 YYY, CEPEL, August.
- [9] CEPEL, 2006, *Manual de Referência do Modelo NEWAVE*. Rio de Janeiro, RJ.
- [10] CIPOLLINI, F., GALLO, G. M., 2010, “Automated variable selection in vector multiplicative error models”, *Computational Statistics & Data Analysis*, v. 54, n. 11, pp. 2470–2486.

- [11] CIPRA, T., 1985, “Periodic moving average process”, *Aplikace matematiky*, v. 30, n. 3, pp. 218–229.
- [12] CIPRA, T., 1985, “Statistical analysis of multiple moving average processes using periodicity”, *Kybernetika*, v. 21, n. 5, pp. 335–345.
- [13] CIPRA, T., TLUSTÝ, P., 1987, “ESTIMATION IN MULTIPLE AUTOREGRESSIVE-MOVING AVERAGE MODELS USING PERIODICITY”, *Journal of Time Series Analysis*, v. 8, n. 3, pp. 293–300.
- [14] CROLEY II, T. E., RAO, K. N. R., 1997, *A manual for hydrologic time series deseasonalization and serial independence reduction*. Relatório técnico, Iowa Institute of Hydraulic Research, The University of Iowa, Iowa City, Iowa.
- [15] DE OLIVEIRA, W. L., PENNA, D. D. J., DAMÁZIO, J. M., et al., 2009, “Análise da Correlação Cruzada Através da Distribuição Lognormal Três Parâmetros”. In: *XVIII Simpósio Brasileiro de Recursos Hídricos*.
- [16] DEMPSTER, A. P., LAIRD, N. M., RUBIN, D. B., 1977, “Maximum likelihood from incomplete data via the EM algorithm”, *Journal of the royal statistical society. Series B (methodological)*, pp. 1–38.
- [17] DUNSMUIR, W., 1981, “Estimation of periodically varying means and standard deviations in time series data”, *Journal of Time Series Analysis*, v. 2, n. 3, pp. 129–153.
- [18] FERREIRA, P. G. C., 2013, *A estocasticidade associada ao Setor Elétrico Brasileiro e uma nova abordagem para a geração de aflúências via Modelos Periódicos Gama*. Tese de Doutorado, PUC-Rio.
- [19] FIERING, M. B., 1964, “Multivariate technique for synthetic hydrology”, *Journal of the Hydraulics Division*, v. 90, n. 5, pp. 43–60.
- [20] FINZI, G., TODINI, E., WALLIS, J., 1975, “Comment upon multivariate synthetic hydrology”, *Water Resources Research*, v. 11, n. 6, pp. 844–850.
- [21] FRANCHINI, M., TODINI, E., GIULIANO, G., et al., 1986, “A multivariate multiseasonal model for long term persistent hydrological series”. In: *Proceedings of the Fourth International Hydrology Symposium on Multivariate Analysis of Hydrologic Processes*.
- [22] FRANSES, P. H., PAAP, R., 2004, *Periodic time series models*. Oxford University Press.

- [23] GLADYSHEV, E. G., 1961, “Periodically correlated random sequences”, *Soviet Mathematics*.
- [24] GLADYSHEV, E., 1963, “Periodically and almost-periodically correlated random processes with a continuous time parameter”, *Theory of Probability & Its Applications*, v. 8, n. 2, pp. 173–177.
- [25] HANNAN, E. J., 2009, *Multiple time series*, v. 38. John Wiley & Sons.
- [26] HAUTSCH, N., 2008, “Capturing common components in high-frequency financial time series: A multivariate stochastic multiplicative error model”, *Journal of Economic Dynamics and Control*, v. 32, n. 12, pp. 3978–4015.
- [27] HAUTSCH, N., MALEC, P., SCHIENLE, M., 2014, “Capturing the zero: a new class of zero-augmented distributions and multiplicative error processes”, *Journal of Financial Econometrics*, v. 12, n. 1, pp. 89–121.
- [28] HIPEL, K. W., MCLEOD, A. I., 1994, *Time series modelling of water resources and environmental systems*. Elsevier.
- [29] JAHEL, A. M. A., 2013, *Generation of PAR Time Series Models Using Periodic Levinson-Durbin Algorithm*. Tese de Doutorado, The Islamic University of Gaza.
- [30] JIMENEZ, C., MCLEOD, A., HIPEL, K., 1989, “Kalman filter estimation for periodic autoregressive-moving average models”, *Stochastic Hydrology and Hydraulics*, v. 3, n. 3, pp. 227–240.
- [31] JONES, R. H., BRELSFORD, W. M., 1967, “Time series with periodic structure”, *Biometrika*, v. 54, n. 3-4, pp. 403–408.
- [32] KAHAN, J. P., 1974, “A method for maintaining cross and serial correlations and the coefficient of skewness under generation in a linear bivariate regression model”, *Water Resources Research*, v. 10, n. 6, pp. 1245–1248.
- [33] KELMAN, J., DAMAZIO, J., COSTA, J., 1985, “A multivariate synthetic daily streamflow generator”. In: *Fourth International Hydrology Symposium, Fort Collins, Colorado*.
- [34] KELMAN, J., 1980, “A stochastic model for daily streamflow”, *Journal of Hydrology*, v. 47, n. 3, pp. 235–249.
- [35] KUCZERA, G., 1987, “On maximum likelihood estimators for the multisite lag-one streamflow model: Complete and incomplete data cases”, *Water Resources Research*, v. 23, n. 4, pp. 641–645.

- [36] LAHIRI, S. N., 2013, *Resampling methods for dependent data*. Springer Science & Business Media.
- [37] LANNE, M., 2006, “A mixture multiplicative error model for realized volatility”, *Journal of Financial Econometrics*, v. 4, n. 4, pp. 594–616.
- [38] LAWRANCE, A., 1976, “A reconsideration of the Fiering two-station model”, *Journal of Hydrology*, v. 29, n. 1, pp. 77–85.
- [39] MAASS, A., HUFSCHMIDT, M. M., DORFMAN, R., et al., 1962, *Design of water-resource systems*. Harvard University Press Cambridge, Massachusetts.
- [40] MAGALHÃES, M. N., DE LIMA, A. C. P., 2000, *Noções de probabilidade e estatística*. IME-USP.
- [41] MATALAS, N., WALLIS, J., 1971, “Statistical properties of multivariate fractional noise process”, *Water Resour Res*, v. 7, pp. 1460–1468.
- [42] MATALAS, N. C., 1967, “Mathematical assessment of synthetic hydrology”, *Water Resources Research*, v. 3, n. 4, pp. 937–945.
- [43] MCLEOD, A. I., HIPEL, K. W., 1978, *Developments in monthly autoregressive modelling*. Relatório técnico, Department of Systems Design Engineering, University of Waterloo, Waterloo, Ontario.
- [44] MCLEOD, A., 1995, “Diagnostic Checking Periodic Autoregression Models With Application”, *The Journal of Time Series Analysis*.
- [45] MEJÍA, J. M., RODRÍGUEZ-ITURBE, I., CORDOVA, J. R., 1974, “Multivariate generation of mixtures of normal and log normal variables”, *Water Resources Research*, v. 10, n. 4, pp. 691–693.
- [46] MINISTÉRIO DE MINAS E ENERGIA, MME, 2015. “Boletim Mensal de Monitoramento do Sistema Elétrico Brasileiro”. Outubro.
- [47] NASSERI, I., 1986, “Multistation stochastic model of hourly rainfall”. In: *Proceedings of the Fourth International Hydrology Symposium*.
- [48] PAGANO, M., 1978, “On periodic and multiple autoregressions”, *The Annals of Statistics*, pp. 1310–1317.
- [49] PEGRAM, G., JAMES, W., 1972, “Multilag multivariate autoregressive model for the generation of operational hydrology”, *Water Resources Research*, v. 8, n. 4, pp. 1074–1076.

- [50] PEREIRA, M. V. F., PINTO, L. M. V. G., 1991, “Multi-stage stochastic optimization applied to energy planning”, *Math. Programming*, v. 52, n. 2, Ser. B, pp. 359–375. ISSN: 0025-5610.
- [51] PEREIRA, M. V. F., 1985, “Optimal scheduling of hydrothermal systems—an overview”. In: da Cunha, S. H. F. (Ed.), *Proceedings of IFAC Symposium on Planning and Operation of Electrical Energy Systems*, v. 1, pp. 1–9, Rio de Janeiro, Brazil.
- [52] PHILPOTT, A. B., DE MATOS, V. L., 2012, “Dynamic sampling algorithms for multi-stage stochastic programs with risk aversion”, *European Journal of Operational Research*, v. 218, n. 2, pp. 470–483.
- [53] RASMUSSEN, P. F., SALAS, J. D., FAGHERAZZI, L., et al., 1996, “Estimation and validation of contemporaneous PARMA models for streamflow simulation”, *Water Resources Research*, v. 32, n. 10, pp. 3151–3160.
- [54] RICHARDSON, C. W., 1981, “Stochastic simulation of daily precipitation, temperature, and solar radiation”, *Water Resources Research*, v. 17, n. 1, pp. 182–190.
- [55] ROCKAFELLAR, R. T., WETS, R. J.-B., 2009, *Variational analysis*, v. 317. Springer Science & Business Media.
- [56] RUSZCZYŃSKI, A., 2003, “Decomposition Methods”. In: Ruszczyński, A., Shapiro, A. (Eds.), *Stochastic Programming*, v. 10, *Handbooks in Operations Research and Management Science*, 1 ed., Elsevier Science B.V., cap. 3, pp. 141–211, Amsterdam, The Netherlands. ISBN: 0 444 50854 6.
- [57] SAKAI, H., 1982, “Circular lattice filtering using Pagano’s method”, *Acoustics, Speech and Signal Processing, IEEE Transactions on*, v. 30, n. 2, pp. 279–287.
- [58] SALAS, J., PEGRAM, G., 1977, “A seasonal multivariate multilag autoregressive model in hydrology”, *Modeling Hydrologic Processes*, pp. 45–60.
- [59] SALAS, J. D., TABIOS, G. Q., BARTOLINI, P., 1985, “Approaches to multivariate modeling of water resources time series1”, *JAWRA Journal of the American Water Resources Association*, v. 21, n. 4, pp. 683–708.
- [60] SHAPIRO, A., 2011. “Topics in Stochastic Programming”. CORE Lecture Series, . Universite Catholique de Louvain.

- [61] SHAPIRO, A., DENTCHEVA, D., RUSZCZYŃSKI, A., 2009, *Lectures on Stochastic Programming: Modeling and Theory*. SIAM.
- [62] SHAPIRO, A., 2011, “Analysis of stochastic dual dynamic programming method”, *European J. Oper. Res.*, v. 209, n. 1, pp. 63–72. ISSN: 0377-2217. doi: 10.1016/j.ejor.2010.08.007. Disponível em: <<http://dx.doi.org/10.1016/j.ejor.2010.08.007>>.
- [63] SHAPIRO, A., 2011, “Analysis of stochastic dual dynamic programming method”, *European Journal of Operational Research*, v. 209, n. 1, pp. 63–72.
- [64] SHAPIRO, A., CABRAL, F. G., DA COSTA, J. P., 2015, *Report for technical cooperation between Georgia Institute of Technology and ONS - Operador Nacional do Sistema Eletrico*. Relatório técnico, Georgia Institute of Technology and ONS - Operador Nacional do Sistema Eletrico, October.
- [65] SILVERMAN, B. W., 1986, *Density estimation for statistics and data analysis*, v. 26. CRC press.
- [66] SRIKANTHAN, R., 1986, “Stochastic simulation of daily climatic data”. In: *Proceedings of the Fourth International Hydrology Symposium*.
- [67] STEDINGER, J. R., 1981, “Estimating correlations in multivariate streamflow models”, *Water Resources Research*, v. 17, n. 1, pp. 200–208.
- [68] STEDINGER, J. R., TAYLOR, M. R., 1982, “Synthetic streamflow generation: 1. Model verification and validation”, *Water resources research*, v. 18, n. 4, pp. 909–918.
- [69] TAO, P.-C., DELLEUR, J. W., 1976, “Seasonal and nonseasonal ARMA models in hydrology”, *Journal of the Hydraulics Division*, v. 102, n. 10, pp. 1541–1559.
- [70] TERRY, L. A., PEREIRA, M. V. F., NETO, T. A. A., et al., 1986, “Coordinating the energy generation of the Brazilian national hydrothermal electrical generating system.” *INTERFACES*, v. 16, n. 1, pp. 16–38.
- [71] THOMPSTONE, R. M., HIPEL, K. W., MCLEOD, A. I., 1985, “Grouping of periodic autoregressive models”, *Time series analysis: Theory and practice*, v. 6, pp. 35–49.
- [72] TIAO, G., GRUPE, M., 1980, “Hidden periodic autoregressive-moving average models in time series data”, *Biometrika*, v. 67, n. 2, pp. 365–373.

- [73] TROUTMAN, B. M., 1979, "Some results in periodic autoregression", *Biometrika*, v. 66, n. 2, pp. 219–228.
- [74] VECCHIA, A., 1985, "Maximum likelihood estimation for periodic autoregressive moving average models", *Technometrics*, v. 27, n. 4, pp. 375–384.
- [75] VECCHIA, A., 1985, "PERIODIC AUTOREGRESSIVE-MOVING AVERAGE (PARMA) MODELING WITH APPLICATIONS TO WATER RESOURCES¹", *JAWRA Journal of the American Water Resources Association*, v. 21, n. 5, pp. 721–730.
- [76] YOUNG, G. K., PISANO, W. C., 1968, "Operational hydrology using residuals", *Journal of the Hydraulics Division*, v. 94, n. 4, pp. 909–924.

Apêndice A

Problema do planejamento de longo prazo da operação

O sistema elétrico interligado brasileiro é um sistema de grande porte projetado e construído considerando a utilização integrada da geração e transmissão de recursos de todos os agentes e o uso inter-regional de intercâmbio energético, de modo a alcançar a redução de custos e a confiabilidade do suprimento de energia. As linhas de transmissão inter-regionais e interbacias permitem um intercâmbio de grandes blocos de energia entre áreas, tornando possível aproveitar a diversidade hidrológica entre bacias hidrográficas. Em outubro de 2015, o parque gerador de energia elétrica brasileiro dispunha de uma capacidade instalada total de 139.272 MW, onde a capacidade instalada hidráulica, térmica e eólica contribuíam com 65,1%, 30,1% e 4,8%, respectivamente. A tabela (A.1) contém os dados do Boletim de Monitoramento do Sistema Elétrico (MINISTÉRIO DE MINAS E ENERGIA, MME [46]) apresentando a capacidade instalada total em outubro de 2015 discriminada por tipo fonte energética. (ONS [1])

Fonte	N. de usinas	Cap. instalada (MW)	(%) Cap. instalada
Hidráulica	1.210	90.620	65,1%
Térmica	2.887	41.951	30,1%
Gás Natural	146	12.917	9,3%
Biomassa	511	13.172	9,4%
Petróleo	2.174	10.105	7,3%
Carvão	23	3.614	2,6%
Nuclear	2	1.990	1,4%
Outros	31	153	0,1%
Eólica	275	6.680	4,8%
Solar	26	21	0,0%
Capacidade Total	4.398	139.272	100,0%

Tabela A.1: Matriz de capacidade instalada de geração de energia elétrica do Brasil.

A.1 Motivação

O propósito do planejamento da operação de sistemas hidrotérmicos é definir uma estratégia de operação que, para cada estágio do período de planejamento, dado o estado do sistema no início do estágio, produz uma meta para cada usina. O objetivo usual é minimizar o valor esperado do custo total ao longo do horizonte de planejamento, de modo a atender os requisitos de continuidade no suprimento energético sujeito a restrições de viabilidade. Os custos de operação compreendem custos de combustível, compras de energia de sistemas vizinhos e penalidades por falhas no atendimento à demanda.

O planejamento da operação hidrotérmica pode ser visto como um problema de decisão sob incerteza, pois são desconhecidas variáveis como aflúncias futuras, demanda, custos de combustível e disponibilidade das instalações. A existência de reservatórios com capacidade de regularização plurianual faz com que o problema do planejamento da operação seja um problema de otimização multiestágio. No Brasil é usual considerar um horizonte de planejamento de 5 anos discretizado em bases mensais. A existência de múltiplas interconexões entre hidrelétricas e restrições de transmissão caracterizam o problema como de grande porte. Além disso, como o valor da energia gerada em usinas hidrelétricas não pode ser medido diretamente como uma função da energia presente gerada, mas sim em termos do custo médio da geração térmica futura evitada, a função objetivo é também não separável (PEREIRA e PINTO [50]).

Em suma, o problema de planejamento da operação hidrotérmica brasileira é um problema de otimização estocástica multiestágio, de grande porte, não linear e não separável. Tal configuração excede em muito a capacidade computacional para se resolver tal problema com uma precisão adequada em um intervalo de tempo razoável. A abordagem padrão para se resolver este problema consiste em recorrer a uma cadeia de modelos abrangendo horizontes de planejamento de longo, médio e curto prazo PEREIRA [51]. Neste documento será apresentado apenas o problema de planejamento de longo prazo.

Para o longo prazo, é usual considerar uma representação aproximada do sistema proposta por ARVANITIDIS e ROSING [2] que agrega todas as usinas hidrelétricas pertencentes a uma mesma região hidrológica homogênea em um único reservatório equivalente de energia, reduzindo assim a complexidade do problema. A capacidade de armazenamento de energia de um reservatório equivalente de energia pode ser estimada como a energia que pode ser produzida ao se deplecionar os reservatórios de um sistema, fornecida uma regra simplificada de operação que aproxima a política de deplecionamento usualmente empregada. Uma descrição detalhada do modelo de sistema agregado aplicado ao sistema elétrico brasileiro pode ser encontrada em

TERRY *et al.* [70].

Para o sistema interligado brasileiro é usual considerar quatro reservatórios equivalentes de energia correspondentes a cada uma das quatro principais regiões interconectadas: SE, S, N, NE e usar a abordagem PDDE para obter as funções de custo futuro para cada estágio do horizonte de planejamento. A política resultante obtida com a representação agregada fornece a condição de fronteira para o problema de médio prazo que, ao considerar a representação individual de cada usina hidrelétrica e termelétrica considerando um horizonte de planejamento de alguns meses, permite obter decisões para cada usina geradora. Essa é a abordagem utilizada atualmente para resolver o planejamento energético de longo prazo no sistema elétrico interligado brasileiro.

A.2 Formulação multi-estágio determinística

Um problema de decisão multi-estágio é um problema cujas decisões são tomadas ao longo do tempo. Esta abordagem se torna importante quando é fundamental descrever a interferência de uma decisão passada numa decisão futura, tanto em termos de custos como em termos de restrições às novas decisões. Uma formulação deste problema é dada via modelo de otimização, isto é, um problema que é basicamente composto por decisões, restrições que caracterizam as possíveis decisões e o custo associado a cada decisão.

O propósito desta seção é descrever três formulações do problema de otimização estocástica multi-estágio: por observações, aninhada e programação dinâmica. Descreve-se inicialmente o problema multi-estágio determinístico, no qual todos os parâmetros do modelo são conhecidos.

As hipóteses adotadas neste documento são: as decisões podem ser representadas por vetores de \mathbb{R}^n , as restrições são representadas por um conjunto de equações e inequações lineares e o custo é representado por uma função linear das decisões. Esta classe de problemas é chamada classe de problemas lineares contínuos. Exemplos de outras hipóteses de modelagem são descritas em BIRGE e LOUVEAUX [3].

Em termos de notação, denota-se por $T \in \mathbb{Z}_+$ o horizonte de planejamento do modelo, por $t \in \{1, 2, \dots, T\}$ um determinado estágio do planejamento e por $x_t \in \mathbb{R}^{n_t}$ uma decisão do estágio t . O conjunto de restrições para as decisões do primeiro estágio é descrito por:

$$Ax_1 = b_1, \quad x_1 \geq 0$$

e para os demais estágios são descritos por:

$$B_t x_{t-1} + A_t x_t = b_t, \quad x_t \geq 0,$$

onde $B_t x_{t-1}$ é a influência da decisão do estágio anterior na decisão do estágio corrente. O custo total das decisões ao longo do horizonte de planejamento é:

$$c_1^\top x_1 + \cdots + c_T^\top x_T$$

Supondo conhecidos todos os parâmetros do modelo, é possível formular o problema de otimização *determinística* multi-estágio como

$$\begin{aligned} \min_{x_1, x_2, \dots, x_T} \quad & c_1^\top x_1 + c_2^\top x_2 + \cdots + c_T^\top x_T \\ \text{s.a.} \quad & A_1 x_1 = b_1 \\ & B_2 x_1 + A_2 x_2 = b_2 \\ & \quad \quad \quad \ddots = \vdots \\ & B_T x_{T-1} + A_T x_T = b_T \\ & x_1 \geq 0, \quad x_2 \geq 0, \quad \cdots \quad x_T \geq 0 \end{aligned} \tag{A.1}$$

onde cada parâmetro de (A.1) pode ser interpretado da seguinte forma:

- $c_t \in \mathbb{R}^{n_t}$: vetor de custo unitário da decisão do estágio t ;
- $b_t \in \mathbb{R}^{m_t}$: vetor de recurso do estágio t ;
- $A_t \in \mathbb{R}^{m_t \times n_t}$: influência unitária da decisão do estágio t no recurso disponível b_t ;
- $B_t \in \mathbb{R}^{m_t \times n_{t-1}}$: influência unitária da decisão do estágio $t - 1$ no recurso disponível b_t .

Esta apresentação das restrições encontra-se na chamada “forma padrão” (SHAPIRO *et al.* [61]) e todo problema linear pode ser reescrito desta forma. É interessante observar que a influência das decisões mais antigas do que as do estágio anterior nas decisões correntes também poderiam ser consideradas nesta formulação geral:

$$B_{tp} x_{t-p} + \cdots + B_{t1} x_{t-1} + A_t x_t = b_t, \quad x_t \geq 0,$$

mas por questões de simplicidade da notação não será considerado nesta apresentação.

Notação e revisão de probabilidade

Antes de prosseguir com o caso estocástico, é fundamental descrever a notação de probabilidade utilizada neste capítulo bem como uma breve revisão do conceito associado:

1. Ω é o espaço amostral do experimento aleatório, isto é, Ω é o espaço de todos os possíveis resultados de tal experimento;

2. \mathcal{F} é a σ -álgebra de Ω . Em termos intuitivos, a σ -álgebra \mathcal{F} é o conjunto formado por subconjuntos do espaço amostral Ω para os quais é permitido associar probabilidades. As propriedades que definem uma σ -álgebra são:

$$(i) \Omega, \emptyset \in \mathcal{F}, \quad (ii) D \in \mathcal{F} \Rightarrow D^c \in \mathcal{F}, \quad \text{e} \quad (iii) \{D_i\}_{i=1}^{\infty} \subset \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} D_i \in \mathcal{F};$$

3. $C \in \mathcal{F}$ é chamado de *evento*;

4. \mathcal{B} é a σ -álgebra de Borel de \mathbb{R}^d . A σ -álgebra de Borel \mathcal{B} é uma σ -álgebra gerada pelos abertos de \mathbb{R}^d , isto é, é a menor σ -álgebra de \mathbb{R}^d que contém os abertos de \mathbb{R}^d . Em termos intuitivos, esta σ -álgebra relaciona o conceito de probabilidade e a ideia de limite, derivada e integral;

5. $\mathbb{P}: \mathcal{F} \rightarrow [0, 1]$ é a função medida de probabilidade que associa cada evento $C \in \mathcal{F}$ a um número entre 0 e 1. As propriedades que definem uma medida de probabilidade \mathbb{P} são:

$$(i) \mathbb{P}(\emptyset) = 0, \quad (ii) \mathbb{P}(\Omega) = 1 \quad \text{e} \quad (iii) \mathbb{P}\left(\biguplus_{i=1}^{\infty} C_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(C_i),$$

onde C_i e C_j são eventos disjuntos, ou seja, $C_i \cap C_j = \emptyset$, se $i \neq j$.

6. $\xi: \Omega \rightarrow \mathbb{R}^d$ é uma variável aleatória que associa o resultado $\omega \in \Omega$ do experimento aleatório a um vetor em \mathbb{R}^d . Para todo elemento D da σ -álgebra de Borel \mathcal{B} a pré-imagem por ξ , $\xi^{-1}(D) = \{\omega \in \Omega \mid \xi(\omega) \in D\}$, deve ser um elemento da σ -álgebra \mathcal{F} . Este requisito é necessário para dar sentido à ideia de probabilidade dos valores de \mathbb{R}^d que estão no contradomínio da variável aleatória ξ ;

7. Ξ é o suporte da variável aleatória ξ . O suporte Ξ de uma variável aleatória ξ é o menor conjunto fechado de \mathbb{R}^d que contém ξ com probabilidade 1, ou seja,

$$\Xi = \bigcap_{\substack{C \text{ é fechado} \\ \mathbb{P}(\xi \in C) = 1}} C,$$

onde $\xi \in C := \{\omega \in \Omega \mid \xi(\omega) \in C\}$;

8. A função distribuição acumulada de uma variável aleatória ξ é definida por

$$F(x_1, \dots, x_d) = \mathbb{P}(\xi_1 \leq x_1, \dots, \xi_d \leq x_d).$$

9. Diz-se que ξ é uma variável aleatória contínua se existe uma função $f : \mathbb{R}^d \rightarrow \mathbb{R}$ tal que

$$F(x_1, \dots, x_d) = \int_{-\infty}^{x_d} \cdots \int_{-\infty}^{x_1} f_\xi(x_1, \dots, x_d) dx_1 \dots dx_d.$$

A função f satisfaz às seguintes propriedades:

$$f_\xi(x_1, \dots, x_d) \geq 0 \quad \text{e} \quad \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_\xi(x_1, \dots, x_d) dx_1 \dots dx_d = 1$$

10. Diz-se que ξ é uma variável aleatória discreta se existe uma função $p : \Xi \rightarrow \mathbb{R}$, onde Ξ é um subconjunto discreto de \mathbb{R}^n , tal que

$$F(x_1, \dots, x_d) = \sum_{\substack{(u_1, \dots, u_d) \in \Xi \\ u_1 \leq x_1, \dots, u_d \leq x_d}} p(u_1, \dots, u_d).$$

A função p satisfaz às seguintes propriedades:

$$p(u_1, \dots, u_d) \geq 0 \quad \text{e} \quad \sum_{(u_1, \dots, u_d) \in \Xi} p(u_1, \dots, u_d) = 1.$$

11. $\mathbb{E}[\xi]$ é o valor esperado da variável aleatória ξ . Se ξ é uma variável aleatória contínua então

$$\mathbb{E}[\xi] := \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \vec{x} f_\xi(x_1, \dots, x_d) dx_1 \dots dx_d$$

e se ξ é uma variável aleatória discreta então

$$\mathbb{E}[\xi] := \sum_{(u_1, \dots, u_d) \in \Xi} \vec{x} p(x_1, \dots, x_d).$$

A.3 Formulações multi-estágio estocásticas

Existem situações nas quais alguns parâmetros do problema determinístico (A.1) não são conhecidos antecipadamente. Uma classe desses problemas é quando os parâmetros são assumidos como aleatórios e as respectivas distribuições de probabilidade podem ser estimadas. Este caso se enquadra nos chamados processos de decisão sob incerteza e a pergunta fundamental que guia a modelagem de tais problemas é:

- Quais informações estão disponíveis no momento que se deve tomar a decisão do estágio t ?

Denotando-se por $\xi_t \in \mathbb{R}^{d_t}$ o vetor aleatório associado aos parâmetros que são desconhecidos até o estágio $t - 1$ e observados apenas no estágio t , o processo de

decisão pode ser ilustrado pelo diagrama abaixo:

$$\begin{aligned} & \text{decisão}(x_1) \rightsquigarrow \text{observação}(\xi_2) \rightsquigarrow \text{decisão}(x_2) \rightsquigarrow \\ & \dots \rightsquigarrow \text{observação}(\xi_T) \rightsquigarrow \text{decisão}(x_T). \end{aligned}$$

Por esta definição, ξ_1 já foi observado e portanto é um vetor aleatório constante. Este documento analisa o caso em que c_t , A_t , B_t e b_t , parâmetros aleatórios do problema (A.1), são observados apenas no estágio t , ou seja, $\xi_t = (c_t, A_t, B_t, b_t)$. Este vetor de parâmetros ξ_t é obtido enfileirando o vetor c_t , as colunas das matrizes A_t e B_t e o vetor b_t .

A.3.1 Formulação por observações

Tendo em vista a natureza dos parâmetros, é importante discorrer sobre a representação das decisões. Pois, como se conhece a priori a distribuição das observações em $t + 1$, ξ_{t+1} , dada a realização das observações até t , $\xi_{[t]}$, é possível criar uma *regra* sobre quais decisões tomar em cada estágio t . Por exemplo, um gestor de um fundo de investimento com a opção de investir em renda variável ou em renda fixa pode criar uma regra sobre o quanto investir em cada aplicação de acordo com a evolução das respectivas taxas de retorno. Considerando este conceito de planejamento das decisões, define-se uma *regra de decisão* ou *política* como uma sequência de funções das observações:

$$\begin{aligned} x_t: \Xi_{[t]} &\rightarrow \mathbb{R}^{n_t} \\ \xi_{[t]} &\mapsto x_t(\xi_{[t]}), \quad t = 1, \dots, T \end{aligned}$$

onde $\Xi_{[t]}$ é o suporte do vetor aleatório $\xi_{[t]} := (\xi_1, \dots, \xi_t)$. Como as regras de decisão são funções das observações, batiza-se a formulação a ser enunciada como *formulação por observações*, pois não se encontrou na literatura um nome apropriado. Observa-se que as regras de decisão do estágio t só podem depender das observações até t . Esta restrição chama-se de *não-antecipatividade*. Além disso, uma regra de decisão $x_1(\cdot), \dots, x_T(\cdot)$ é *viável* se satisfaz a:

$$\begin{aligned} A_1(\xi_1)x_1(\xi_{[1]}) &= b_1(\xi_1) \\ B_2(\xi_2)x_1(\xi_{[1]}) + A_2(\xi_2)x_2(\xi_{[2]}) &= b_2(\xi_2) \\ &\vdots \\ B_T(\xi_T)x_{T-1}(\xi_{[T-1]}) + A_T(\xi_T)x_T(\xi_{[T]}) &= b_T(\xi_T) \\ x_1(\xi_{[1]}) \geq 0, \quad x_2(\xi_{[2]}) \geq 0, \quad \dots \quad x_T(\xi_{[T]}) \geq 0 \end{aligned} \tag{A.2}$$

A.3.2 Formulação aninhada

A formulação aninhada pode ser vista como uma maneira de estabelecer uma ponte entre a intuição criada pela formulação por observações e a descrição sucinta da formulação por programação dinâmica. Em geral, a formulação por programação dinâmica é a mais adequada para enunciar problemas reais.

Para se obter a formulação aninhada é necessário reescrever a formulação por observações (A.3) de maneira adequada. Com esta finalidade, denota-se por \mathcal{X}_t o subconjunto de \mathbb{R}^{n_t} gerado pelas restrições de (A.3) do estágio t :

$$\mathcal{X}_1(\xi_1) = \{x_1 \in \mathbb{R}^{n_1} \mid A_1(\xi_1)x_1 = b_1(\xi_1), x_1 \geq 0\} \quad (\text{A.4})$$

$$\mathcal{X}_t(x_{t-1}, \xi_t) = \{x_t \in \mathbb{R}^{n_t} \mid B_t(\xi_t)x_{t-1} + A_t(\xi_t)x_t = b_t(\xi_t), x_t \geq 0\}, \quad t = 2, \dots, T.$$

Com isso, a formulação por observações (A.3) pode ser descrita por

$$\begin{aligned} \min_{x_1(\cdot), x_2(\cdot), \dots, x_T(\cdot)} \quad & \mathbb{E} [c_1^\top(\xi_1)x_1(\xi_{[1]}) + c_2^\top(\xi_2)x_2(\xi_{[2]}) + \dots + c_T^\top(\xi_T)x_T(\xi_{[T]})] \\ & x_1(\xi_{[1]}) \in \mathcal{X}_1(\xi_1), x_t(\xi_{[t]}) \in \mathcal{X}_t(x_{t-1}(\xi_{[t-1]}), \xi_t), \quad t = 2, \dots, T. \end{aligned} \quad (\text{A.5})$$

Ainda no intuito de produzir uma formulação adequada, denota-se por \mathcal{M}_t o espaço das regras de decisão viáveis do estágio t gerado pelas restrições de (A.5):

$$\mathcal{M}_1 = \{x_1 : \Xi_1 \rightarrow \mathbb{R}^{n_1} \mid x_1(\xi_{[1]}) \in \mathcal{X}_1(\xi_1) \text{ q.c.}\}$$

$$\mathcal{M}_t(x_{t-1}(\cdot)) = \{x_t : \Xi_{[t]} \rightarrow \mathbb{R}^{n_t} \mid x_t(\xi_{[t]}) \in \mathcal{X}_t(x_{t-1}(\xi_{[t-1]}), \xi_t) \text{ q.c.}\}, \quad t = 2, \dots, T.$$

Com isso, a formulação (A.5) pode ser descrita como:

$$\begin{aligned} \min_{x_1(\cdot), x_2(\cdot), \dots, x_T(\cdot)} \quad & \mathbb{E} [c_1^\top(\xi_1)x_1(\xi_{[1]}) + c_2^\top(\xi_2)x_2(\xi_{[2]}) + \dots + c_T^\top(\xi_T)x_T(\xi_{[T]})] \\ & x_1(\cdot) \in \mathcal{M}_1, x_t(\cdot) \in \mathcal{M}_t(x_{t-1}(\cdot)), \quad t = 2, \dots, T. \end{aligned} \quad (\text{A.6})$$

Um princípio necessário para a construção da formulação aninhada é a *permutabilidade* entre a esperança e o mínimo. Sejam $\mathcal{X}(\xi) \subseteq \mathbb{R}^n$ um conjunto que representa as decisões viáveis para cada observação ξ e

$$\mathcal{M} = \{x : \Xi \rightarrow \mathbb{R}^n \mid x(\xi) \in \mathcal{X}(\xi), \text{ q.c.}\}$$

o conjunto das regras de decisão viáveis. Então, vale a seguinte relação:

$$\min_{x(\cdot) \in \mathcal{M}} \mathbb{E} [F(x(\xi), \xi)] = \mathbb{E} \left[\min_{\tilde{x} \in \mathcal{X}(\xi)} F(\tilde{x}, \xi) \right], \quad (\text{A.7})$$

ou seja, o *mínimo da esperança* considerando o espaço das regras de decisão viáveis

(funções de \mathcal{M}) é igual à *esperança do mínimo* considerando o espaço das decisões viáveis (variáveis de $\mathcal{X}(\xi) \subseteq \mathbb{R}^n$) SHAPIRO [60]. Além disso, as soluções ótimas de ambos os problemas se relacionam da seguinte forma: $x^*(\cdot)$ é uma regra de decisão ótima, isto é,

$$x^*(\cdot) \in \arg \min_{x(\cdot) \in \mathcal{M}} \mathbb{E} [F(x(\xi), \xi)] \quad (\text{A.8})$$

se, e somente, para cada observação $\xi(\omega)$ a decisão correspondente $x^*(\xi(\omega))$ é ótima, isto é,

$$x^*(\xi(\omega)) \in \arg \min_{\tilde{x} \in \mathcal{X}(\xi)} F(\tilde{x}, \xi), \quad \forall \omega \in \Omega, \text{ q.c..} \quad (\text{A.9})$$

Este princípio é válido sob condições bem gerais, como descrito em SHAPIRO [60].

No caso do problema (A.6), o princípio da permutabilidade entre a esperança e o mínimo (A.7) é aplicado T vezes:

$$\begin{aligned} & \min_{\substack{x_1(\cdot) \in \mathcal{M}_1 \\ x_t(\cdot) \in \mathcal{M}_t(x_{t-1}(\cdot)) \\ t=2, \dots, T}} \mathbb{E} [c_1^\top(\xi_1)x_1(\xi_{[1]}) + \dots + c_T^\top(\xi_T)x_T(\xi_{[T]})] = \\ &= \min_{\substack{x_1(\cdot) \in \mathcal{M}_1 \\ x_t(\cdot) \in \mathcal{M}_t(x_{t-1}(\cdot)) \\ t=2, \dots, T-1}} \mathbb{E} \left[\min_{\tilde{x}_T \in \mathcal{X}_T(x_{T-1}(\xi_{[T-1]}), \xi_T)} [c_1^\top(\xi_1)x_1(\xi_{[1]}) + \dots + c_T^\top(\xi_T)\tilde{x}_T] \right] \\ &= \dots = \mathbb{E} \left[\min_{\substack{\tilde{x}_1 \in \mathcal{X}_1(\xi_1) \\ \tilde{x}_t \in \mathcal{X}_t(\tilde{x}_{t-1}, \xi_t) \\ t=2, \dots, T}} [c_1^\top(\xi_1)\tilde{x}_1 + \dots + c_T^\top(\xi_T)\tilde{x}_T] \right] \end{aligned} \quad (\text{A.10})$$

Em seguida, aplica-se o princípio da decomposição dos mínimos (ROCKAFELLAR e WETS [55]) para o argumento da esperança (A.10):

$$\min_{\substack{\tilde{x}_1 \in \mathcal{X}_1(\xi_1) \\ \tilde{x}_t \in \mathcal{X}_t(\tilde{x}_{t-1}, \xi_t) \\ t=2, \dots, T}} [\cdot] = \min_{\tilde{x}_1 \in \mathcal{X}_1(\xi_1)} \left[\min_{\tilde{x}_2 \in \mathcal{X}_2(\tilde{x}_1, \xi_2)} \left[\dots \min_{\tilde{x}_T \in \mathcal{X}_T(\tilde{x}_{T-1}, \xi_T)} [\cdot] \right] \right].$$

Com isso, o mínimo da soma pode ser reescrito em sua forma aninhada:

$$\begin{aligned} & \min_{\substack{\tilde{x}_1 \in \mathcal{X}_1(\xi_1) \\ \tilde{x}_t \in \mathcal{X}_t(\tilde{x}_{t-1}, \xi_t) \\ t=2, \dots, T}} [c_1^\top(\xi_1)\tilde{x}_1 + \dots + c_T^\top(\xi_T)\tilde{x}_T] = \\ &= \min_{\tilde{x}_1 \in \mathcal{X}_1(\xi_1)} \left[c_1^\top(\xi_1)\tilde{x}_1 + \min_{\tilde{x}_2 \in \mathcal{X}_2(\tilde{x}_1, \xi_2)} \left[\dots + \min_{\tilde{x}_T \in \mathcal{X}_T(\tilde{x}_{T-1}, \xi_T)} [c_T^\top(\xi_T)\tilde{x}_T] \right] \right]. \end{aligned} \quad (\text{A.11})$$

Por último, a lei das expectativas iteradas (MAGALHÃES e DE LIMA [40])

$$\mathbb{E}[\cdot] = \mathbb{E}_{|\xi_{[1]}}[\mathbb{E}_{|\xi_{[2]}}[\dots \mathbb{E}_{|\xi_{[T-1]}}[\cdot]]]]$$

aplicada em (A.10) com o argumento (A.11) resulta em

$$\begin{aligned} \min_{\tilde{x}_1 \in \mathcal{X}_1(\xi_1)} c_1^\top(\xi_1)\tilde{x}_1 + \mathbb{E}_{|\xi_{[1]}} \left[\min_{\tilde{x}_2 \in \mathcal{X}_2(\tilde{x}_1, \xi_2)} c_2^\top(\xi_2)\tilde{x}_2 + \mathbb{E}_{|\xi_{[2]}} \left[\dots \right. \right. \\ \left. \left. + \mathbb{E}_{|\xi_{[T]}} \left[\min_{\tilde{x}_T \in \mathcal{X}_T(\tilde{x}_{T-1}, \xi_T)} c_T^\top(\xi_T)\tilde{x}_T \right] \right] \right]. \end{aligned} \quad (\text{A.12})$$

Substituindo a definição dos conjuntos \mathcal{X}_t em (A.12), para t desde 1 até T , obtém-se a *formulação aninhada*:

$$\begin{aligned} \min_{\substack{A_1(\xi_1)\tilde{x}_1=b_1(\xi_1) \\ \tilde{x}_1 \geq 0}} c_1^\top(\xi_1)\tilde{x}_1 + \mathbb{E}_{|\xi_{[1]}} \left[\min_{\substack{B_2(\xi_2)\tilde{x}_1+A_2(\xi_2)\tilde{x}_2=b_2(\xi_2) \\ \tilde{x}_2 \geq 0}} c_2^\top(\xi_2)\tilde{x}_2 + \mathbb{E}_{|\xi_{[2]}} \left[\dots \right. \right. \\ \left. \left. + \mathbb{E}_{|\xi_{[T]}} \left[\min_{\substack{B_T(\xi_T)\tilde{x}_{T-1}+A_T(\xi_T)\tilde{x}_T=b_T(\xi_T) \\ \tilde{x}_T \geq 0}} c_T^\top(\xi_T)\tilde{x}_T \right] \right] \right] \end{aligned} \quad (\text{A.13})$$

A apresentação usual da formulação aninhada omite a variável ξ de modo a produzir uma notação menos carregada:

$$\min_{\substack{A_1x_1=b_1 \\ x_1 \geq 0}} c_1^\top x_1 + \mathbb{E} \left[\min_{\substack{B_2x_1+A_2x_2=b_2 \\ x_2 \geq 0}} c_2^\top x_2 + \mathbb{E} \left[\dots + \mathbb{E} \left[\min_{\substack{B_Tx_{T-1}+A_Tx_T=b_T \\ x_T \geq 0}} c_T^\top x_T \right] \right] \right] \quad (\text{A.14})$$

Porém, para uma primeira leitura, omitir estes detalhes pode obscurecer o entendimento. Uma interpretação da formulação aninhada (A.14) é: a primeira decisão x_1 será aquela que minimiza o custo imediato $c_1^\top x_1$ somado ao custo médio das decisões influenciada por x_1 a partir do segundo estágio, considerando que no segundo estágio, ao observar ξ_2 , a decisão x_2 será aquela que minimiza o seu custo imediato $c_2^\top x_2$ somado ao custo médio das decisões influenciada por x_2 a partir do terceiro estágio, e assim por diante até que no último estágio T , ao observar ξ_T , a decisão x_T será aquela que minimiza o seu custo imediato $c_T^\top x_T$.

A.3.3 Formulação por programação dinâmica

A *formulação por programação dinâmica*, em geral, é conveniente para a descrição dos algoritmos. A ideia da programação dinâmica é construir uma relação recursiva entre funções que são resultados de problemas de otimização, onde geralmente a base da recursão é uma função que considera todo o problema. Com o intuito de criar essa recursão, utiliza-se a formulação aninhada (A.14) para definir o custo médio \bar{Q}_t

a partir do estágio t :

$$\bar{Q}_t(x_{t-1}, \xi_{[t-1]}) := \mathbb{E} \left[\min_{\substack{B_t x_{t-1} + A_t x_t = b_t \\ x_t \geq 0}} c_t^\top x_t + \mathbb{E} \left[\cdots + \mathbb{E} \left[\min_{\substack{B_T x_{T-1} + A_T x_T = b_T \\ x_T \geq 0}} c_T^\top x_T \right] \right] \middle| \xi_{[t-1]} \right] \quad (\text{A.15})$$

$t = 2, \dots, T$. Considerando a definição de \bar{Q}_t , nota-se a seguinte relação recursiva

$$\bar{Q}_t(x_{t-1}, \xi_{[t-1]}) = \mathbb{E} \left[\min_{\substack{B_t x_{t-1} + A_t x_t = b_t \\ x_t \geq 0}} c_t^\top x_t + \bar{Q}_{t+1}(x_t, \xi_{[t]}) \middle| \xi_{[t-1]} \right], \quad (\text{A.16})$$

$t = 2, \dots, T$, onde $\bar{Q}_{T+1}(\cdot, \cdot) := 0$. Dessa forma, o problema para a primeira decisão x_1 pode ser descrito como

$$\min_{\substack{A_1 x_1 = b_1 \\ x_1 \geq 0}} c_1^\top x_1 + \bar{Q}_2(x_1, \xi_{[1]}). \quad (\text{A.17})$$

A função $\bar{Q}_t(\cdot, \cdot)$, chamada de *função de custo futuro*, é interpretada como o custo médio futuro condicionado às observações até o tempo $t - 1$ e à decisão em $t - 1$.

Ainda no intuito de criar a recursão da programação dinâmica, define-se por Q_t o custo total do estágio t :

$$Q_t(x_{t-1}, \xi_{[t]}) := \min_{\substack{B_t x_{t-1} + A_t x_t = b_t \\ x_t \geq 0}} c_t^\top x_t + \bar{Q}_{t+1}(x_t, \xi_{[t]}), \quad t = 1, \dots, T. \quad (\text{A.18})$$

A função $Q_t(\cdot, \cdot)$, chamada de *função de custo total*, é interpretada como o menor custo obtido considerando a soma do custo imediato $c_t^\top x_t$ e o custo médio futuro $\bar{Q}_{t+1}(x_t, \xi_{[t]})$. Vale ressaltar que, pela definição da função de custo total (A.18), a recursão (A.16) pode ser reescrita como:

$$\bar{Q}_t(x_{t-1}, \xi_{[t-1]}) = \mathbb{E} [Q_t(x_{t-1}, \xi_{[t]}) \mid \xi_{[t-1]}], \quad t = 2, \dots, T, \quad (\text{A.19})$$

o que significa que o custo médio futuro em t é a média dos possíveis custos totais do estágio t .

Tendo em vista as definições das funções de custo total e a de custo futuro, é possível enunciar a *formulação por programação dinâmica*:

$$Q_t(x_{t-1}, \xi_{[t]}) = \min_{\substack{B_t x_{t-1} + A_t x_t = b_t \\ x_t \geq 0}} c_t^\top x_t + \bar{Q}_{t+1}(x_t, \xi_{[t]}), \quad t \in \{1, \dots, T\} \quad (\text{A.20})$$

$$\bar{Q}_{t+1}(x_t, \xi_{[t]}) = \begin{cases} \mathbb{E} [Q_{t+1}(x_t, \xi_{[t+1]}) \mid \xi_{[t]}] & , \quad t \in \{1, \dots, T-1\} \\ 0 & , \quad t = T \end{cases}.$$

A base da recursão é a função de custo total $Q_1(\xi_{[1]})$, descrita em (A.17), que

considera todo o problema.

Caso particular de um processo independente

Um caso particular importante para a descrição do algoritmo SDDP a ser enunciado adiante é quando as observações $\{\xi_t\}_{t \in \mathbb{N}}$ formam um processo independente. Neste caso particular, as funções de custo futuro não dependem de $\xi_{[t]}$ e as funções de custo total só dependem de ξ_t . Com efeito, se $\{\xi_t\}_{t \in \mathbb{N}}$ é um processo independente então a definição (A.15) da função de custo futuro pode ser escrita sem o condicionante:

$$\bar{Q}_t(x_{t-1}) = \mathbb{E} \left[\min_{\substack{B_t x_{t-1} + A_t x_t = b_t \\ x_t \geq 0}} c_t^\top x_t + \mathbb{E} \left[\cdots + \mathbb{E} \left[\min_{\substack{B_T x_{T-1} + A_T x_T = b_T \\ x_T \geq 0}} c_T^\top x_T \right] \right] \right], \quad (\text{A.21})$$

pois as restrições e custos de (A.15) são referentes às observações a partir de t , (ξ_t, \dots, ξ_T) , e estão condicionadas às observações até $t-1$, $\xi_{[t-1]}$. Pela hipótese de independência, a esperança condicionada (A.15) é igual à não condicionada (A.21). Com isso, a relação recursiva (A.16) se reduz a

$$\bar{Q}_t(x_{t-1}) = \mathbb{E} \left[\min_{\substack{B_t x_{t-1} + A_t x_t = b_t \\ x_t \geq 0}} c_t^\top x_t + \bar{Q}_{t+1}(x_t) \right], \quad t = 2, \dots, T, \quad (\text{A.22})$$

e a função de custo total (A.18) só depende da observação em t :

$$Q_t(x_{t-1}, \xi_t) = \min_{\substack{B_t x_{t-1} + A_t x_t = b_t \\ x_t \geq 0}} c_t^\top x_t + \bar{Q}_{t+1}(x_t) \quad t = 1, \dots, T. \quad (\text{A.23})$$

Substituindo (A.23) em (A.22), obtém-se a função de custo futuro como a média dos custos totais para o caso particular em que as observações são independentes:

$$\bar{Q}_t(x_{t-1}) = \mathbb{E} [Q_t(x_{t-1}, \xi_t)], \quad t = 2, \dots, T. \quad (\text{A.24})$$

A *formulação por programação dinâmica* para o caso em que as observações $\{\xi_t\}_{t \in \mathbb{N}}$ são um processo independente é, portanto,

$$Q_t(x_{t-1}, \xi_t) = \min_{\substack{B_t x_{t-1} + A_t x_t = b_t \\ x_t \geq 0}} c_t^\top x_t + \bar{Q}_{t+1}(x_t), \quad t \in \{1, \dots, T\} \quad (\text{A.25})$$

$$\bar{Q}_{t+1}(x_t) = \begin{cases} \mathbb{E} [Q_{t+1}(x_t, \xi_{t+1})] & , \quad t \in \{1, \dots, T-1\} \\ 0 & , \quad t = T \end{cases} ,$$

A.4 Formulação do problema de planejamento de longo prazo da operação

Como visto na seção A.1, o objetivo do planejamento da operação de sistemas hidrotérmicos é determinar uma meta de geração para cada usina em cada estágio do período de planejamento e estado do sistema. Essa estratégia deve minimizar o valor esperado do custo de operação ao longo do período de planejamento que compreende custos de combustível e penalidades por falha no atendimento à demanda.

A disponibilidade de quantidades limitadas de energia hidroelétrica, sob a forma de água armazenada nos reservatórios do sistema, faz com que o problema de operação seja muito complexo, porque cria uma conexão entre as decisões de um dado estágio e as futuras consequências desta decisão. Por exemplo, ao se esgotar os estoques de energia hidroelétrica e no futuro ocorrerem baixos volumes de afluência pode ser necessário o uso de geração térmica cara ou falhar no atendimento a demanda. Por outro lado, ao se manter os reservatórios cheios através de um uso mais intenso de geração térmica e no futuro ocorrer um grande volume de afluência pode haver vertimento, o que significa um desperdício de energia, e consequentemente custos de operação mais altos. Como é impossível ter uma previsão perfeita de afluências futuras, o problema de operação é essencialmente estocástico.

A seguir é apresentado, na formulação por programação dinâmica, o problema de planejamento energético de longo prazo como descrito em SHAPIRO *et al.* [64], considerando a representação a reservatório equivalente de energia:

$$\begin{aligned}
 Q_t(v_t, \mathbf{a}_{[t]}) = & \min \sum_{k \in K} \sum_{j \in T_k} c_j g_{tj} + \beta \bar{Q}_{t+1}(v_{t+1}, \mathbf{a}_{[t]}) \\
 \text{s.a. } & v_{t+1} = v_t + \mathbf{a}_t - q_t - s_t \\
 & q_{tk} + \sum_{j \in T_k} g_{tj} + \sum_{l \in \Omega_k} (f_{tlk} - f_{tkl}) = d_{tk}, \quad \forall k \in K \\
 & 0 \leq v_{t+1} \leq \bar{v}, \quad 0 \leq q_t \leq \bar{q}, \quad 0 \leq s_t, \\
 & q_t + s_t \geq \underline{q}, \quad \underline{g} \leq g_t \leq \bar{g}, \quad \underline{f} \leq f_t \leq \bar{f} \\
 \bar{Q}_{t+1}(v_{t+1}, \mathbf{a}_{[t]}) = & \begin{cases} \mathbb{E}[Q_{t+1}(v_{t+1}, \mathbf{a}_{[t+1]}) \mid \mathbf{a}_{[t]}] & , t \in \{1, \dots, T-1\} \\ 0 & , t = T \end{cases} ,
 \end{aligned}$$

para todo $t = 1, \dots, T$.

A função objetivo

$$\sum_{k \in K} \sum_{j \in T_k} c_j g_{tj} + \beta \bar{Q}_{t+1}(v_{t+1}, \mathbf{a}_{[t]})$$

representa a soma do custo total de geração térmica, déficit e o custo futuro $\bar{Q}_{t+1}(v_{t+1}, \mathbf{a}_{[t]})$ convertido em seu valor presente:

- β fator de desconto;
- K conjunto de subsistemas; e
- T_k conjunto de térmicas do subsistema k .

A equação de *balanço energético* para cada reservatório é:

$$v_{t+1} = v_t + \mathbf{a}_t + q_t + s_t$$

- v_t vetor de energia armazenada no início do estágio t ;
- \mathbf{a}_t vetor de energia afluyente durante o estágio t ;
- q_t vetor de energia turbinada durante o estágio t ; e
- s_t vetor de energia vertida durante o estágio t .

A equação de *balanço de carga*, em MWmês, em cada subsistema k e estágio t é:

$$q_{tk} + \sum_{j \in T_k} g_{tj} + \sum_{l \in \Omega_k} (f_{tlk} - f_{tkl}) = d_{tk}$$

- d_{tk} carga;
- q_{tk} geração hidráulica;
- $\sum_{j \in T_k} g_{tj}$ geração térmica (incluindo o deficit);
- f_{tlk} fluxo de energia do subsistema l para o subsistema k ; e
- Ω_k conjunto de subsistemas diretamente conectados ao subsistema k .
- $\sum_{l \in \Omega_k} (f_{tlk} - f_{tkl})$ intercâmbio líquido de energia;

O não atendimento a demanda é representado por uma geração térmica fictícia de capacidade infinita e custo associado elevado. A interpretação desse custo é o impacto econômico unitário associado ao corte de carga.

Os limites das variáveis são:

- $0 \leq v_{t+1} \leq \bar{v}$, energia armazenável mínima e máxima;
- $0 \leq q_t \leq \bar{q}$, geração hidráulica mínima e máxima;
- $0 \leq s_t$, vertimento mínimo;
- $q_t + s_t \geq \underline{q}$, energia defluente mínima;
- $\underline{g} \leq g_t \leq \bar{g}$, geração térmica mínima e máxima; e
- $\underline{f} \leq f_t \leq \bar{f}$, intercâmbio energético mínimo e máximo.

Para cada estágio t , o vetor de decisão é dado por $x_t = (v_{t+1}, q_t, s_t, g_t, f_t)$. No problema de planejamento da operação a única incerteza considerada é a afluência, isto é, $\xi_t = \mathbf{a}_t$, e a função de custo total e futuro dependem somente da componente

v_t da decisão anterior x_{t-1} . Neste sentido, é usual escrever $Q_t(v_t, \mathbf{a}_{[t]})$ e $\bar{Q}_t(v_t, \mathbf{a}_{[t]})$ ao invés de $Q_t(x_{t-1}, \mathbf{a}_{[t]})$ e $\bar{Q}_t(x_{t-1}, \mathbf{a}_{[t]})$.

A.5 Árvore de cenários

Como em geral, para variáveis aleatórias *contínuas*, não é possível calcular analiticamente a função de custo futuro como a esperança das funções de total, aproxima-se a respectiva distribuição por uma distribuição *discreta* finita dada por uma árvore de cenários. Para tal procedimento, existem várias técnicas como: Monte Carlo, quasi-Monte Carlo, importance sampling, Latin Hypercube, entre outros, descritos em SHAPIRO *et al.* [61]. Neste ponto, assume-se que o processo aleatório já foi discretizado.

O propósito desta seção é definir o que é uma árvore de cenários, sugerir uma notação para se referir às componentes dessa árvore e descrever a formulação por programação dinâmica (A.20) e (A.25) para um processo discreto dependente e independente representado nestes termos. Para ilustrar as definições que serão apresentadas a seguir é dado um exemplo de uma árvore de cenários dependente e uma independente.

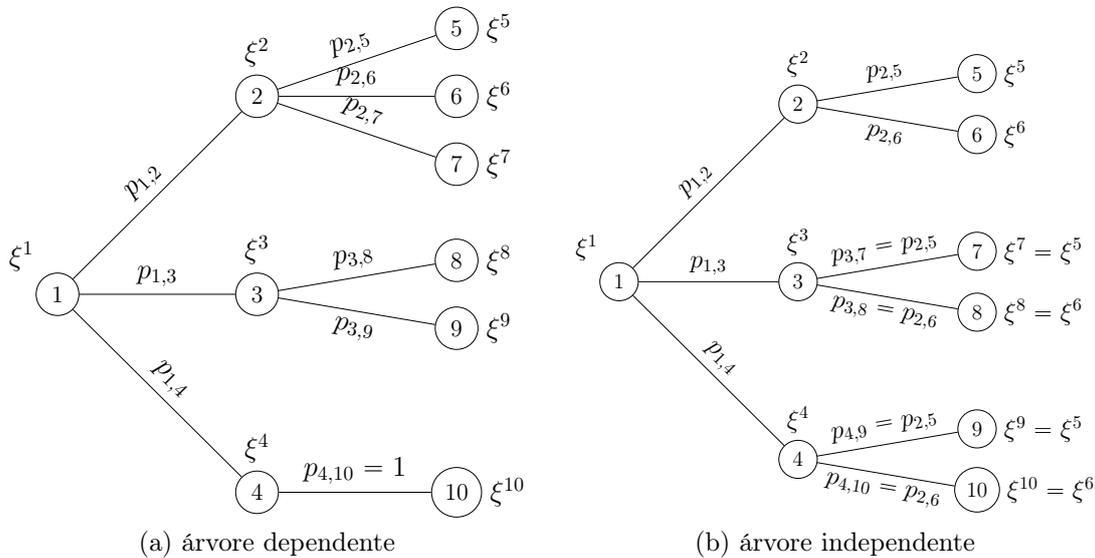


Figura A.1: Árvores de cenários

De modo geral, um processo aleatório discreto pode ser representado por uma *árvore* cujos os vértices possuem a informação das possíveis realizações das variáveis aleatórias e as arestas possuem a informação das probabilidades condicionais do processo, ou seja, a informação da probabilidade de um dado vértice condicionado à todo passado observado até o vértice antecessor. Esta árvore é usualmente chamada de *árvore de cenários*.

Formalmente, denota-se por \mathcal{N} o conjunto finito de vértices da árvore, por 1 o vértice raiz e por $a(\cdot)$ a função antecessora que a cada vértice j associa o respectivo vértice antecessor $a(j)$, onde apenas o vértice raiz não possui antecessor:

$$a: \mathcal{N} \setminus \{1\} \rightarrow \mathcal{N}$$

$$j \mapsto a(j).$$

A propriedade que define a função antecessora $a(\cdot)$ é a de *não possuir ciclos*, isto é, não existe vértice j tal que seu r -ésimo antecessor seja ele mesmo, isto é, $a^r(j) \neq j$, $\forall r \geq 1$. Ressalta-se que cada vértice deve possuir um único índice. Assim, define-se uma *árvore* como uma trinca $(\mathcal{N}, 1, a(\cdot))$, onde \mathcal{N} é um conjunto finito, 1 é um elemento especial de \mathcal{N} chamado de vértice raiz e $a(\cdot)$ é a função antecessora tal como definida acima. No caso das figuras A.1a e A.1b, tem-se $\mathcal{N}_{\text{dep}} = \mathcal{N}_{\text{indep}} = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ e as respectivas funções antecessoras $a(\cdot)$ descritas na tabela A.2a e A.2b abaixo.

j	2	3	4	5	6	7	8	9	10
$a(j)$	1	1	1	2	2	2	3	3	4

j	2	3	4	5	6	7	8	9	10
$a(j)$	1	1	1	2	2	3	3	4	4

(a) Função antecessora da árvore dependente (b) Função antecessora da árvore independente

Tabela A.2: Exemplo de funções antecessoras

A seguir, são descritas as diversas estruturas inerentes a uma *árvore* $(\mathcal{N}, 1, a(\cdot))$:

- (i) O conjunto de *arestas* \mathcal{A} é formado por todos os pares $(a(j), j)$, onde j é um vértice diferente da raiz 1. Por exemplo, nas figuras A.1a e A.1b os conjuntos de arestas são:

$$\mathcal{A}_{\text{dep}} = \{(1, 2), (1, 3), (1, 4), (2, 5), (2, 6), (2, 7), (3, 8), (3, 9), (4, 10)\}$$

$$\mathcal{A}_{\text{indep}} = \{(1, 2), (1, 3), (1, 4), (2, 5), (2, 6), (3, 7), (3, 8), (4, 9), (4, 10)\};$$

- (ii) O conjunto de vértices *filhos* ou também denominado *sucessores* de um vértice i é o conjunto formado pelos vértices que tem i como antecessor, ou seja, $\mathcal{S}(i) := \{j \in \mathcal{N} \mid i = a(j)\}$. Por exemplo, na figura A.1a o vértice 2 possui como conjunto de sucessores $\mathcal{S}(2) = \{5, 6, 7\}$.
- (iii) Diz-se que um vértice j pertence ao *estágio* t , se seu $(t-1)$ -ésimo antecessor é o vértice raiz: $a^{t-1}(j) = 1$. Por convenção, assume-se que a raiz pertence ao primeiro estágio. O conjunto dos vértices pertencentes ao estágio t é denotado por \mathcal{N}_t . Por exemplo, na figura A.1b o vértice 6 pertence ao terceiro estágio, pois $a^2(6) = a(2) = 1$, e o conjunto dos vértices do terceiro estágio é $\mathcal{N}_3 = \{5, 6, 7, 8, 9, 10\}$;

- (iv) Um vértice j é uma *folha* se pertence ao último estágio T ;
- (v) Diz-se que um vértice i é *ancestral* do vértice j , se existir $r \geq 0$ tal que $i = a^r(j)$. Por convenção, define-se $a^0(j)$ como o próprio vértice j . Logo, todo vértice é ancestral de si mesmo (não é antecessor). Por exemplo, na figura A.1a o vértice 9 tem como ancestrais os vértices 9, 3 e 1;
- (vi) O conjunto de vértices da *sub-árvore* com raiz i é o conjunto formado pelos vértices que tem i como ancestral, ou seja, $\mathcal{T}(i) = \{j \in \mathcal{N} \mid i = a^r(j), r \geq 0\}$. A sub-árvore com raiz i é dada por $(\mathcal{T}(i), i, a_{|\mathcal{T}(i)\setminus\{i\}})$, onde $a_{|\mathcal{T}(i)\setminus\{i\}}$ é a função antecessora $a(\cdot)$ restrita ao conjunto de vértices $\mathcal{T}(i)\setminus\{i\}$. Por exemplo, na figura A.1a o conjunto de vértices da sub-árvore dependente com raiz em 3 é $\mathcal{T}(3) = \{3, 8, 9\}$ e a restrição da função antecessora $a_{|\mathcal{T}(3)\setminus\{3\}}(j)$ é dada pela tabela A.3.

j	8	9
$a_{ \mathcal{T}(3)\setminus\{3\}}(j)$	3	3

Tabela A.3: Função antecessora da sub-árvore dependente

- (vii) Se o vértice j pertence ao estágio t , o *ancestral do estágio* τ de j ($\tau \leq t$) é definido por $j_\tau = a^{t-\tau}(j)$. Decorre da definição que se j pertence ao estágio t então seu ancestral do primeiro estágio é o vértice raiz: $j_1 = 1$. Por exemplo, na figura A.1b o ancestral do segundo estágio do vértice 10 é 4 e o do primeiro estágio é 1.

Para que a estrutura de árvore acima descrita constitua uma árvore de cenários é necessário incorporar a informação das realizações do processo aleatório $\{\xi_t\}_{t=1}^T$ e as respectivas probabilidades na estrutura de árvore. Neste sentido, associa-se a informação de uma possível realização da variável aleatória ξ_t a um vértice j do estágio t . Esta informação é denotada por ξ^j . Os vértices desde a raiz até o vértice j do estágio t são denotados por $[j] := (a^{t-1}(j), \dots, a(j), j)$ e estão associados a uma possível realização da variável aleatória $\xi_{[j]}$. Esta informação é denotada por $\xi^{[j]}$. Por exemplo, na figura A.1a o conjunto de valores que as variáveis aleatórias ξ_2 e $\xi_{[2]}$ podem assumir são $\Xi_2 = \{\xi^2, \xi^3, \xi^4\}$ e $\Xi_{[2]} = \{\xi^{[2]}, \xi^{[3]}, \xi^{[4]}\} = \{(\xi^1, \xi^2), (\xi^1, \xi^3), (\xi^1, \xi^4)\}$, respectivamente.

Com relação às probabilidades, a informação da probabilidade de ξ_t assumir o valor ξ^j condicionado ao fato de que nos estágios anteriores a t se observou $\xi_{[t-1]}$ igual a $\xi^{[a(i)]}$, isto é, $\mathbb{P}(\xi_t = \xi^j \mid \xi_{[t-1]} = \xi^{[a(j)]})$, está associada à aresta $(a(j), j)$, onde j é um vértice do estágio $t \geq 2$. Esta informação é denotada por $p_{a(j),j}$. Pela definição de probabilidade condicional, destaca-se para um dado vértice i que

$p_{i,j} \geq 0$ para todo j sucessor de i e $\sum_{j \in \mathcal{S}(i)} p_{i,j} = 1$. Por exemplo, na figura A.1a $p_{2,5} + p_{2,6} + p_{2,7} = 1$ e $p_{2,5}, p_{2,6}, p_{2,7} \geq 0$.

Com isso, uma árvore de cenários é definida como uma árvore $(\mathcal{N}, 1, a(\cdot))$ associada às possíveis realizações do processo aleatório $\{\xi^j\}_{j \in \mathcal{N}}$ e probabilidades condicionais $\{p_{ij}\}_{(i,j) \in \mathcal{A}}$. A seguir, são descritas algumas estruturas inerentes a uma *árvore de cenários*:

- (a) O vértice raiz representa a informação do primeiro estágio que se supõe conhecida e por isso a probabilidade associada é 1, ou seja, $p^1 := \mathbb{P}(\xi_1 = \xi^1) = 1$. Com este fato e a definição de probabilidade da aresta, tem-se a motivação para definir a probabilidade associada a um vértice j do estágio t como a probabilidade de $\xi_{[t]}$ ser igual a $\xi^{[j]}$, ou seja, $p^j := \mathbb{P}(\xi_{[t]} = \xi^{[j]})$. Com isso é possível interpretar a decomposição da probabilidade conjunta em produto de probabilidades condicionais:

$$\mathbb{P}(\xi_{[t]} = \xi^{[j]}) = \mathbb{P}(\xi_1 = \xi^1) \prod_{\tau=2}^t \mathbb{P}(\xi_{[\tau]} = \xi^{[j_\tau]} \mid \xi_{[\tau-1]} = \xi^{[j_{(\tau-1)}]})$$

como a decomposição da probabilidade do vértice j em produto das probabilidades das arestas relacionadas ao caminho que liga o vértice raiz a j :

$$p^j = p^1 \prod_{\tau=2}^t p_{j_{(\tau-1)}, j_\tau},$$

onde j_τ é *ancestral* de j do estágio τ . Destaca-se também a interpretação da probabilidade condicional em função da razão de probabilidades conjuntas:

$$\mathbb{P}(\xi_t = \xi^j \mid \xi_{[t-1]} = \xi^{[a(j)]}) = \frac{\mathbb{P}(\xi_{[t]} = \xi^{[j]})}{\mathbb{P}(\xi_{[t-1]} = \xi^{[a(j)]})}$$

como a probabilidade da aresta em função da razão da probabilidade dos vértices de sua extremidade: $p^j = p_{a(j), j} p^{a(j)}$. Por exemplo, na figura A.1a $p^9 = p_{3,9} p^3 = p_{1,3} p_{3,9}$.

- (b) Um *cenário* é uma particular realização dos coeficientes aleatórios do primeiro ao último estágio T , isto é, um cenário é uma realização de $\xi_{[T]}$. Se a distribuição dos coeficientes $\xi_{[T]}$ é descrita por uma árvore de cenários, um cenário corresponde às realizações ao longo do único caminho da raiz até uma de suas folhas, ou seja, os cenários são dados por $\xi^{[j]}$, onde j são vértices do último estágio T .

Considerando a notação de árvore de cenários, é conveniente escrever a formulação por programação dinâmica (A.20) e (A.25) para processos dependentes

e independentes nesta notação. Dessa forma, as funções de custo futuro e total são descritas pelos possíveis valores do processo aleatório $\{\xi_t\}_{t=1}^T$ e as respectivas probabilidades ao invés do processo aleatório em si.

Neste sentido, seja $(\mathcal{N}, 1, a(\cdot))$, $\{\xi^j\}_{j \in \mathcal{N}}$ e $\{p_{ij}\}_{(i,j) \in \mathcal{A}}$ a representação em árvore de cenários do processo aleatório $\{\xi_t\}_{t=1}^T$ e ξ^i uma possível realização de ξ_{t-1} . Com isso, a função de custo futuro em (A.20) pode ser descrita por:

$$\begin{aligned}\bar{Q}_t(x_{t-1}, \xi^{[i]}) &= \mathbb{E} [Q_t(x_{t-1}, \xi_{[t]}) \mid \xi_{[t-1]} = \xi^{[i]}] \\ &= \sum_{j \in \mathcal{S}(i)} \mathbb{P}(\xi_t = \xi^j \mid \xi_{[t-1]} = \xi^{[i]}) Q_t(x_{t-1}, \xi^{[j]}) \\ &= \sum_{j \in \mathcal{S}(i)} p_{ij} Q_t(x_{t-1}, \xi^{[j]}).\end{aligned}$$

A formulação por programação dinâmica (A.20) para árvore de cenários fica:

$$\begin{aligned}Q_t(x_{t-1}, \xi^{[j]}) &= \min_{\substack{B^j x_{t-1} + A^j x_t = b^j \\ x_t \geq 0}} c^j x_t + \bar{Q}_{t+1}(x_t, \xi^{[j]}) & \quad (\text{A.26}) \\ \bar{Q}_{t+1}(x_t, \xi^{[j]}) &= \begin{cases} \sum_{k \in \mathcal{S}(j)} p_{jk} Q_{t+1}(x_t, \xi^{[k]}) & , t \in \{1, \dots, T-1\} \\ 0 & , t = T \end{cases}\end{aligned}$$

para todo $j \in \mathcal{N}_t$ e $t = 1, \dots, T$. De modo a não sobrecarregar a notação, omite-se o símbolo de transposto $(\cdot)^\top$ para o produto interno entre c^j e x_t . Por exemplo, as figuras A.2a e A.3a representam as funções de custo total e futuro em cada vértice da árvore de cenários, ou seja, as funções de custo total e futuro associadas à informação disponível no dado estágio. De modo análogo, para o caso independente (A.25) a mesma relação para as funções de custo futuro se descreve por:

$$\begin{aligned}\bar{Q}_t(x_{t-1}) &= \mathbb{E} [Q_t(x_{t-1}, \xi_t)] \\ &= \mathbb{E} [Q_t(x_{t-1}, \xi_t) \mid \xi_{[t-1]} = \xi^{[i]}] \\ &= \sum_{j \in \mathcal{S}(i)} p_{ij} Q_t(x_{t-1}, \xi^j),\end{aligned}$$

onde a segunda igualdade decorre da propriedade de independência da árvore de cenários. Com isso, a formulação por programação dinâmica do caso independente (A.25) é:

$$\begin{aligned}Q_t(x_{t-1}, \xi^j) &= \min_{\substack{B^j x_{t-1} + A^j x_t = b^j \\ x_t \geq 0}} c^j x_t + \bar{Q}_{t+1}(x_t) & \quad (\text{A.27}) \\ \bar{Q}_{t+1}(x_t) &= \begin{cases} \sum_{k \in \mathcal{S}(j)} p_{jk} Q_{t+1}(x_t, \xi^k) & , t \in \{1, \dots, T-1\} \\ 0 & , t = T \end{cases}\end{aligned}$$

para todo $j \in \mathcal{N}_t$ e $t = 1, \dots, T$. Por exemplo, as figuras A.2b e A.3b representam as funções de custo total e futuro em cada vértice da árvore de cenários independente.

Uma convenção que facilita a associação de um vértice com o respectivo estágio é reservar o uso de certas letras para certos estágios, tendo como referência o valor de t . Neste documento, as letras i , j e k são utilizadas para os vértices do estágio $t - 1$, t e $t + 1$, respectivamente: $i \in \mathcal{N}_{t-1}$, $j \in \mathcal{N}_t$ e $k \in \mathcal{N}_{t+1}$.

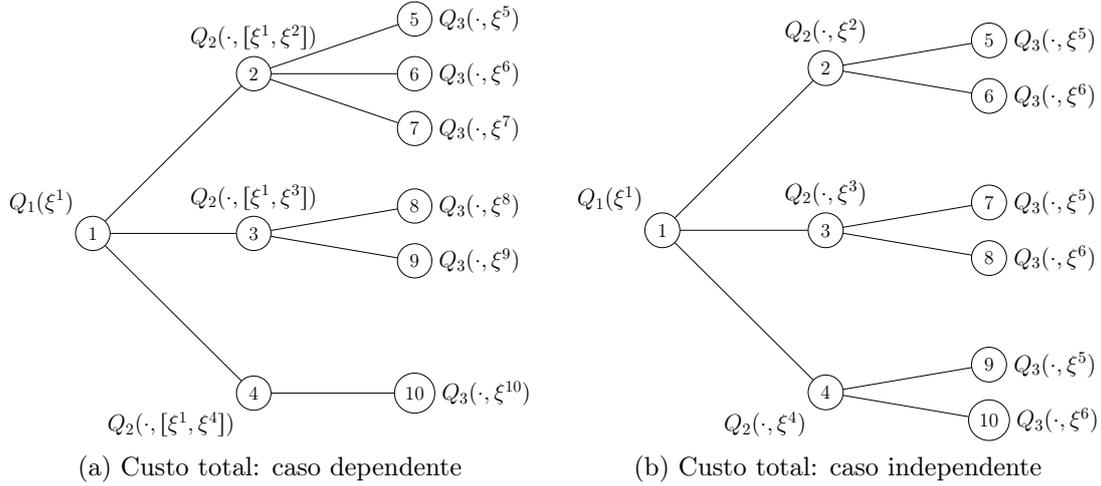


Figura A.2: Funções de custo total

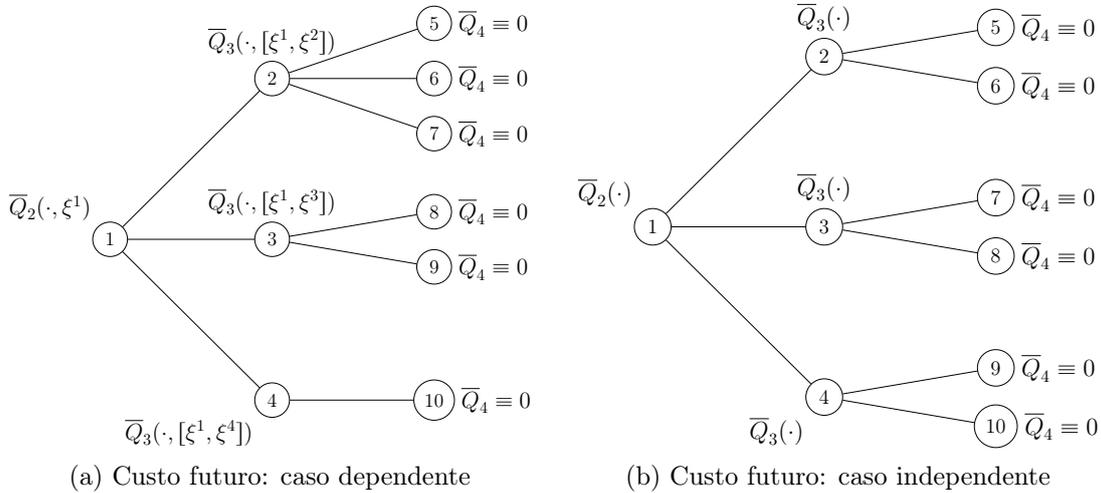


Figura A.3: Funções de custo futuro

A.6 Elementos de análise convexa e algoritmos em otimização estocástica

O conhecimento das funções de custo futuro e total para qualquer estágio e histórico de observações permite obter a decisão ótima em qualquer instante por meio da

solução de (A.20) em cada estágio. Apesar de se observar as relações entre $Q_t(\cdot, \cdot)$ e $\bar{Q}_t(\cdot, \cdot)$, não é conhecida uma fórmula analítica geral para se obter estas funções, mesmo quando o processo aleatório é descrito por uma árvore de cenários.

Neste sentido, torna-se necessário criar algoritmos que forneçam aproximações para as funções de custo futuro e total. Para esta finalidade, descrevem-se algumas propriedades dessas funções, considerando as definições a seguir e a ilustração da figura A.4. Seja $f : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$ onde $\bar{\mathbb{R}} = \mathbb{R} \cup \{+\infty, -\infty\}$

- I. O *epígrafo* de $f(\cdot)$, $\text{epi}(f)$, é a região definida acima do gráfico de f , ou seja, $\text{epi}(f) := \{(x, w) \in \mathbb{R}^{n+1} \mid f(x) \leq w\}$;
- II. Diz-se que $f(\cdot)$ é *convexa* se seu epígrafo é um conjunto convexo;
- III. O *domínio efetivo* de $f(\cdot)$, $\text{dom}(f)$, é o conjunto de pontos $x \in \mathbb{R}^n$ em que $f(x)$ é menor do que $+\infty$;
- IV. Diz-se que $f(\cdot)$ é *própria* se seu domínio efetivo é não vazio e se em todo ponto tal função é maior do que $-\infty$;
- V. Diz-se que $f(\cdot)$ é *poliedral* se é própria, seu domínio efetivo é um conjunto poliedral convexo fechado e em seu domínio efetivo tal função é o máximo de um número finito de funções afins. Em outras palavras:

$$f(x) = \begin{cases} \max_{l=1, \dots, L} (r_l^\top x + \gamma_l) & , \text{ se } Ax \leq b \\ +\infty & , \text{ caso contrário} \end{cases}$$

e $\text{dom}(f) = \{x \in \mathbb{R}^n \mid Ax \leq b\} \neq \emptyset$. Uma função poliedral é em particular uma função convexa como ilustrado na figura A.4.

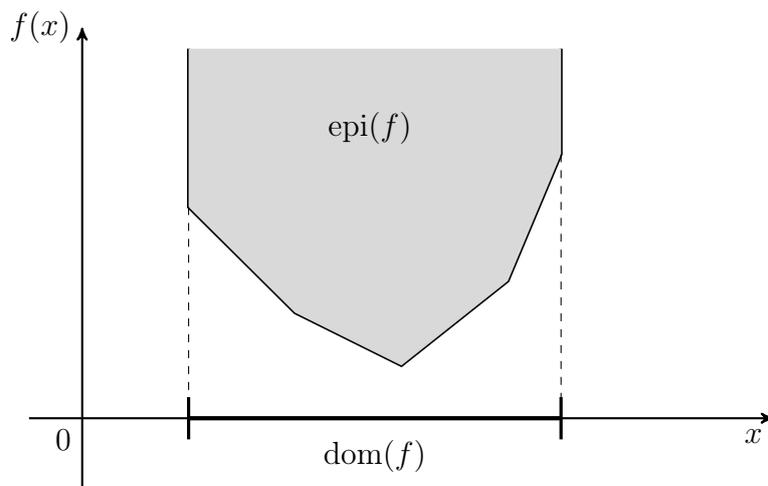


Figura A.4: função poliedral

Por convenção, o valor ótimo de um problema de minimização é $+\infty$ se o respectivo conjunto de soluções viáveis é vazio. Como as funções de custo total (A.18) e futuro (A.15) são definidas como valores ótimos de problemas de minimização, ao mudar os argumentos dessas funções mudam o conjunto de soluções viáveis dos respectivos problemas. Logo, é possível que para certas observações e decisões passadas tais conjuntos de viabilidade sejam vazios e com isso as funções de custo futuro ou total assumam $+\infty$.

Por outro lado, assume-se que as funções de custo total são maiores do que $-\infty$, para toda decisão passada e realização da incerteza. Por isso, essa hipótese precisa ser verificada para cada problema. Em geral os recursos físicos são finitos, o que significa que na maioria dos problemas não existe uma decisão “infinitamente boa” e portanto essa hipótese é quase sempre verdadeira.

Visando construir um algoritmo que forneça aproximações para as funções de custo futuro e total, é importante compreender melhor as propriedades dessas funções. Uma primeira propriedade é o seguinte

Fato 1: $\bar{Q}_t(\cdot, \xi_{[t-1]})$ e $Q_t(\cdot, \xi_{[t]})$ são funções poliedrais.

Considerando a estrutura de uma função poliedral $f(\cdot)$, uma ideia natural é aproximar $f(\cdot)$ em um dado ponto x^* pelo plano tangente que passa por $f(x^*)$, pois em seu domínio efetivo $f(\cdot)$ é uma função convexa, linear por partes e portanto pode ser representada por um número finito de planos. A maneira tradicional de se parametrizar cada plano tangente é pela expansão de Taylor de primeira ordem: $f(x^*) + \nabla f(x^*)^\top (x - x^*)$.

Entretanto, funções poliedrais, assim como funções convexas em geral, podem não ser diferenciáveis em alguns pontos, como os bicos da figura A.4 sugerem, e por isso é necessário que se estenda o conceito de diferenciabilidade. Dada uma função convexa própria $f : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}$ e um ponto x^* tal que $f(x^*)$ é finito, diz-se que $g \in \mathbb{R}^n$ é um *subgradiente* de $f(\cdot)$ em x^* se

$$f(x) \geq f(x^*) + g^\top (x - x^*), \quad \forall x \in \mathbb{R}^n \quad (\text{A.28})$$

O conjunto de todos os subgradientes de $f(\cdot)$ em x^* é chamado de *subdiferencial* e é denotado por $\partial f(x^*)$. Quando $f(x^*)$ é igual a $+\infty$, define-se $\partial f(x^*)$ como sendo vazio. Pode-se mostrar que $f(\cdot)$ é diferenciável em $x^* \in \text{int}(\text{dom } f)$ se, e somente se, $\partial f(x^*)$ possui um único elemento. Neste caso, $\partial f(x^*) = \{\nabla f(x^*)\}$.

Considerando essa generalização de derivada, observa-se que $f(x^*) + g^\top (x - x^*)$ é uma parametrização de um plano tangente em $(x^*, f(x^*))$ que contém $\text{epi}(f)$ no lado oposto ao vetor normal $(g, -1)$:

$$g^\top x^* - f(x^*) \geq g^\top x - f(x), \quad \forall x \in \mathbb{R}^n,$$

como ilustrado na figura A.5. Por exemplo, a função módulo $|\cdot|$ de um número real x

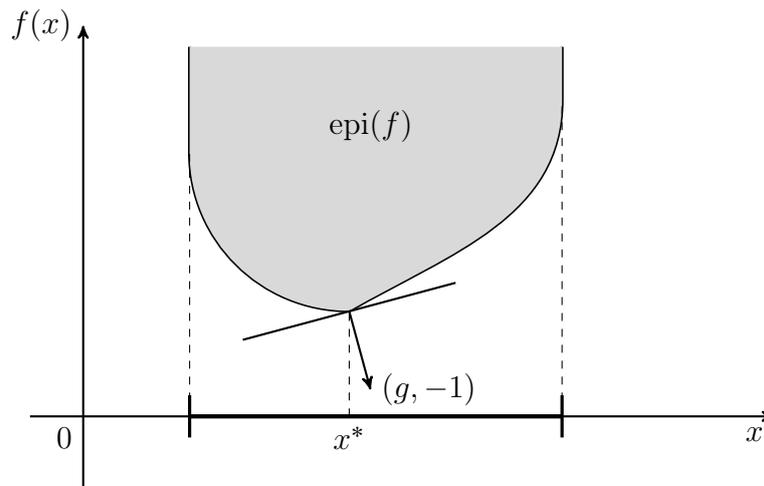


Figura A.5: Plano tangente e vetor subgradiente

é uma função poliedral, pois $|x| = \max(x, -x)$, e conseqüentemente uma função convexa própria. Para valores de x maiores que 0, tem-se $f(x) = x$ e portanto f é diferenciável em x com $f'(x) = 1$. De modo análogo, para valores de x menores que 0, tem-se $f(x) = -x$ e por isso f é diferenciável em x com $f'(x) = -1$. Quando x é igual a 0, a função f é não diferenciável e por isso é necessário aplicar a definição de subgradiente: $|x| \geq g \cdot x, \forall x \in \mathbb{R}$. Com isso, aplicando valores positivos e negativos para x , obtem-se que $1 \geq g$ e $-1 \leq g$, respectivamente. Além disso, todo g entre -1 e 1 satisfaz a definição de subgradiente em 0. Portanto, o conjunto de subgradientes de $f(\cdot)$ em 0 são os números entre -1 e 1 , ou seja, $\partial f(0) = [-1, 1]$. Resumindo,

$$\partial f(x) = \begin{cases} \{-1\} & , \text{ se } x < 0 \\ [-1, 1] & , \text{ se } x = 0 \\ \{1\} & , \text{ se } x > 0 \end{cases}$$

Na figura A.6 é ilustrada a função módulo e os respectivos subgradientes.

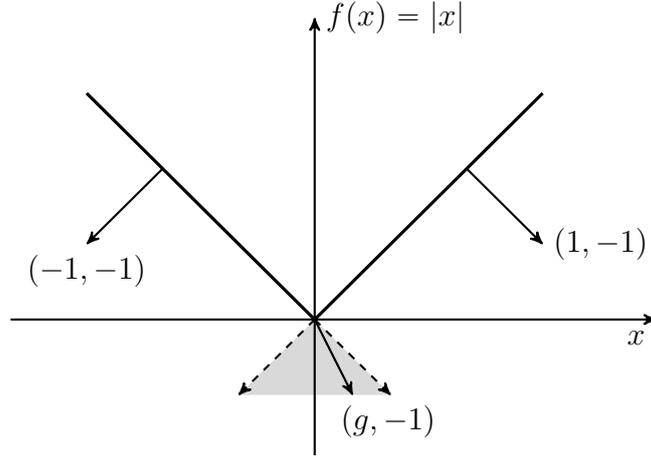


Figura A.6: Função módulo

Neste momento, levanta-se a seguinte pergunta: como calcular os subgradientes das funções de custo total e futuro? Os resultados são enunciados a seguir.

Fato 2: Quando $Q_t(x_{t-1}, \xi_{[t]})$ é finito, o subdiferencial de $Q_t(\cdot, \xi_{[t]})$ no ponto x_{t-1} é

$$\partial Q_t(x_{t-1}, \xi_{[t]}) = -B_t^\top \mathfrak{D}_t(x_{t-1}, \xi_{[t]}), \quad (\text{A.29})$$

onde $\mathfrak{D}_t(x_{t-1}, \xi_{[t]})$ é o conjunto de soluções duais ótimas associados ao problema (A.18) que define a função de custo total $Q_t(\cdot, \xi_{[t]})$. Obs: o conjunto de soluções duais ótimas é igual a menos o conjunto de *multiplicadores de Lagrange* ótimos.

Fato 3 (Subdiferencial da soma): Quando $Q_t(x_{t-1}, \xi^{[j]})$ é finito para todo vértice j sucessor de i , tem-se que $\bar{Q}_t(x_{t-1}, \xi^{[i]})$ é finito e o subdiferencial de $\bar{Q}_t(\cdot, \xi^{[i]})$ no ponto x_{t-1} é:

$$\partial \bar{Q}_t(x_{t-1}, \xi^{[i]}) = \sum_{j \in S(i)} p_{ij} \partial Q_t(x_{t-1}, \xi^{[j]}) \quad (\text{A.30})$$

Uma análise desses resultados indica que o subdiferencial da função de custo futuro $\bar{Q}_t(\cdot, \xi^{[i]})$ depende do subdiferencial das funções de custo total $Q_t(\cdot, \xi^{[j]})$ de cada vértice j sucessor de i e o subdiferencial da função de custo total $Q_t(\cdot, \xi^{[j]})$ depende das soluções duais ótimas referentes ao problema de minimização (A.18) que o define. Entretanto, a função objetivo do problema de minimização que define a função de custo total $Q_t(\cdot, \xi^{[j]})$ depende do custo futuro $\bar{Q}_{t+1}(\cdot, \xi^{[j]})$ que é referente ao estágio posterior e não é conhecido previamente, a menos do caso $t = T$, onde $\bar{Q}_{T+1}(\cdot, \cdot) = 0$. Por esse motivo, o único caso em que se pode calcular subgradientes com (A.29) e (A.30) é para o último estágio T . Com isso, a ideia de aproximar o custo total e o custo futuro em cada estágio por planos tangentes precisa ser reavaliada.

Uma abordagem para a questão levantada é calcular *todos* os planos tangentes da função de custo futuro e total do último estágio T . Feito isso, tem-se a descrição

completa da função de custo futuro $\bar{Q}_T(\cdot, \xi^{[i]})$ e a partir daí gera-se com (A.29) e (A.30) todos os planos tangentes da função de custo total e futuro do estágio $T - 1$. Este procedimento continua até que se conheça todas as funções de todos os estágios. Porém, existem duas principais dificuldades que tornam este algoritmo impraticável:

- (a) O desconhecimento de um bom critério que verifica se um dado conjunto de planos tangentes representa a respectiva função em todos os pontos;
- (b) O tempo computacional para o cálculo de todos os planos tangentes ou o espaço necessário para o armazenamento dos mesmos pode ser proibitivo.

A.6.1 Algoritmo Nested Cutting Plane

Um aperfeiçoamento destas duas últimas ideias é calcular *alguns* planos que aproximam inferiormente as funções de custo futuro e total, mas que não são necessariamente *tangentes*, e acrescentar esses planos à aproximação da função de custo futuro de cada estágio. A ideia é que a aproximação inferior da função de custo futuro seja definida pelo máximo dos planos inferiores gerados até a iteração corrente, o que por definição é uma função poliedral. Com isso, a cada novo plano inferior acrescentado uma *nova* aproximação inferior da função de custo futuro é produzida, maior ou igual à anterior. Este processo de acrescentar planos inferiores é chamado de refinamento da aproximação da função de custo futuro.

Fixada uma família de funções poliedrais $\{\bar{\Omega}_t(\cdot, \xi^{[i]} \mid i \in \mathcal{N}_t, t = 1, \dots, T\}$ que aproximam inferiormente as respectivas funções de custo futuro:

$$\bar{\Omega}_t(x_{t-1}, \xi^{[i]}) \leq \bar{Q}_t(x_{t-1}, \xi^{[i]}), \quad \forall x_{t-1}$$

é possível induzir uma família de funções $\{\Omega_t(\cdot, \xi^{[j]} \mid j \in \mathcal{N}_t, t = 1, \dots, T\}$ que aproximam inferiormente as respectivas funções de custo total ao substituir na definição (A.26) a função de custo futuro pela sua aproximação inferior:

$$\Omega_t(x_{t-1}, \xi^{[j]}) := \min_{\substack{B^j x_{t-1} + A^j x_t = b^j \\ x_t \geq 0}} c^j x_t + \bar{\Omega}_{t+1}(x_t, \xi^{[j]}), \quad \forall x_{t-1}. \quad (\text{A.31})$$

A função $\Omega_t(\cdot, \xi^{[j]})$ é uma aproximação inferior de $Q_t(\cdot, \xi^{[j]})$, pois os problemas de otimização (A.31) e (A.26) que as definem são idênticos a menos da função objetivo, onde a função objetivo de (A.31) é menor do que ou igual a de (A.26). Assim:

$$\Omega_t(x_{t-1}, \xi^{[j]}) \leq Q_t(x_{t-1}, \xi^{[j]}), \quad \forall x_{t-1}.$$

A aproximação da função de custo total $\Omega_t(\cdot, \xi^{[j]})$ possui propriedades importantes que serão usadas na construção do algoritmo Nested Cutting Plane e que estão

descritas no Fato 4.

Fato 4: $\mathfrak{Q}_t(\cdot, \xi^{[j]})$ é uma função poliedral. Quando $\mathfrak{Q}_t(x_{t-1}, \xi^{[j]})$ é finito, o subdiferencial de $\mathfrak{Q}_t(\cdot, \xi^{[j]})$ no ponto x_{t-1} é

$$\partial \mathfrak{Q}_t(x_{t-1}, \xi^{[j]}) = -B_t^\top \widehat{\mathfrak{D}}_t(x_{t-1}, \xi^{[j]}), \quad (\text{A.32})$$

onde $\widehat{\mathfrak{D}}_t(x_{t-1}, \xi^{[j]})$ é o conjunto de soluções duais ótimas de (A.31).

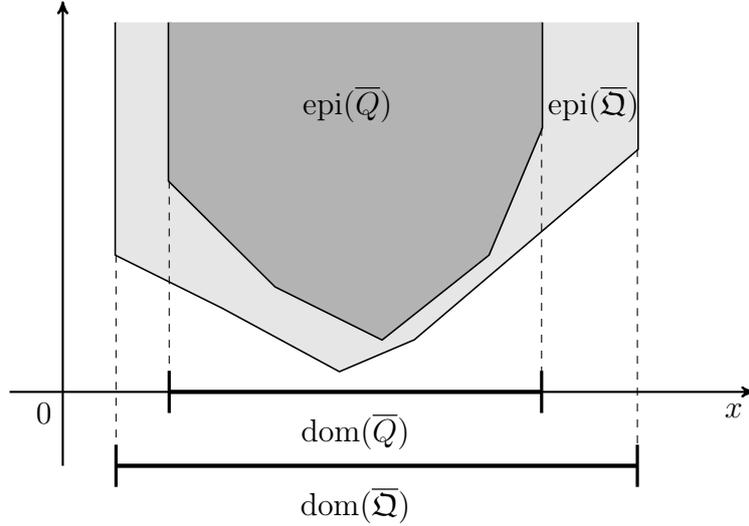


Figura A.7: Aproximação da função de custo futuro

De modo a simplificar esta exposição, é interessante que a função de custo futuro e total sejam sempre finitas. Para esta finalidade, é suficiente supor que o problema original (A.26) seja *relativamente completo*, isto é, toda decisão viável do estágio anterior torna o conjunto viável do estágio posterior não vazio para qualquer possível realização passada $\xi_{[t]}$ representada na árvore de cenários. Com isso, as funções de custo futuro e total são sempre finitas para toda decisão viável.

Neste ponto a pergunta é como refinar a aproximação inferior das funções de custo futuro e total. Para responder essa pergunta, é importante lembrar a limitação da abordagem de cálculo de planos tangentes. Como mencionado no Fato 3, a função de custo futuro do estágio t , $\overline{Q}_t(\cdot, \xi^{[i]})$, depende das funções de custo total do estágio t , $\{Q_t(\cdot, \xi^{[j]})\}_{j \in \mathcal{S}(i)}$, e cada função de custo total $Q_t(\cdot, \xi^{[j]})$ depende da função de custo futuro do estágio $t + 1$, $\overline{Q}_{t+1}(\cdot, \xi^{[j]})$, que é conhecida apenas no caso $t = T$ em que é identicamente nula. Em função do Fato 3, a pergunta sobre como refinar as aproximações inferiores das funções do último estágio T já está respondida. A questão importante agora é: a partir da aproximação das funções de custo futuro e total de um dado estágio $t + 1$, como melhorar a aproximação das funções do estágio t ? Com a resposta desta pergunta, responde-se também a pergunta inicial que gerou esta análise, pois por indução ao longo dos estágios se refina as aproximações das

funções de um dado estágio por meio das aproximações dos estágios posteriores até que todas as funções de todos os estágios estejam bem aproximadas.

Com o intuito de propor uma forma de trazer a informação das aproximações de um dado estágio para o estágio anterior, observa-se que a função de custo futuro pode ser escrita como a esperança condicional dos custos totais e cada custo total possui uma aproximação inferior por planos. Com isso, a esperança condicional da aproximação inferior por planos é uma cota inferior para a função de custo futuro. Com efeito, dada uma decisão viável x_{t-1}^* é possível gerar um refinamento da aproximação do custo futuro $\bar{\mathfrak{Q}}_t(\cdot, \xi^{[j]})$ observando que:

$$\begin{aligned}
\bar{Q}_t(x_{t-1}, \xi^{[i]}) &= \sum_{j \in S(i)} p_{ij} Q_t(x_{t-1}, \xi^{[j]}) \\
&\geq \sum_{j \in S(i)} p_{ij} \mathfrak{Q}_t(x_{t-1}, \xi^{[j]}) \\
&\geq \sum_{j \in S(i)} p_{ij} (\mathfrak{Q}_t(x_{t-1}^*, \xi^{[j]}) + g^j(x_{t-1} - x_{t-1}^*)) \\
&\geq \sum_{j \in S(i)} p_{ij} \mathfrak{Q}_t(x_{t-1}^*, \xi^{[j]}) + \sum_{j \in S(i)} (p_{ij} g^j) (x_{t-1} - x_{t-1}^*), \quad (\text{A.33}) \\
&\quad \forall x_{t-1}
\end{aligned}$$

onde g^j é o subgradiente de $\mathfrak{Q}_t(\cdot, \xi^{[j]})$ no ponto x_{t-1}^* , ou seja, $g^j \in \partial \mathfrak{Q}_t(x_{t-1}^*, \xi^{[j]})$. O plano definido por (A.33) é chamado de *corte médio* e pode ser agregado ao conjunto de planos que aproximam a função de custo futuro do estágio t :

$$\bar{\mathfrak{Q}}_t(x_{t-1}, \xi^{[i]}) := \max \left\{ \bar{\mathfrak{Q}}_t(x_{t-1}, \xi^{[i]}), \sum_{j \in S(i)} p_{ij} \mathfrak{Q}_t(x_{t-1}^*, \xi^{[j]}) + \sum_{j \in S(i)} (p_{ij} g^j) (x_{t-1} - x_{t-1}^*) \right\}$$

Após atualizar a aproximação das funções de custo futuro, atualiza-se a aproximação das funções de custo total:

$$\mathfrak{Q}_{t-1}(x_{t-2}, \xi^{[i]}) := \min_{\substack{B^i x_{t-2} + A^i x_{t-1} = b^i \\ x_t \geq 0}} c^i x_{t-1} + \bar{\mathfrak{Q}}_t(x_{t-1}, \xi^{[i]}).$$

Isso responde a pergunta de como refinar as aproximações das funções de custo total e futuro num *ponto* x_{t-1}^* .

Uma outra pergunta natural é como inicializar as aproximações das funções de custo futuro. Uma ideia simples é inicializá-las como *constant*es que são cotas inferiores para a função de custo futuro. Para isso é preciso verificar duas hipóteses:

1. As funções de custo futuro $\bar{Q}_t(\cdot, \xi^{[i]})$, $t = 2, \dots, T$, são limitadas inferiormente para cada realização da incerteza $\xi^{[i]}$; $i \in \mathcal{N}_{t-1}$.

2. O problema que define a função de custo total (A.18) sem a função de custo futuro deve ser limitado inferiormente para toda decisão viável x_{t-1} :

$$-\infty < \min_{\substack{B^j x_{t-1} + A^j x_t = b^j \\ x_t \geq 0}} c^j x_t \quad t = 1, \dots, T. \quad (\text{A.34})$$

Por fim, é preciso definir uma maneira de escolher os pontos x_{t-1}^* em torno dos quais se obterá uma aproximação linear para a função de custo futuro $\bar{Q}_t(\cdot, \xi^{[t]})$, como sugerido em (A.33). Vários algoritmos podem ser criados apenas escolhendo diferentes formas de se obter soluções viáveis x_{t-1}^* para cada estágio. Em geral, o algoritmo mais utilizado é o Nested Cutting Plane que se baseia em duas etapas e um teste de parada:

- i) **Etapa “forward”**: encontra-se soluções viáveis em torno das quais se obterá linearizações que aproximam inferiormente a função de custo futuro. Isto é feito percorrendo a árvore de cenários em ordem *crescente* de estágio e passando por *cada* vértice j de modo a obter a solução ótima x^j de cada problema de minimização que define $\mathfrak{Q}_t(x^{a(j)}, \xi^{[j]})$:

$$x^j \in \arg \min_{\substack{B^j x^{a(j)} + A^j x_t = b^j \\ x_t \geq 0}} c^j x_t + \bar{\mathfrak{Q}}_{t+1}(x_t, \xi^{[j]}),$$

onde $x^{a(j)}$ é a solução ótima do problema do vértice antecessor. Após calculadas todas as soluções viáveis de todos os vértices da árvore, estima-se uma cota superior para o custo total do primeiro estágio $Q_1(\xi^1)$ com $\bar{z} := \sum_{j \in \mathcal{N}} p^j c^j x^j$ (a afirmação “superior” será justificada adiante).

- ii) **Etapa “backward”**: calcula-se as aproximações lineares inferiores das funções de custo futuro centradas nas soluções viáveis obtidas na etapa forward. Isto é feito percorrendo a árvore de cenários em ordem *decrecente* de estágio, calculando em torno de cada ponto viável x^i a aproximação do custo total $\mathfrak{Q}_t(x^i, \xi^{[j]})$ e o respectivo subgradiente $g^j \in \partial \mathfrak{Q}_t(x^i, \xi^{[j]})$, para todo vértice j sucessor de i . É importante observar que no último estágio T é calculado o próprio custo total $Q_T(x^i, \xi^{[j]})$ e o respectivo subgradiente $g^j \in \partial Q_T(x^i, \xi^{[j]})$. Essas informações são utilizadas na construção do corte médio (A.33) que é agregado ao conjunto de estimativas da função de

custo futuro $\bar{\Omega}_t(\cdot, \xi^{[i]})$:

$$\bar{\Omega}_t(x_{t-1}, \xi^{[i]}) := \max \left\{ \bar{\Omega}_t(x_{t-1}, \xi^{[i]}), \sum_{j \in S(i)} p_{ij} \Omega_t(x^j, \xi^{[j]}) + \sum_{j \in S(i)} (p_{ij} g^j) (x_{t-1} - x^j) \right\}.$$

Em seguida, atualiza-se a aproximação das funções de custo total:

$$\Omega_{t-1}(x_{t-2}, \xi^{[i]}) := \min_{\substack{B^i x_{t-2} + A^i x_{t-1} = b^i \\ x_t \geq 0}} c^i x_{t-1} + \bar{\Omega}_t(x_{t-1}, \xi^{[i]}).$$

Após refinadas todas as aproximações das funções de custo futuro e total, estima-se uma cota inferior para o custo total do primeiro estágio $Q_1(\xi^1)$ com $\underline{z} := \Omega_1(\xi^1)$.

- iii) **Teste de parada:** Verifica-se a condição $\bar{z} - \underline{z} < \epsilon$. Caso verdadeira, o algoritmo termina. Caso falsa, o algoritmo retorna à etapa forward.

A demonstração de convergência do Nested Cutting Plane em um número finito de iterações utiliza fortemente o fato de que as funções de custo futuro e total são funções poliedrais. Um panorama da demonstração é dado observando que a estratégia para encontrar soluções viáveis na etapa forward gera ou a solução ótima, e com isso satisfaz ao critério de parada, ou um novo plano é acrescentado a cada aproximação da função de custo futuro na etapa backward. Supondo que em cada iteração um novo plano seja acrescentado, em um número finito de passos terão sido gerados todos os planos das funções de custo futuro do último estágio, ou seja, $\bar{\Omega}_T(\cdot, \cdot) = \bar{Q}_T(\cdot, \cdot)$. A partir daí, todos os planos gerados pelo corte médio (A.33) para as aproximações do custo futuro do estágio $T - 1$ serão planos tangentes e do mesmo modo em um número finito de passos terão sido gerados todos. Por indução, em um número finito de passos terão sido geradas todas as funções de custo total e futuro de todos os estágios e a partir deste momento a solução da etapa forward será ótima. Felizmente, na prática, a solução ótima é encontrada com muito menos planos. Uma explicação é que o algoritmo constroi iterativamente aproximações lineares das funções de custo total e futuro em torno de uma solução ótima e não em todos os pontos do seu domínio efetivo.

Uma das formas de verificar que $\bar{z} = \sum_{j \in N} p^j c^j x^j$ é uma cota superior para o problema original é *provar* que a função de custo total em um dado estágio t possui

a seguinte cota superior:

$$Q_t(x^i, \xi^{[j]}) \leq \frac{1}{p^j} \left(\sum_{k \in \mathcal{T}(j)} p^k c^k x^k \right), \quad (\text{A.35})$$

onde $\mathcal{T}(j)$ é o conjunto de vértices da sub-árvore com raiz em j . Para o primeiro estágio, a desigualdade (A.35) equivale a:

$$z^* = Q_1(\xi^1) \leq \frac{1}{p^1} \left(\sum_{k \in \mathcal{T}(1)} p^k c^k x^k \right) = \sum_{j \in \mathcal{N}} p^j c^j x^j = \bar{z}.$$

A demonstração da relação (A.35) pode ser feita por indução no estágio. Com efeito, para o último estágio T , a função de custo futuro $Q_{T+1}(\cdot, \cdot)$ é nula e por isso

$$Q_T(x^i, \xi^{[j]}) = \min_{\substack{B^j x^i + A^j x_T = b^j \\ x_T \geq 0}} c^j x_T \leq c^j x^j,$$

onde j é um vértice sucessor de i . Pois, pela construção da etapa forward, x^j é viável (ótima) para o problema que define $Q_T(x^i, \xi^{[j]})$, já que x^j é solução do problema de minimização que define $\mathfrak{Q}_T(x^i, \xi^{[j]})$ e ambos os problemas possuem o mesmo conjunto viável (e neste caso, mesma função objetivo). Supondo a desigualdade (A.35) válida para os vértices k do estágio $t + 1$:

$$Q_{t+1}(x^j, \xi^{[k]}) \leq \frac{1}{p^k} \left(\sum_{l \in \mathcal{T}(k)} p^l c^l x^l \right), \quad (\text{A.36})$$

mostrar-se que (A.35) também é válida para os vértices j do estágio t , completando assim a prova por indução. Inicialmente, observa-se que (A.36) induz uma cota superior para a função de custo futuro associada ao vértice j do estágio $t + 1$:

$$\begin{aligned} \bar{Q}_{t+1}(x^j, \xi^{[j]}) &= \sum_{k \in \mathcal{S}(j)} p_{jk} Q_{t+1}(x^j, \xi^{[j]}) \\ &\leq \sum_{k \in \mathcal{S}(j)} \frac{p_{jk}}{p^k} \left(\sum_{l \in \mathcal{T}(k)} p^l c^l x^l \right) \\ &= \sum_{k \in \mathcal{S}(j)} \frac{1}{p^j} \left(\sum_{l \in \mathcal{T}(k)} p^l c^l x^l \right), \end{aligned} \quad (\text{A.37})$$

onde na última igualdade (A.37) se utilizou que a probabilidade p_k de um vértice k é igual a probabilidade p_j do vértice j antecessor a k multiplicada pela probabilidade condicional p_{jk} do vértice k condicionado ao vértice j , ou seja, $p^k = p^j p_{jk}$. Por sua

vez, a cota (A.37) da função de custo futuro induz uma cota para a função de custo total:

$$\begin{aligned}
Q_t(x^i, \xi^{[j]}) &= \min_{\substack{B^j x^i + A^j x_t = b^j \\ x_t \geq 0}} c^j x_t + \bar{Q}_{t+1}(x_t, \xi^{[j]}) \\
&\leq c^j x^j + \bar{Q}_{t+1}(x^j, \xi^{[j]}) \\
&\leq c^j x^j + \sum_{k \in S(j)} \frac{1}{p^j} \left(\sum_{l \in \mathcal{T}(k)} p^l c^l x^l \right) \\
&= \frac{1}{p^j} \left(\sum_{k \in \mathcal{T}(j)} p^k c^k x^k \right),
\end{aligned}$$

onde x^j é uma solução viável do problema que define $Q_t(x^i, \xi^{[j]})$, visto que x^j é solução do problema de minimização que define $\mathfrak{Q}_t(x^i, \xi^{[j]})$ e ambos os problemas possuem o mesmo conjunto viável. Isto prova o passo de indução e portanto a relação (A.35).

Apesar do Nested Cutting Plane ser um algoritmo geral e de convergência finita, na prática funciona apenas para casos em que a árvore de cenários é pequena. Pois, no caso de uma árvore de cenários com T estágios cujos vértices possuem um número constante n de sucessores, o número total de vértices é $1 + n^1 + n^2 + \dots + n^{T-1} = (n^T - 1)/(n - 1)$. Uma árvore dependente possui uma função de custo futuro por vértice e por isso o número de funções a serem estimadas cresce exponencialmente com relação ao número de estágio. Por exemplo, se o planejamento for para um horizonte de $T = 10$ estágios e com um número de sucessores constante $n = 20$, o número de funções de custo futuro a serem estimadas é $(20^{10} - 1)/(20 - 1) \approx 5,38 \times 10^{11}$. Se a resolução de cada problema referente a aproximação do custo total (A.31) demora em média 0,01 segundos, o tempo médio para a execução da etapa forward seria em torno de 170 anos.

No caso de uma árvore de cenários independente, o número de funções de custo futuro a serem estimadas cai para $T - 1$, uma em cada estágio $\{\bar{Q}_t(\cdot) \mid t = 2, \dots, T\}$, e o custo total se reduz a:

$$Q_t(x_{t-1}, \xi^j) = \min_{\substack{B^j x_{t-1} + A^j x_t = b^j \\ x_t \geq 0}} c^j x_t + \bar{Q}_{t+1}(x_t),$$

como descrito em (A.27). Com isso, as aproximações inferiores das funções de custo futuro são *únicas* por estágio $\{\bar{\mathfrak{Q}}_t(\cdot) \mid t = 1, \dots, T\}$ e as do custo total são tantas

quanto forem o número de aberturas de cada estágio:

$$\Omega_t(x_{t-1}, \xi^j) = \min_{\substack{B^j x_{t-1} + A^j x_t = b^j \\ x_t \geq 0}} c^j x_t + \bar{\Omega}_{t+1}(x_t).$$

Apesar dessa redução de complexidade gerada pela estrutura especial de uma árvore independente, o Nested Cutting Plane ainda é impraticável no caso de muitos estágios, pois tal algoritmo visita todos os vértices da árvore de cenários na etapa forward e backward e neste caso o número de vértices também cresce exponencialmente com relação ao número de estágios.

A.6.2 Aplicação: problema do planejamento de longo prazo da operação

A finalidade desta seção é mostrar como o algoritmo Nested Cutting Plane pode ser aplicado ao problema de planejamento de longo prazo da operação visto na seção A.4. Vale ressaltar que, na prática, o tempo computacional necessário para se resolver um problema de otimização estocástica com o algoritmo Nested Cutting Plane pode ser proibitivo, quando considerado um número moderado de estágios.

Como visto ao longo da seção A.6.1, os requisitos para a aplicação do Nested Cutting Plane são:

- 1) O recurso deve ser relativamente completo, ou seja, toda decisão viável do problema associado a função de custo total de um estágio deve produzir um problema viável no estágio posterior;
- 2) As funções de custo futuro devem ser limitadas inferiormente, isto é, deve existir uma constante $M \in \mathbb{R}$ tal que $\bar{Q}_t(\cdot, \cdot) \geq M$.
- 3) O problema que define a função de custo total sem o termo da função de custo futuro deve ser limitado inferiormente:

$$-\infty < \min_{\substack{B_t x_{t-1} + A_t x_t = b_t \\ x_t \geq 0}} c_t x_t \quad ; \quad (\text{A.38})$$

- 4) Deve-se explicitar o cálculo dos subgradientes da aproximação da função de custo total.

Para verificar 2) e 3) é necessário relembrar a formulação por programação

dinâmica do problema de planejamento de longo prazo da operação:

$$\begin{aligned}
Q_t(v_t, \mathbf{a}_{[t]}) = \min & \sum_{k \in K} \sum_{j \in T_k} c_j g_{tj} + \beta \bar{Q}_{t+1}(v_{t+1}, \mathbf{a}_{[t]}) \\
\text{s.a.} & v_{t+1} = v_t + \mathbf{a}_t - q_t - s_t \\
& q_{tk} + \sum_{j \in T_k} g_{tj} + \sum_{l \in \Omega_k} (f_{tlk} - f_{tkl}) = d_{tk}, \quad \forall k \in K \\
& 0 \leq v_{t+1} \leq \bar{v}, \quad 0 \leq q_t \leq \bar{q}, \quad 0 \leq s_t, \\
& q_t + s_t \geq \underline{q}, \quad \underline{g} \leq g_t \leq \bar{g}, \quad \underline{f} \leq f_t \leq \bar{f} \\
\bar{Q}_{t+1}(v_{t+1}, \mathbf{a}_{[t]}) = & \begin{cases} \mathbb{E}[Q_{t+1}(v_{t+1}, \mathbf{a}_{[t+1]}) \mid \mathbf{a}_{[t]}] & , t \in \{1, \dots, T-1\} \\ 0 & , t = T \end{cases},
\end{aligned}$$

para todo $t = 1, \dots, T$. No último estágio T , a função de custo total $Q_T(v_T, \mathbf{a}_{[T]})$ é positiva, pois a respectiva função objetivo compreende apenas o custo das térmicas e este custo é positivo. Como cada função de custo total $Q_T(v_T, \mathbf{a}_{[T]})$ é positiva, a função de custo futuro $\bar{Q}_T(v_T, \mathbf{a}_{[T-1]})$ também é, pois é a esperança condicional de $Q_T(v_T, \mathbf{a}_{[T]})$. No penúltimo estágio $T-1$, a função de custo total $Q_{T-1}(v_{T-1}, \mathbf{a}_{[T-1]})$ é também maior do que ou igual a zero, pois a respectiva função objetivo compreende o custo das térmicas e também o custo futuro $\bar{Q}_T(v_T, \mathbf{a}_{[T-1]})$. Por indução no estágio, mostra-se que as funções de custo total e futuro são todas positivas e por isso $M = 0$ atende ao requisito 2). Além disso, o requisito 3) é atendido, pois 0 é também o limite inferior de (A.38).

Com relação ao requisito 1), é preciso verificar se o problema referente à função de custo total $Q_t(v_t, \mathbf{a}_{[t]})$ é viável para todo par viável $(v_t, \mathbf{a}_{[t]})$, ou seja, para qualquer energia armazenável que respeite os limites operativos $0 \leq v_t \leq \bar{v}$ e para qualquer afluência energética $\mathbf{a}_{[t]}$ que o modelo estocástico considere. Porém, a restrição de defluência mínima $q_t + s_t \geq \underline{q}$ pode não ser atendida caso cenários de baixa afluência ocorram. Uma maneira de contornar este problema é considerar esta restrição na função objetivo com uma penalidade associada ou implementar a abordagem por cortes de viabilidade do Nested Cutting Plane (RUSZCZYŃSKI [56]). Entretanto, isso não resolve a questão de modelagem, pois a restrição $q_t + s_t \geq \underline{q}$ não é garantidamente satisfeita, podendo resultar em um custo alto e artificial, no caso da penalidade, ou num certificado de inviabilidade, no caso dos cortes de viabilidade. Na opinião do autor, é necessária uma análise mais profunda para que se possa descrever esta restrição de maneira mais coerente. Como não existe nenhuma sugestão satisfatória na literatura, adota-se a formulação por penalidades que é a

abordagem tradicional do setor:

$$\begin{aligned}
Q_t(v_t, \mathbf{a}_{[t]}) = & \min \sum_{k \in K} \sum_{j \in T_k} c_j g_{tj} + \rho_t^\top u_t + \beta \bar{Q}_{t+1}(v_{t+1}, \mathbf{a}_{[t]}) \\
s.a. & v_{t+1} = v_t + \mathbf{a}_t - q_t - s_t \\
& q_{tk} + \sum_{j \in T_k} g_{tj} + \sum_{l \in \Omega_k} (f_{tlk} - f_{tkl}) = d_{tk}, \quad \forall k \in K \quad (\text{A.39}) \\
& 0 \leq v_{t+1} \leq \bar{v}, \quad 0 \leq q_t \leq \bar{q}, \quad 0 \leq s_t, \quad 0 \leq u_t, \\
& q_t + s_t + u_t \geq \underline{q}, \quad \underline{g} \leq g_t \leq \bar{g}, \quad \underline{f} \leq f_t \leq \bar{f} \\
\bar{Q}_{t+1}(v_{t+1}, \mathbf{a}_{[t]}) = & \begin{cases} \mathbb{E}[Q_{t+1}(v_{t+1}, \mathbf{a}_{[t+1]}) \mid \mathbf{a}_{[t]}] & , t \in \{1, \dots, T-1\} \\ 0 & , t = T \end{cases} ,
\end{aligned}$$

onde o vetor de decisão é dado por $x_t = (v_{t+1}, q_t, s_t, g_t, f_t, u_t)$ e ρ_t é o vetor de penalidades. Observa-se que as análises feitas para os requisitos 2) e 3) continuam válidas para esta nova formulação, pois apenas um custo positivo foi acrescentado e por isso todo argumento segue de maneira análoga.

Considerando essa nova formulação por penalidades, é possível satisfazer ao requisito 1) admitindo a hipótese de que, para a geração térmica mínima e o intercâmbio mínimo, o total de geração térmica (sem deficit) somado ao intercâmbio líquido é menor ou igual do que a carga de cada subsistema e estágio, ou seja,

$$\sum_{j \in T_k \setminus \{e_k\}} \underline{g}_j + \sum_{l \in \Omega_k} (\underline{f}_{lk} - \underline{f}_{kl}) \leq d_{tk}, \quad \forall k \in K,$$

para todo $t = 1, \dots, T$. Com isso, a partir de qualquer par viável $(v_t, \mathbf{a}_{[t]})$ é possível produzir um vetor de decisão viável da seguinte forma:

- mantendo a mesma energia armazenada ($v_{t+1} = v_t$);
- não turbinando nada ($q_t = 0$);
- vertendo tudo que afluir ($s_t = \mathbf{a}_t$);
- produzindo a geração térmica mínima ($g_{tj} = \underline{g}_j$), para toda térmica j diferente do deficit e_k ;
- produzindo o intercâmbio mínimo ($f_t = \underline{f}$)
- atendendo a carga restante $d_{tk} - \sum_{j \in T_k \setminus \{e_k\}} \underline{g}_j + \sum_{l \in \Omega_k} (\underline{f}_{lk} - \underline{f}_{kl})$ com deficit $g_{e_k t}$;
- e definindo a variável de folga u_t usual: $u_t = \max\{\underline{q} - q_t - s_t, 0\}$.

Obviamente esta operação não será escolhida pelo modelo, mas o vetor de decisão assim definido satisfaz as restrições do problema referente à função de custo total

$Q_t(v_t, \mathbf{a}_{[t]})$. Isso mostra que o problema com penalidades é relativamente completo, atendendo ao requisito 1).

Por último, para atender ao requisito 4) é oportuno relembrar o algoritmo Nested Cutting Plane. Como mencionado, é necessário que o processo estocástico $\{\mathbf{a}_t\}_{t=1}^T$ seja discretizado e representado por uma árvore de cenários. Além disso, inicializa-se as aproximações das funções de custo futuro por constantes que são cotas inferiores das funções originais. Estas constantes são 0 para o planejamento de longo prazo, ou seja, inicialmente $\bar{\mathfrak{Q}}_{t+1}(\cdot, \cdot) := 0$. Numa iteração qualquer, a aproximação da função de custo futuro é dada por

$$\bar{\mathfrak{Q}}_{t+1}(v_{t+1}, a^{[j]}) = \max_{l \in L^j} (r_l^j v_{t+1} + \gamma_l^j)$$

e a aproximação da função de custo total é

$$\begin{aligned} \mathfrak{Q}_t(v_t, a^{[j]}) = & \min \sum_{k \in K} \sum_{j \in T_k} c_j g_{tj} + \rho_t^\top u_t + \beta \bar{\mathfrak{Q}}_{t+1}(v_{t+1}, a^{[j]}) \\ \text{s.a. } & v_{t+1} = v_t + a^j - q_t - s_t \\ & q_{tk} + \sum_{j \in T_k} g_{tj} + \sum_{l \in \Omega_k} (f_{tlk} - f_{tkl}) = d_{tk}, \quad \forall k \in K . \\ & 0 \leq v_{t+1} \leq \bar{v}, \quad 0 \leq q_t \leq \bar{q}, \quad 0 \leq s_t, \quad 0 \leq u_t, \\ & q_t + s_t + u_t \geq \underline{q}, \quad \underline{g} \leq g_t \leq \bar{g}, \quad \underline{f} \leq f_t \leq \bar{f} \end{aligned}$$

Com o intuito de facilitar o cálculo da aproximação da função de custo total e os respectivos subgradientes, é interessante reformular sua definição como um problema de programação linear:

$$\begin{aligned} \mathfrak{Q}_t(v_t, a^{[j]}) = & \min \sum_{k \in K} \sum_{j \in T_k} c_j g_{tj} + \rho_t^\top u_t + \beta \bar{\mathfrak{Q}}_{t+1} \\ \text{s.a. } & v_{t+1} = v_t + a^j - q_t - s_t \\ & q_{tk} + \sum_{j \in T_k} g_{tj} + \sum_{l \in \Omega_k} (f_{tlk} - f_{tkl}) = d_{tk}, \quad \forall k \in K \\ & 0 \leq v_{t+1} \leq \bar{v}, \quad 0 \leq q_t \leq \bar{q}, \quad 0 \leq s_t, \quad 0 \leq u_t, \\ & q_t + s_t + u_t \geq \underline{q}, \quad \underline{g} \leq g_t \leq \bar{g}, \quad \underline{f} \leq f_t \leq \bar{f} \\ & r_l^j v_{t+1} + \gamma_l^j \leq \bar{\mathfrak{Q}}_{t+1}, \quad \forall l \in L^j \end{aligned} \quad , \quad (\text{A.40})$$

onde agora as variáveis de decisão são $x_t = (v_{t+1}, q_t, s_t, g_t, f_t, u_t, \bar{\mathfrak{Q}}_{t+1})$.

A seguir, apresenta-se uma fórmula de referência para o cálculo de subgradientes

dada pelo valor ótimo de um problema de programação linear:

$$\begin{aligned}
V(x) = \min_y \quad & \tilde{c}^\top y \\
\text{s.a.} \quad & A_1 y = b_1 \quad (\pi_{b_1}) \\
& A_2 y \leq b_2 \quad (\pi_{b_2}) , \\
& T_1 x + W_1 y = h_1 \quad (\pi_{h_1}) \\
& T_2 x + W_2 y \leq h_2 \quad (\pi_{h_2})
\end{aligned} \tag{A.41}$$

tal que y é a variável de decisão de (A.41), $(\pi_{b_1}, \pi_{b_2}, \pi_{h_1}, \pi_{h_2})$ são as soluções duais ótimas associadas as restrições indicadas e x é a variável de estado, isto é, a variável de interesse que uma vez fixada define as restrições do problema (A.41). Abaixo seguem as propriedades da função $V(x)$.

Fato 5: $V(\cdot)$ é uma função poliedral. Quando $V(x)$ é finito, o subdiferencial de $V(\cdot)$ no ponto x é:

$$\begin{aligned}
\partial V(x) = & - [0 \ 0 \ T_1^\top \ T_2^\top] \widehat{\mathcal{D}}(x), \\
= & \left\{ -T_1^\top \pi_{h_1} - T_2^\top \pi_{h_2} \mid \begin{array}{l} \pi_{h_1} \text{ e } \pi_{h_2} \text{ são as soluções duais ótimas} \\ \text{de } T_1 x + W_1 y = h_1 \text{ e } T_2 x + W_2 y \leq h_2 \end{array} \right\},
\end{aligned} \tag{A.42}$$

onde $\widehat{\mathcal{D}}(x)$ é o conjunto das soluções duais ótimas $(\pi_{b_1}, \pi_{b_2}, \pi_{h_1}, \pi_{h_2})$ associadas a (A.41).

Com este resultado, o que precisa ser feito para o cálculo dos subgradientes de $\mathcal{Q}_t(v_t, a^{[j]})$ é identificar as restrições de (A.40) com os blocos do problema (A.41). Os blocos $A_1 y = b_1$ e $A_2 y \leq b_2$ compreendem às restrições de igualdade e desigualdade que não possuem relação com o estado, respectivamente, e os blocos $T_1 x + W_1 y = h_1$ e $T_2 x + W_2 y \leq h_2$ compreendem as restrições que possuem relação com o estado x . De forma a facilitar a construção desses blocos, é interessante que *todas* as restrições de (A.40) sejam escritas em notação vetorial e não em notação indicial.

Notação indicial	Notação vetorial
$\sum_{k \in K} \sum_{j \in T_k} c_j g_{tj}$	$c^\top g_t$
$q_{tk} + \sum_{j \in T_k} g_{tj} + \sum_{l \in \Omega_k} (f_{tlk} - f_{tkl}) = d_{tk}, \quad \forall k \in K$	$q_t + M_I g_t + M_D f_t = d_t$
$r_l^j v_{t+1} + \gamma_l^j \leq \bar{\mathcal{Q}}_{t+1}, \quad \forall l \in L^j$	$R^j v_{t+1} + \gamma^j \leq \bar{\mathcal{Q}}_{t+1} \cdot \mathbf{1}$

Tabela A.4: Conversão de notação indicial para vetorial.

Seguindo a notação descrita na tabela A.4, onde $\mathbf{1} := (1, \dots, 1)^\top \in \mathbb{R}^{|K|}$ e $|K|$ é a cardinalidade do conjunto de subsistemas K , é possível reformular o problema

(A.40) com restrições apenas vectoriais:

$$\begin{aligned}
\mathfrak{Q}_t(v_t, a^{[j]}) = \min \quad & c^\top g_t + \rho_t^\top u_t + \beta \bar{\mathfrak{Q}}_{t+1} \\
s.a. \quad & v_{t+1} = v_t + a^j - q_t - s_t \\
& q_t + M_I g_t + M_D f_t = d_t \\
& 0 \leq v_{t+1} \leq \bar{v}, \quad 0 \leq q_t \leq \bar{q}, \quad 0 \leq s_t, \quad 0 \leq u_t, \\
& q_t + s_t + u_t \geq \underline{q}, \quad \underline{g} \leq g_t \leq \bar{g}, \quad \underline{f} \leq f_t \leq \bar{f} \\
& R^j v_{t+1} + \gamma^j \leq \bar{\mathfrak{Q}}_{t+1} \cdot \mathbf{1}
\end{aligned} \tag{A.43}$$

Com o intuito de descrever o problema (A.43) nos moldes de (A.41), observa-se que:

- a variável de decisão y é igual a $(v_{t+1}, q_t, s_t, g_t, f_t, u_t, \bar{\mathfrak{Q}}_{t+1})$;
- a variável de estado x é igual a $x = v_t$;
- o custo associado à decisão y é

$$\tilde{c} = \begin{pmatrix} v_{t+1} & q_t & s_t & g_t & f_t & u_t & \bar{\mathfrak{Q}}_{t+1} \\ 0 & 0 & 0 & c & 0 & \rho_t & \beta \end{pmatrix};$$

- o bloco referente à restrição de igualdade $A_1 y = b_1$ é

$$A_1 = \begin{pmatrix} v_{t+1} & q_t & s_t & g_t & f_t & u_t & \bar{\mathfrak{Q}}_{t+1} \\ 0 & I & 0 & M_I & M_D & 0 & 0 \end{pmatrix}, \quad b_1 = d_t$$

- o bloco referente à restrição de desigualdade $A_2 y \leq b_2$ é

$$A_2 = \begin{pmatrix} v_{t+1} & q_t & s_t & g_t & f_t & u_t & \bar{\Omega}_{t+1} \\ I & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I & 0 & 0 \\ -I & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -I & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -I & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -I & 0 \\ 0 & -I & -I & 0 & 0 & -I & 0 \\ 0 & 0 & 0 & -I & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -I & 0 & 0 \\ R^j & 0 & 0 & 0 & 0 & 0 & -\mathbf{1} \end{pmatrix}, \quad b_2 = \begin{pmatrix} \bar{v} \\ \bar{q} \\ \bar{g} \\ \bar{f} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ -\underline{q} \\ -\underline{g} \\ -\underline{f} \\ -\gamma^j \end{pmatrix}$$

- o bloco referente à restrição de igualdade $T_1 x + W_1 y = h_1$ é

$$T_1 = \begin{pmatrix} v_t \\ -I \end{pmatrix}, \quad W_1 = \begin{pmatrix} v_{t+1} & q_t & s_t & g_t & f_t & u_t & \bar{\Omega}_{t+1} \\ I & I & I & 0 & 0 & 0 & 0 \end{pmatrix}, \quad h_1 = \mathbf{a}_t$$

- o bloco referente à restrição de desigualdade $T_1 x + W_1 y \leq h_1$ não existe, pois neste caso específico, não existe nenhuma restrição de desigualdade que envolva v_t e as outras variáveis.

Portanto, pelo Fato 5), obtém-se a seguinte expressão para o subdiferencial de $\Omega_t(v_t, a^{[j]})$:

$$\begin{aligned} \partial \Omega_t(v_t, a^{[j]}) &= - [0 \ 0 \ T_1^\top] \widehat{\mathcal{D}}(x) = \\ &= \left\{ \pi_v \in \mathbb{R}^{|K|} \left| \begin{array}{l} \pi_v \text{ é soluções dual ótima da restrição} \\ \text{hidráulica } v_{t+1} = v_t + a^j - q_t - s_t \text{ de (A.43)} \end{array} \right. \right\}, \end{aligned}$$

Com isso, está atendido o requisito 4) e é possível aplicar o Nested Cutting Plane ao problema de planejamento de longo prazo da operação reformulado (A.39).

A.6.3 Algoritmo SDDP

Como mencionado ao final da seção sobre o Nested Cutting Plane (A.6.1), a aplicação desse algoritmo é impraticável no caso de muitos estágios, mesmo quando a árvore de cenários é uma árvore independente, na qual o número de funções de custo futuro

é reduzido para uma por estágio. A razão do Nested Cutting Plane ser impraticável é o fato dele visitar todos os vértice na etapa forward e backward.

Um aperfeiçoamento do Nested Cutting Plane para o caso particular de uma árvore independente é um algoritmo que não visita todos os vértices na etapa forward e backward, mas sim *alguns* deles. Com base neste ideia, uma estratégia consiste em sortear, na etapa forward, alguns cenários obtendo do primeiro ao último estágio soluções viáveis ao longo das observações sorteadas e, na etapa backward, percorrer tais observações do último ao primeiro estágio refinando a aproximação da *única* função de custo futuro de cada estágio em torno da solução viável obtida.

Relembrando a descrição da seção A.5 de árvore de Cenários, um cenário é uma observação ao longo de todos estágio, isto é, é uma realização da variável aleatória $\xi_{[T]}$. Seguindo a notação de árvore, o conjunto de valores que a variável aleatória $\xi_{[T]}$ pode assumir é $\{\xi^{[j]} \mid j \in \mathcal{N}_T\}$ e o respectivo conjunto de probabilidades é $\{p^j \mid j \in \mathcal{N}_T\}$. Além disso, denota-se por $\xi_{[t]}$ a variável aleatória associada aos t primeiros estágios de $\xi_{[T]}$, cujo o conjunto de realizações e probabilidades associadas são $\{\xi^{[j_t]} \mid j \in \mathcal{N}_T\}$ e $\{p^{j_t} \mid j \in \mathcal{N}_T\}$, respectivamente, onde j_t é o ancestral do vértice j no estágio t .

Com o intuito de descrever as soluções viáveis obtidas ao longo dos cenários sorteados, denota-se por $x_t(\xi_{[t]})$ a solução ótima encontrada no estágio t para o problema que define a aproximação do custo total $\mathfrak{Q}_t(x_{t-1}(\xi_{[t-1]}), \xi_t)$:

$$x_t(\xi_{[t]}) \in \arg \min_{\substack{B_t x_{t-1}(\xi_{[t-1]}) + A_t x_t = b_t \\ x_t \geq 0}} c_t^\top x_t + \bar{\mathfrak{Q}}_{t+1}(x_t),$$

onde $x_{t-1}(\xi_{[t-1]})$ é uma solução ótima encontrada no estágio anterior, lembrando que ξ_t é a variável aleatória que representa a incerteza nos parâmetros (c_t, A_t, B_t, b_t) .

Neste contexto de soluções viáveis sorteadas, a ideia de se refinar a aproximação da função de custo futuro por meio do *corte médio* ao longo das decisões encontradas na etapa forward precisa ser reexaminada. Seja $\xi^{[i]}$ o cenário sorteado na etapa forward, isto é, $\xi_{[T]} = \xi^{[i]}$, onde i é um vértice do último estágio T . De forma semelhante, a média dos planos tangentes das aproximações da função de custo

total é uma cota inferior para a função de custo futuro:

$$\begin{aligned}
\bar{Q}_t(x_{t-1}) &= \sum_{j \in S(i_{t-1})} p_{i_{t-1}j} Q_t(x_{t-1}, \xi^j) \\
&\geq \sum_{j \in S(i_{t-1})} p_{i_{t-1}j} \mathfrak{Q}_t(x_{t-1}, \xi^j) \\
&\geq \sum_{j \in S(i_{t-1})} p_{i_{t-1}j} (\mathfrak{Q}_t(x_{t-1}^*, \xi^j) + g^j(x_{t-1} - x_{t-1}^*)) \\
&\geq \sum_{j \in S(i_{t-1})} p_{i_{t-1}j} \mathfrak{Q}_t(x_{t-1}^*, \xi^j) + \sum_{j \in S(i_{t-1})} (p_{i_{t-1}j} g^j) (x_{t-1} - x_{t-1}^*) \quad (\text{A.44}) \\
&\forall x_{t-1},
\end{aligned}$$

onde i_{t-1} é o ancestral do vértice i referente ao estágio $t - 1$ e x_{t-1}^* é uma solução viável obtida no estágio $t - 1$ para o cenário $\xi^{[i]}$, ou seja, $x_{t-1}^* = x_{t-1}(\xi^{[i_{t-1}]})$.

Portanto, é possível calcular soluções viáveis ao longo dos cenários *sorteados* e obter aproximações inferiores em torno destas soluções para cada função de custo futuro via corte médio. Porém, perde-se com isso a cota superior $\bar{z} = \sum_{j \in \mathcal{N}} p^j c^j x^j$, onde x^j é a solução viável obtida no vértice j da árvore de cenários

$$x^j \in \arg \min_{\substack{B^j x^{a(j)} + A^j x_t = b^j \\ x_t \geq 0}} c_t^\top x_t + \bar{\mathfrak{Q}}_{t+1}(x_t)$$

pois, para se obter \bar{z} é necessário calcular soluções viáveis para *todos* os vértices da árvore. Desta forma, a regra de parada precisa ser reavaliada.

Uma forma de superar esta dificuldade é construindo um estimador para a cota superior $\bar{z} = \sum_{j \in \mathcal{N}} p^j c^j x^j$. Uma inspiração para essa construção é a seguinte identidade:

$$\mathbb{E} \left[\sum_{t=1}^T c_t x_t(\xi_{[t]}) \right] = \sum_{j \in \mathcal{N}} p^j c^j x^j. \quad (\text{A.45})$$

Supondo a relação (A.45), seja $\{\xi_{[T],n}\}_{n=1}^N$ uma amostra i.i.d dos cenários da árvore de cenários, $c_{t,n}$ a componente de custos de $\xi_{t,n}$ e $x_{t,n}$ a solução obtida no estágio t para $\xi_{[T],n}$, ou seja, $x_{t,n} := x_t(\xi_{[t],n})$. Então, pela Lei dos Grandes Números, o estimador $\bar{z}_N := \frac{1}{N} \sum_{n=1}^N \left(\sum_{t=1}^T c_{t,n} x_{t,n} \right)$ converge quase certamente (q.c) para a cota superior desejada:

$$\frac{1}{N} \sum_{n=1}^N \left(\sum_{t=1}^T c_{t,n} x_{t,n} \right) \xrightarrow{q.c} \mathbb{E} \left[\sum_{t=1}^T c_t x_t(\xi_{[t]}) \right]. \quad (\text{A.46})$$

Portanto, \bar{z}_N é um estimador consistente para \bar{z} . Adicionalmente, uma forma de

estimar a incerteza do estimador \bar{z}_N é pelo seu desvio padrão amostral

$$\sigma_N^2 := \frac{1}{N} \sum_{n=1}^N \left(\bar{z}_N - \sum_{t=1}^T c_{t,n} x_{t,n} \right)^2$$

ou por seu intervalo de confiança 95% assintoticamente normal $[\bar{z}_N - 1.96\sigma_N, \bar{z}_N + 1.96\sigma_N]$.

Para mostrar a relação (A.45), basta decompor a esperança em uma soma de esperanças condicionais multiplicadas pelas respectivas probabilidades. Esta forma de decompor a esperança é chamada de Lei das Expectativas Iteradas. Com efeito, pela linearidade da esperança, o lado esquerdo de (A.45) pode ser decomposto em uma soma:

$$\mathbb{E} \left[\sum_{t=1}^T c_t x_t(\xi_{[t]}) \right] = \sum_{t=1}^T \mathbb{E} [c_t x_t(\xi_{[t]})]$$

e, pela Lei das Expectativas Iteradas, cada componente pode ser reescrita por

$$\begin{aligned} \mathbb{E} [c_t x_t(\xi_{[t]})] &= \sum_{j \in \mathcal{N}_t} \mathbb{E} [c_t x_t(\xi_{[t]}) \mid \xi_{[t]} = \xi^{[j]}] \mathbb{P} (\xi_{[t]} = \xi^{[j]}) \\ &= \sum_{j \in \mathcal{N}_t} \mathbb{E} [c_t x_t(\xi_{[t]}) \mid \xi_{[t]} = \xi^{[j]}] p^j \\ &= \sum_{j \in \mathcal{N}_t} p^j c^j x^j \end{aligned}$$

onde a identidade $\mathbb{E} [c_t x_t(\xi_{[t]}) \mid \xi_{[t]} = \xi^{[j]}] = c^j x^j$ decorre ao observar que $c^j = c_t(\xi^j)$ e $x^j = x_t(\xi^{[j]})$. Somando do estágio 1 até o horizonte T , obtém-se o lado direito da equação (A.45).

O algoritmo que leva em consideração essa amostragem dos cenários para o cálculo de soluções viáveis e a dada modificação no cálculo do corte médio para a aproximação da função de custo futuro numa árvore de cenários independente é chamado de SDDP, (PEREIRA e PINTO [50]). Este algoritmo pode ser descrito em duas etapas e um critério de parada:

- i) **Etapa “forward”:** encontra-se soluções viáveis em torno das quais se obterá linearizações que aproximam inferiormente a função de custo futuro. Isto é feito *sorteando* cenários $\{\xi_{[T],n}\}_{n=1}^N$ da árvore e percorrendo-os em ordem crescente de estágio de modo a obter a solução ótima $x_{t,n}^* = x_t(\xi_{[t],n})$ de cada problema de minimização

que define $\mathfrak{Q}_t(x_{t-1,n}^*, \xi_{t,n})$:

$$\arg \min_{\substack{B_{t,n}x_{t-1,n}^* + A_{t,n}x_t = b_{t,n} \\ x_t \geq 0}} c_{t,n}x_t + \bar{\mathfrak{Q}}_{t+1}(x_t),$$

onde $x_{t-1,n}^*$ é a solução ótima do problema do estágio anterior, ou seja, $x_{t-1,n}^* = x_{t-1}(\xi_{[t-1],n})$. Após calculadas todas as soluções viáveis de todos cenários sorteados, constrói-se um estimador \bar{z}_N para a cota superior do custo total do primeiro estágio $Q_1(\xi^1)$.

- ii) **Etapa “backward”**: calcula-se as aproximações lineares inferiores das funções de custo futuro centradas nas soluções viáveis obtidas na etapa forward. Isto é feito percorrendo cada cenário *sorteado* $\xi^{[i]}$ em ordem *decrecente* de estágio, calculando em torno de cada ponto viável $x_{t-1}^* = x_{t-1}(\xi^{[i_{t-1}]})$ a aproximação do custo total $\mathfrak{Q}_t(x_{t-1}^*, \xi^j)$ e o respectivo subgradiente $g^j \in \partial \mathfrak{Q}_t(x_{t-1}^*, \xi^j)$, para todo vértice j sucessor de i_{t-1} . Essas informações são utilizadas na construção do corte médio (A.44) que é agregado ao conjunto de estimativas da função de custo futuro $\bar{\mathfrak{Q}}_t(\cdot)$:

$$\begin{aligned} \bar{\mathfrak{Q}}_t(x_{t-1}) := & \max \left\{ \bar{\mathfrak{Q}}_t(x_{t-1}), \right. \\ & \left. \sum_{j \in S(i_{t-1})} p_{i_{t-1}j} \mathfrak{Q}_t(x_{t-1}^*, \xi^j) + \sum_{j \in S(i_{t-1})} (p_{i_{t-1}j} g^j) (x_{t-1} - x_{t-1}^*) \right\} \end{aligned} \quad (\text{A.47})$$

Após refinadas todas as aproximações das funções de custo futuro e total, estima-se uma cota inferior para o custo total do primeiro estágio $Q_1(\xi^1)$ com $\underline{z} := \mathfrak{Q}_1(\xi^1)$.

- iii) **Teste de parada**: Verifica-se a condição $\underline{z} \in [\bar{z}_N - 2\sigma_N, \bar{z}_N + 2\sigma_N]$. Caso verdadeira, o algoritmo termina. Caso falsa, o algoritmo retorna à etapa forward.

A justificativa do artigo original PEREIRA e PINTO [50] para este critério de parada é que a cota inferior \underline{z} deve ser comparada com a precisão do estimador \bar{z}_N da cota superior \bar{z} . Entretanto, este critério de parada pode ser um tanto otimista, como descrito em SHAPIRO [62]. Um critério de parada menos otimista para o SDDP pode ser a estabilização da cota inferior \underline{z} ao longo das iterações:

$$\frac{\underline{z}_{\text{new}} - \underline{z}_{\text{old}}}{|\underline{z}_{\text{old}}|} < \Delta,$$

onde $\underline{z}_{\text{old}}$ é uma cota inferior obtida em uma dada iteração, $\underline{z}_{\text{new}}$ é uma outra cota

obtida em uma iteração posterior e Δ é o parâmetro do ganho relativo usado para definir a estabilização das cotas. É importante observar que a estabilização das cotas inferiores não significa que o problema tenha sido resolvido à otimalidade e sim que mais iterações do algoritmo SDDP não melhorariam significativamente a qualidade da aproximação das funções de custo futuro.

Extensão do algoritmo SDDP - caso de árvore dependente

O leitor poderia se perguntar se esta estratégia de sortear cenários pode ser adaptada ao caso de uma árvore de cenários dependente. A princípio sim, mas no caso geral o número de funções de custo futuro é exponencial em relação ao número de estágios. Logo, esta estratégia amostral não produz nenhuma melhora significativa no caso geral, pois a qualidade da aproximação do custo futuro em um dado estágio depende da qualidade da aproximação do custo futuro dos estágios posteriores e com um número grande de estágios não se teria uma boa aproximação da solução do problema.

Entretanto, para árvores de cenários dependentes que possuam uma estrutura particular é possível estender o SDDP de maneira estatística. Por exemplo, é possível estender o algoritmo SDDP para árvores K -Markovianas, ou seja, uma árvore dependente de no máximo K termos passados. O artigo de PHILPOTT e DE MATOS [52] considera uma árvore 1-Markoviana, isto é, uma cadeia de Markov de estados e tempos discretos.

Extensão da regra de decisão obtida via SDDP

Considerando a definição da aproximação da função de custo futuro para o caso de uma árvore independente, é possível estendê-la de forma simples para outros cenários $\xi_{[T]}$ que não estão contidos na árvore de cenários apenas substituindo ξ_t na sua expressão original:

$$\Omega_t(x_{t-1}, \xi_t) := \min_{\substack{B_t x_{t-1} + A_t x_t = b_t \\ x_t \geq 0}} c_t^\top x_t + \bar{\Omega}_{t+1}(x_t) \quad t = 1, \dots, T. \quad (\text{A.48})$$

Com isso, ao final do processo iterativo do SDDP, é possível estender as regras de decisão $x_t(\xi_{[t]})$ (política) para outros valores de incerteza $\xi_{[t]}$. A forma que isto é feito é análoga ao cálculo da etapa forward do SDDP para um cenário $\xi_{[T]}$:

$$x_t(\xi_{[t]}) \in \arg \min_{\substack{B_t x_{t-1}(\xi_{[t-1]}) + A_t x_t = b_t \\ x_t \geq 0}} c_t^\top x_t + \bar{\Omega}_{t+1}(x_t).$$

A.6.4 Aplicação: problema do planejamento de longo prazo da operação considerando os modelos PVAR e PVAR_m

Como observado na seção A.6.3, é necessário que o processo aleatório do problema de otimização estocástica possua componentes independentes para a aplicação do algoritmo SDDP. Porém, no caso do problema de planejamento de longo prazo da operação assumir que as afluências $\{\mathbf{a}_1, \mathbf{a}_2, \dots\}$ sejam independentes desconsidera uma característica importante do fenômeno que é a correlação temporal.

Uma maneira de incorporar a informação da correlação temporal da afluência sem violar a hipótese de independência necessária para o SDDP é construir um modelo de série temporal com *erro* independente para as afluências $\{\mathbf{a}_1, \mathbf{a}_2, \dots\}$ e incluir este modelo na formulação do problema de otimização estocástica. Isto é feito reformulando o problema original de modo a considerar a afluência \mathbf{a}_t como variável, a equação do modelo de série temporal como restrição e o erro independente como o *único* parâmetro aleatório.

O problema de otimização estocástica reformulado deve ser convexo, pois a propriedade de convexidade é fundamental para os algoritmos baseados em planos cortantes como é o caso do Nested Cutting Plane e do SDDP. Para garantir a convexidade do problema, é necessário que a equação que define o modelo de série temporal seja expressa por meio de uma fórmula linear, ou seja, a afluência \mathbf{a}_t deve ser uma função *afim* das afluências passadas. Uma restrição *não linear* de igualdade destrói a propriedade de convexidade de um problema de otimização matemática.

O modelo de série temporal considerado hoje em dia é um caso *particular* do modelo Periódico Vetorial Autorregressivo (PVAR), e em sua versão original é um modelo linear nas afluências e com erro aditivo independente. Apesar de satisfazer a linearidade, o artifício utilizado na distribuição de erro do modelo PVAR para garantir afluências sempre positivas destrói a independência das componentes de erro. Para maiores detalhes vide CEPEL [8].

Como apresentado no capítulo 2 - seção 2.2, o PVAR consiste em um modelo Vetorial Autorregressivo (VAR) para cada mês que leva em conta a correlação temporal e a sazonalidade do processo. Relembrando, o modelo PAR de período \mathbb{S} e ordem $\mathbf{p} = (p_1, \dots, p_{\mathbb{S}})$, PVAR(\mathbf{p}), para a afluência \mathbf{a}_t é definido por

$$\mathbf{a}_t = \zeta_t + \sum_{\nu=1}^{p_t} \Phi_{t,\nu} \mathbf{a}_{t-\nu} + \boldsymbol{\epsilon}_t, \quad (\text{A.49})$$

para todo inteiro positivo t , onde os erros *aditivos* $\{\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2, \dots\}$ são independentes. Para cada tempo t , os coeficientes $\Phi_{t,\nu}$ e as ordens do modelo são iguais módulo \mathbb{S} ; os erros são identicamente distribuídos módulo \mathbb{S} e com média zero. Em outras palavras:

- $\zeta_t = \zeta_{t+\mathbb{S}}$ e $\Phi_{t,\nu} = \Phi_{t+\mathbb{S},\nu}$;
- $p_t = p_{t+\mathbb{S}}$;
- ϵ_t possui a mesma distribuição de $\epsilon_{t+\mathbb{S}}$; e
- $\mathbb{E}[\epsilon_t] = 0$.

Destaca-se que o modelo PVAR apresentado é vetorial, pois $\{\mathbf{a}_1, \mathbf{a}_2, \dots\}$ é uma série vetorial. Consequentemente, $\Phi_{t,1}, \dots, \Phi_{t,p_t}$ são matrizes e ζ_t, ϵ_t são vetores.

Por ser um modelo com erro *aditivo*, a principal limitação do PVAR quando aplicado às afluências é a possibilidade de gerar valores negativos de afluência. Como mencionado no capítulo 2 - seção 2.5, as tentativas para contornar esse problema introduziram como efeito colateral dependência temporal dos ruídos.

Com o intuito de superar as dificuldades do PVAR, propôs-se no capítulo 3 - seção 3.1 o modelo Periódico Vetorial Autorregressivo Multiplicativo (PVAR_m). Em sua concepção, o PVAR_m representa a correlação temporal e a sazonalidade de maneira análoga ao PVAR, porém com erros *multiplicativos*. A modificação do conceito de erro aliada a restrições nos coeficientes que garantam previsões sempre positivas possibilita ao PVAR_m superar as limitações apresentadas pelo PVAR.

Relembrando, o modelo PVAR_m de período \mathbb{S} e ordem $\mathbf{p} = (p_1, \dots, p_{\mathbb{S}})$, PVAR_m(\mathbf{p}), para a afluência \mathbf{a}_t é definido por

$$\mathbf{a}_t = \left(\zeta_t + \sum_{\nu=1}^{p_t} \Phi_{t,\nu} \mathbf{a}_{t-\nu} \right) \bullet \boldsymbol{\eta}_t, \quad (\text{A.50})$$

para todo inteiro positivo t , onde os erros *multiplicativos* $\{\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \dots\}$ são independentes. O símbolo \bullet representa o produto de Hadamard que é definido como um produto entrada a entrada entre vetores ou matrizes, ou seja, $(\mathbf{a} \bullet \mathbf{b})_i = a_i b_i$ e $(\mathbf{A} \bullet \mathbf{B})_{ij} = A_{ij} B_{ij}$. Para cada tempo t , os coeficientes e as ordens do modelo são iguais módulo \mathbb{S} ; os erros são identicamente distribuídos módulo \mathbb{S} com média unitária e variância constante módulo \mathbb{S} . Em outras palavras:

- $\zeta_t = \zeta_{t+\mathbb{S}}$ e $\Phi_{t,\nu} = \Phi_{t+\mathbb{S},\nu}$;
- $p_t = p_{t+\mathbb{S}}$;
- $\boldsymbol{\eta}_t$ possui a mesma distribuição de $\boldsymbol{\eta}_{t+\mathbb{S}}$; e
- $\mathbb{E}[\boldsymbol{\eta}_t] = \mathbf{1}$.

Vale ressaltar que o modelo PVAR_m apresentado é vetorial, pois $\{\mathbf{a}_1, \mathbf{a}_2, \dots\}$ é uma série vetorial. Por conseguinte, $\Phi_{t,1}, \dots, \Phi_{t,p_t}$ são matrizes e $\zeta_t, \boldsymbol{\eta}_t$ são vetores.

No modelo PVAR_m a afluência também é uma função afim das afluências passadas e o erro multiplicativo $\boldsymbol{\eta}_t$ é independente. Uma expressão equivalente a equação (A.50) que enfatiza a relação linear entre as afluências é dada abaixo:

$$\mathbf{a}_t = \text{diag}(\boldsymbol{\eta}_t) \left(\zeta_t + \sum_{\nu=1}^{p_t} \Phi_{t,\nu} \mathbf{a}_{t-\nu} \right), \quad (\text{A.51})$$

onde $\text{diag}(\boldsymbol{\eta}_t)$ é a matriz diagonal criada a partir do vetor $\boldsymbol{\eta}_t$. Portanto, a utilização do PVAR_m na reformulação do problema de planejamento de longo prazo da operação para a aplicação do algoritmo SDDP é *legítima*.

O objetivo desta seção é reformular o problema de planejamento de longo prazo da operação considerando os modelos PVAR e PVAR_m como restrição e o cálculo de subgradientes das respectivas funções de custo futuro e total. Recapitulando, a seção A.6.2 - equação (A.43) apresenta o problema de planejamento de longo prazo da operação em sua forma vetorial final e que por conveniência é escrito abaixo:

$$\begin{aligned} Q_t(v_t, \mathbf{a}_{[t]}) = \min \quad & c^\top g_t + \rho_t^\top u_t + \beta \bar{Q}_{t+1}(v_{t+1}, \mathbf{a}_{[t]}) \\ \text{s.a.} \quad & v_{t+1} = v_t + \mathbf{a}_t - q_t - s_t \\ & q_t + M_I g_t + M_D f_t = d_t \\ & 0 \leq v_{t+1} \leq \bar{v}, \quad 0 \leq q_t \leq \bar{q}, \quad 0 \leq s_t, \quad 0 \leq u_t, \\ & q_t + s_t + u_t \geq \underline{q}, \quad \underline{g} \leq g_t \leq \bar{g}, \quad \underline{f} \leq f_t \leq \bar{f} \end{aligned} \quad (\text{A.52})$$

$$\bar{Q}_{t+1}(v_{t+1}, \mathbf{a}_{[t]}) = \begin{cases} \mathbb{E}[Q_{t+1}(v_{t+1}, \mathbf{a}_{[t+1]}) \mid \mathbf{a}_{[t]}] & , t \in \{1, \dots, T-1\} \\ 0 & , t = T \end{cases},$$

onde o vetor de decisão é dado por $x_t = (v_{t+1}, q_t, s_t, g_t, f_t, u_t)$. Como observado, esse problema satisfaz a todos os requisitos necessários para o uso do algoritmo Nested Cutting Plane, desde que a afluência \mathbf{a}_t seja *sempre* positiva.

Para facilitar a exposição da reformulação do problema (A.52), é conveniente representar os modelos PVAR (A.49) e PVAR_m (A.50) em termos da ordem máxima:

$$\mathbf{a}_t = \zeta_t + \sum_{\nu=1}^{\hat{p}} \Phi_{t,\nu} \mathbf{a}_{t-\nu} + \boldsymbol{\epsilon}_t, \quad (\text{A.53})$$

$$\mathbf{a}_t = \text{diag}(\boldsymbol{\eta}_t) \left(\zeta_t + \sum_{\nu=1}^{\hat{p}} \Phi_{t,\nu} \mathbf{a}_{t-\nu} \right), \quad (\text{A.54})$$

onde \hat{p} é a maior ordem obtida de $\{p_1, \dots, p_S\}$. Se p_t é menor do que \hat{p} então os coeficientes $\Phi_{t,p_t+1}, \Phi_{t,p_t+2}, \dots, \Phi_{t,\hat{p}}$ são todos nulos. Esta representação simplifica a descrição dos argumentos da função de custo total e futuro no problema de planejamento da operação.

O problema de planejamento de longo prazo da operação (A.52) considerando o modelo PVAR para as afluências pode ser reformulado transformando o parâmetro aleatório \mathbf{a}_t em variável, ou seja, transformando o vetor de decisão $x_t = (v_{t+1}, q_t, s_t, g_t, f_t, u_t)$ em $x_t = (v_{t+1}, q_t, s_t, g_t, f_t, u_t, \mathbf{a}_t)$, substituindo o processo aleatório de \mathbf{a}_t para $\boldsymbol{\epsilon}_t$ e adicionando a equação do modelo PVAR (A.53) como uma restrição do problema de otimização estocástica:

$$\begin{aligned}
Q_t([v_t, \mathbf{a}_{[t-\hat{p}, t-1]}], \boldsymbol{\epsilon}_t) = & \min \quad c^\top g_t + \rho_t^\top u_t + \beta \bar{Q}_{t+1}([v_{t+1}, \mathbf{a}_{[t+1-\hat{p}, t]}) \\
\text{s.a.} \quad & v_{t+1} = v_t + \mathbf{a}_t - q_t - s_t \\
& \mathbf{a}_t = \zeta_t + \sum_{\nu=1}^{\hat{p}} \Phi_{t,\nu} \mathbf{a}_{t-\nu} + \boldsymbol{\epsilon}_t \\
& q_t + M_I g_t + M_D f_t = d_t \\
& 0 \leq v_{t+1} \leq \bar{v}, \quad 0 \leq q_t \leq \bar{q}, \quad 0 \leq s_t, \quad 0 \leq u_t, \\
& q_t + s_t + u_t \geq \underline{q}, \quad \underline{g} \leq g_t \leq \bar{g}, \quad \underline{f} \leq f_t \leq \bar{f}
\end{aligned} \tag{A.55}$$

$$\bar{Q}_{t+1}([v_{t+1}, \mathbf{a}_{[t+1-\hat{p}, t]}) = \begin{cases} \mathbb{E} [Q_{t+1}([v_{t+1}, \mathbf{a}_{[t+1-\hat{p}, t]})] & , t \in \{1, \dots, T-1\} \\ 0 & , t = T \end{cases} .$$

Vale lembrar que num problema de otimização estocástica cujo processo aleatório $\{\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2, \dots\}$ possui componentes independentes o condicionante “ $|\boldsymbol{\epsilon}_{[t]}$ ” da esperança condicional que define a função de custo futuro é dispensável. A justificativa é dada por uma indução, desde o último estágio até o primeiro, observando que cada função de custo total do estágio t necessita somente da informação de $\boldsymbol{\epsilon}_t$. Esta observação é trivial para o caso t igual a T . Com isso, é possível dispensar o condicionante “ $|\boldsymbol{\epsilon}_{[t]}$ ” da esperança condicional que define a função de custo futuro do estágio $t-1$. Consequentemente, a função de custo total do estágio $t-1$ depende somente da informação de $\boldsymbol{\epsilon}_{t-1}$ e assim sucessivamente.

De forma análoga, é possível utilizar a mesma técnica para a reformulação de (A.52) no caso em que a afluência \mathbf{a}_t é representada pelo modelo PVAR $_m$. Considerando o modelo PVAR $_m$ para as afluências, o problema (A.52) pode ser reformulado transformando o parâmetro aleatório \mathbf{a}_t em variável, ou seja, transformando o vetor de decisão $x_t = (v_{t+1}, q_t, s_t, g_t, f_t, u_t)$ em $x_t = (v_{t+1}, q_t, s_t, g_t, f_t, u_t, \mathbf{a}_t)$, substituindo o processo aleatório de \mathbf{a}_t para $\boldsymbol{\eta}_t$ e adicionando a equação do modelo PVAR $_m$ (A.54)

como uma restrição do problema de otimização estocástica:

$$\begin{aligned}
Q_t([v_t, \mathbf{a}_{[t-\hat{p}, t-1]}], \boldsymbol{\eta}_t) = & \min \quad c^\top g_t + \rho_t^\top u_t + \beta \bar{Q}_{t+1}([v_{t+1}, \mathbf{a}_{[t+1-\hat{p}, t]}]) \\
\text{s.a.} \quad & v_{t+1} = v_t + \mathbf{a}_t - q_t - s_t \\
& \mathbf{a}_t = \text{diag}(\boldsymbol{\eta}_t) \left(\zeta_t + \sum_{\nu=1}^{\hat{p}} \Phi_{t,\nu} \mathbf{a}_{t-\nu} \right) \\
& q_t + M_I g_t + M_D f_t = d_t \\
& 0 \leq v_{t+1} \leq \bar{v}, \quad 0 \leq q_t \leq \bar{q}, \quad 0 \leq s_t, \quad 0 \leq u_t, \\
& q_t + s_t + u_t \geq \underline{q}, \quad \underline{g} \leq g_t \leq \bar{g}, \quad \underline{f} \leq f_t \leq \bar{f}
\end{aligned} \tag{A.56}$$

$$\bar{Q}_{t+1}([v_{t+1}, \mathbf{a}_{[t+1-\hat{p}, t]}]) = \begin{cases} \mathbb{E} [Q_{t+1}([v_{t+1}, \mathbf{a}_{[t+1-\hat{p}, t]}], \boldsymbol{\eta}_{t+1})] & , t \in \{1, \dots, T-1\} \\ 0 & , t = T \end{cases} .$$

Os argumentos das funções de custo total e futuro de (A.55) e (A.56) podem ser justificados por indução. A justificativa será feita para o problema (A.55), porém todos os passos do desenvolvimento são válidos apenas trocando $\boldsymbol{\epsilon}_t$ por $\boldsymbol{\eta}_t$ e (A.55) por (A.56). Com efeito, seja t igual ao último estágio T . Neste caso, a função de custo futuro \bar{Q}_{T+1} é identicamente nula e o problema de otimização (A.55) está determinado uma vez fixados os parâmetros $v_T, \mathbf{a}_{T-\hat{p}}, \dots, \mathbf{a}_{T-1}$ e $\boldsymbol{\epsilon}_T$. Com isso, a função de custo total do estágio T é da forma $Q_T([v_T, \mathbf{a}_{[T-\hat{p}, T-1]}], \boldsymbol{\epsilon}_T)$ e como o custo futuro pode ser visto como a esperança do custo total então a função de custo futuro do estágio T é da forma $\bar{Q}_T([v_T, \mathbf{a}_{[T-\hat{p}, T-1]}])$. As hipóteses sobre os argumentos da função de custo total e futuro do estágio $t+1$ são $Q_{t+1}([v_{t+1}, \mathbf{a}_{[t+1-\hat{p}, t]}], \boldsymbol{\epsilon}_{t+1})$ e $\bar{Q}_{t+1}([v_{t+1}, \mathbf{a}_{[t+1-\hat{p}, t]}])$, respectivamente. Assim, o problema de otimização (A.55) está determinado uma vez fixados os parâmetros $v_t, \mathbf{a}_{t-\hat{p}}, \dots, \mathbf{a}_{t-1}$ e $\boldsymbol{\epsilon}_t$. Portanto, a função de custo total do estágio t é da forma $Q_t([v_t, \mathbf{a}_{[t-\hat{p}, t-1]}], \boldsymbol{\epsilon}_t)$ e como o custo futuro pode ser interpretado como a esperança do custo total então a função de custo futuro do estágio t é da forma $\bar{Q}_t([v_t, \mathbf{a}_{[t-\hat{p}, t-1]}])$, concluindo a prova por indução.

Como visto na seção A.6.3, o algoritmo SDDP é baseado nas mesmas técnicas do Nested Cutting Plane para a obtenção de soluções viáveis e cálculo de subgradientes. Portanto, o SDDP possui os mesmos requisitos do algoritmo Nested Cutting Plane com a exigência adicional do processo aleatório possuir componentes independente. Por comodidade, lista-se abaixo os requisitos do algoritmo SDDP:

- 1) O processo aleatório $\{\xi_1, \xi_2, \dots\}$ deve possuir componentes independentes;
- 2) O recurso deve ser relativamente completo, ou seja, toda decisão viável do problema associado a função de custo total de um estágio deve produzir um problema viável no estágio posterior;

- 3) As funções de custo futuro devem ser limitadas inferiormente, isto é, deve existir uma constante $M \in \mathbb{R}$ tal que $\bar{Q}_t(\cdot, \cdot) \geq M$.
- 4) O problema que define a função de custo total sem o termo da função de custo futuro deve ser limitado inferiormente:

$$-\infty < \min_{\substack{B_t x_{t-1} + A_t x_t = b_t \\ x_t \geq 0}} c_t x_t \quad ; \quad (\text{A.57})$$

- 5) Os subgradientes da aproximação da função de custo total devem ser calculáveis.

O requisito 1) é atendido para os problemas de otimização estocástica (A.55) e (A.56) considerando o modelo PVAR e PVAR_m, respectivamente, pois por hipótese os parâmetros aleatórios $\xi_t = \epsilon_t$ e $\xi_t = \eta_t$ formam um processo estocástico de termos independentes. A demonstração de que os problemas (A.55) e (A.56) atendem aos requisitos 2), 3) e 4) é análoga à mostrada na seção A.6.2 sobre a aplicação do Nested Cutting Plane. Vale ressaltar que a argumentação utilizada para justificar o item 2) se baseia no fato da afluência \mathbf{a}_t ser *sempre* positiva. Entretanto, como em geral o modelo PVAR pode gerar afluências negativas é possível que em algumas instâncias o problema (A.55) seja inviável. Já o problema (A.56) é relativamente completo, pois o modelo PVAR_m com restrição nos coeficientes gera sempre afluências positivas.

Em relação ao requisito 5), a apresentação do cálculo dos subgradientes para os problemas (A.55) e (A.56) é também análoga a apresentação feita na seção A.6.2 sobre a aplicação do Nested Cutting Plane. Por uma questão de completude da exposição, será deduzido em detalhes o cálculo dos subgradientes das aproximações das funções de custo total para cada um dos problemas (A.55) e (A.56).

Para facilitar a exposição do cálculo de tais subgradientes, representa-se a ordem do modelo PVAR e PVAR_m de cada mês como fixa e igual a ordem máxima, $\hat{p} = \max_s p_s$. Neste caso específico, também é conveniente que o valor de \hat{p} seja maior do que ou igual à 2. Assim, se a ordem de um mês é menor do que \hat{p} então os coeficientes associados aos lags inexistentes são considerados zero, mas denotados por uma notação cômoda para fins de cálculo.

Cálculo de subgradientes: caso PVAR(p)

Como enunciado em (A.55), o problema de planejamento de longo prazo da operação que considera a equação do modelo PVAR como uma restrição é dado por:

$$\begin{aligned}
Q_t([v_t, \mathbf{a}_{[t-\hat{p}, t-1]}], \boldsymbol{\epsilon}_t) = & \min \quad c^\top g_t + \rho_t^\top u_t + \beta \bar{Q}_{t+1}([v_{t+1}, \mathbf{a}_{[t+1-\hat{p}, t]}]) \\
s.a. \quad & v_{t+1} = v_t + \mathbf{a}_t - q_t - s_t \\
& \mathbf{a}_t = \zeta_t + \sum_{\nu=1}^{\hat{p}} \Phi_{t,\nu} \mathbf{a}_{t-\nu} + \boldsymbol{\epsilon}_t \\
& q_t + M_I g_t + M_D f_t = d_t \\
& 0 \leq v_{t+1} \leq \bar{v}, \quad 0 \leq q_t \leq \bar{q}, \quad 0 \leq s_t, \quad 0 \leq u_t, \\
& q_t + s_t + u_t \geq \underline{q}, \quad \underline{g} \leq g_t \leq \bar{g}, \quad \underline{f} \leq f_t \leq \bar{f}
\end{aligned} \tag{A.58}$$

$$\bar{Q}_{t+1}([v_{t+1}, \mathbf{a}_{[t+1-\hat{p}, t]}]) = \begin{cases} \mathbb{E} [Q_{t+1}([v_{t+1}, \mathbf{a}_{[t+1-\hat{p}, t]}], \boldsymbol{\epsilon}_{t+1})] & , t \in \{1, \dots, T-1\} \\ 0 & , t = T \end{cases} .$$

Numa iteração qualquer do SDDP, a aproximação da função de custo futuro é dada por

$$\bar{\mathcal{Q}}_{t+1}([v_{t+1}, \mathbf{a}_{[t+1-\hat{p}, t]}]) = \max_{l \in L_t} (r_{l,t} v_{t+1} + \mathcal{A}_{t,l} \mathbf{a}_{[t+1-\hat{p}, t-1]} + \mathcal{K}_{t,l} \mathbf{a}_t + \gamma_{t,l}) \tag{A.59}$$

A partir da aproximação da função de custo futuro (A.59), é possível construir um problema de minimização capaz representá-la de maneira equivalente:

$$\begin{aligned}
\bar{\mathcal{Q}}_{t+1}([v_{t+1}, \mathbf{a}_{[t+1-\hat{p}, t]}]) = & \min \quad \bar{\mathcal{Q}}_{t+1} \\
s.a. \quad & r_{l,t} v_{t+1} + \mathcal{A}_{t,l} \mathbf{a}_{[t+1-\hat{p}, t-1]} + \mathcal{K}_{t,l} \mathbf{a}_t + \gamma_{t,l} \leq \bar{\mathcal{Q}}_{t+1} \\
& \forall l \in L_t
\end{aligned}$$

Este mesmo problema pode ser escrito em uma forma vetorial:

$$\begin{aligned}
\bar{\mathcal{Q}}_{t+1}([v_{t+1}, \mathbf{a}_{[t+1-\hat{p}, t]}]) = & \min \quad \bar{\mathcal{Q}}_{t+1} \\
s.a. \quad & R_t v_{t+1} + \mathcal{A}_t \mathbf{a}_{[t+1-\hat{p}, t-1]} + \mathcal{K}_t \mathbf{a}_t + \gamma_t \leq \bar{\mathcal{Q}}_{t+1} \cdot \mathbf{1}
\end{aligned} \tag{A.60}$$

Com o intuito de descrever o problema que define a aproximação da função de custo total de modo mais conveniente para o cálculo de subgradientes, é interessante

representar a equação do modelo PVAR da seguinte forma:

$$\begin{aligned}
\mathbf{a}_t &= \zeta_t + \sum_{\nu=1}^{\hat{p}} \Phi_{t,\nu} \mathbf{a}_{t-\nu} + \boldsymbol{\epsilon}_t \\
&= \zeta_t + \begin{bmatrix} \Phi_{t,1} & \cdots & \Phi_{t,\hat{p}-1} \end{bmatrix} \begin{bmatrix} \mathbf{a}_{t-1} \\ \cdots \\ \mathbf{a}_{t+1-\hat{p}} \end{bmatrix} + \Phi_{t,\hat{p}} \mathbf{a}_{t-\hat{p}} + \boldsymbol{\epsilon}_t \\
&= \zeta_t + \Phi_{t,[1,\hat{p}-1]} \mathbf{a}_{[t+1-\hat{p},t-1]} + \Phi_{t,\hat{p}} \mathbf{a}_{t-\hat{p}} + \boldsymbol{\epsilon}_t
\end{aligned} \tag{A.61}$$

Reunindo as equações (A.60), (A.61) e apoiando-se na formulação (A.58) é possível descrever a aproximação da função de custo total para o problema em questão como:

$$\begin{aligned}
\mathfrak{Q}_t([v_t, \mathbf{a}_{[t-\hat{p},t-1]}], \boldsymbol{\epsilon}_t) &= \min \quad c^\top g_t + \rho_t^\top u_t + \beta \bar{\mathfrak{Q}}_{t+1} \\
s.a. \quad v_{t+1} &= v_t + \mathbf{a}_t - q_t - s_t \\
\mathbf{a}_t &= \zeta_t + \Phi_{t,[1,\hat{p}-1]} \mathbf{a}_{[t+1-\hat{p},t-1]} + \Phi_{t,\hat{p}} \mathbf{a}_{t-\hat{p}} + \boldsymbol{\epsilon}_t \\
q_t + M_I g_t + M_D f_t &= d_t \\
0 \leq v_{t+1} \leq \bar{v}, \quad 0 \leq q_t \leq \bar{q}, \quad 0 \leq s_t, \quad 0 \leq u_t, \\
q_t + s_t + u_t \geq \underline{q}, \quad \underline{g} \leq g_t \leq \bar{g}, \quad \underline{f} \leq f_t \leq \bar{f} \\
R_t v_{t+1} + \mathcal{D}_t \mathbf{a}_{[t+1-\hat{p},t-1]} + \mathcal{K}_t \mathbf{a}_t + \gamma_t &\leq \bar{\mathfrak{Q}}_{t+1} \cdot \mathbf{1}
\end{aligned} \tag{A.62}$$

A seguir, relembra-se a fórmula de referência para o cálculo de subgradientes de funções que são valor ótimo de um problemas otimização linear:

$$\begin{aligned}
V(x) &= \min_y \quad \tilde{c}^\top y \\
s.a. \quad A_1 y &= b_1 \quad (\pi_{b_1}) \\
A_2 y &\leq b_2 \quad (\pi_{b_2}), \\
T_1 x + W_1 y &= h_1 \quad (\pi_{h_1}) \\
T_2 x + W_2 y &\leq h_2 \quad (\pi_{h_2})
\end{aligned} \tag{A.63}$$

onde y é a variável de decisão de (A.63), $(\pi_{b_1}, \pi_{b_2}, \pi_{h_1}, \pi_{h_2})$ são as soluções duais ótimas associadas às restrições indicadas e x é a variável de estado, isto é, a variável de interesse que uma vez fixada define as restrições do problema (A.63). Abaixo relembra-se as propriedades da função $\mathfrak{Q}(x)$.

Fato 5: $V(\cdot)$ é uma função poliedral. Quando $V(x)$ é finito, o subdiferencial de

$V(\cdot)$ no ponto x é:

$$\begin{aligned} \partial V(x) &= - [0 \ 0 \ T_1^\top \ T_2^\top] \widehat{\mathcal{D}}(x), \\ &= \left\{ -T_1^\top \pi_{h_1} - T_2^\top \pi_{h_2} \left| \begin{array}{l} \pi_{h_1} \text{ e } \pi_{h_2} \text{ são as soluções duais ótimas} \\ \text{de } T_1 x + W_1 y = h_1 \text{ e } T_2 x + W_2 y \leq h_2 \end{array} \right. \right\}, \end{aligned} \quad (\text{A.64})$$

onde $\widehat{\mathcal{D}}(x)$ é o conjunto das soluções duais ótimas $(\pi_{b_1}, \pi_{b_2}, \pi_{h_1}, \pi_{h_2})$ associadas á (A.63).

Dessa forma, pode-se obter a expressão do subdiferencial de $\mathcal{Q}_t([v_t, \mathbf{a}_{[t-\hat{p}, t-1]}], \epsilon_t)$ associando as variáveis e restrições do problema (A.62) do mesmo modo que o problema (A.63). Neste sentido, observa-se que

- a variável de decisão y é igual a $(v_{t+1}, q_t, s_t, g_t, f_t, u_t, \mathbf{a}_t, \bar{\mathcal{Q}}_{t+1})$;
- a variável de estado x é igual a $(v_t, \mathbf{a}_{[t-\hat{p}, t-1]})$;
- o custo associado à decisão y é:

$$\tilde{c} = \begin{pmatrix} v_{t+1} & q_t & s_t & g_t & f_t & u_t & \mathbf{a}_t & \bar{\mathcal{Q}}_{t+1} \\ 0 & 0 & 0 & c & 0 & \rho_t & 0 & \beta \end{pmatrix};$$

- o bloco referente à restrição de igualdade $A_1 y = b_1$ é

$$A_1 = \begin{pmatrix} v_{t+1} & q_t & s_t & g_t & f_t & u_t & \mathbf{a}_t & \bar{\mathcal{Q}}_{t+1} \\ 0 & I & 0 & M_I & M_D & 0 & 0 & 0 \end{pmatrix} \quad b_1 = (d_t);$$

- o bloco referente à restrição de desigualdade $A_2 y \leq b_2$ é

$$A_2 = \begin{pmatrix} v_{t+1} & q_t & s_t & g_t & f_t & u_t & \mathbf{a}_t & \bar{\mathcal{Q}}_{t+1} \\ I & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I & 0 & 0 & 0 \\ -I & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -I & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -I & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -I & 0 & 0 \\ 0 & -I & -I & 0 & 0 & -I & 0 & 0 \\ 0 & 0 & 0 & -I & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -I & 0 & 0 & 0 \end{pmatrix}, \quad b_2 = \begin{pmatrix} \bar{v} \\ \bar{q} \\ \bar{g} \\ \bar{f} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ -\underline{q} \\ -\underline{g} \\ -\underline{f} \end{pmatrix},$$

- o bloco referente à restrição de igualdade $T_1x + W_1y = h_1$ é

$$T_1 = \begin{pmatrix} v_t & \mathbf{a}_{[t-\hat{p}+1,t-1]} & \mathbf{a}_{t-\hat{p}} \\ -I & 0 & 0 \\ 0 & -\Phi_{t,[1,\hat{p}-1]} & -\Phi_{t,\hat{p}} \end{pmatrix}$$

$$W_1 = \begin{pmatrix} v_{t+1} & q_t & s_t & g_t & f_t & u_t & \mathbf{a}_t & \bar{\Sigma}_{t+1} \\ I & I & I & 0 & 0 & 0 & -I & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & I & 0 \end{pmatrix}$$

$$h_1 = \begin{pmatrix} 0 \\ \zeta_t + \epsilon_t \end{pmatrix};$$

- o bloco referente à restrição de desigualdade $T_2x + W_2y \leq h_2$ é

$$T_2 = \begin{pmatrix} v_t & \mathbf{a}_{[t-\hat{p}+1,t-1]} & \mathbf{a}_{t-\hat{p}} \\ 0 & \mathbb{A}_t & 0 \end{pmatrix}$$

$$W_2 = \begin{pmatrix} v_{t+1} & q_t & s_t & g_t & f_t & u_t & \mathbf{a}_t & \bar{\Sigma}_{t+1} \\ R_t & 0 & 0 & 0 & 0 & 0 & \mathbb{K}_t & -\mathbf{1} \end{pmatrix}$$

$$h_2 = (-\gamma) ,$$

Como resultado do Fato 5), a expressão do subgradiente da aproximação da função de custo total (A.62) é:

$$\begin{aligned} -T_1^\top \begin{pmatrix} \pi_a \\ \pi_v \end{pmatrix} - T_2^\top \begin{pmatrix} \pi_c \end{pmatrix} &= \begin{pmatrix} I & 0 \\ 0 & \Phi_{t,[1,\hat{p}-1]}^\top \\ 0 & \Phi_{t,\hat{p}}^\top \end{pmatrix} \begin{pmatrix} \pi_v \\ \pi_a \end{pmatrix} + \begin{pmatrix} 0 \\ -\mathbb{A}_t^\top \\ 0 \end{pmatrix} \begin{pmatrix} \pi_c \end{pmatrix} \\ &= \begin{pmatrix} \pi_v \\ \Phi_{t,[1,\hat{p}-1]}^\top \pi_a - \mathbb{A}_t^\top \pi_c \\ \Phi_{t,\hat{p}}^\top \pi_a \end{pmatrix}. \end{aligned}$$

Portanto, a expressão do subdiferencial de $\mathfrak{Q}_t([v_t, \mathbf{a}_{[t-\hat{p}, t-1]}], \boldsymbol{\epsilon}_t)$ é dada por:

$$\partial \mathfrak{Q}_t([v_t, \mathbf{a}_{[t-\hat{p}, t-1]}], \boldsymbol{\epsilon}_t) = - \begin{bmatrix} 0 & 0 & T_1^\top & T_2^\top \end{bmatrix} \widehat{\mathfrak{D}}(x) =$$

$$= \left\{ \begin{array}{l} \left(\begin{array}{c} \pi_v \\ \Phi_{t, [1, \hat{p}-1]}^\top \pi_a - \mathbb{D}_t^\top \pi_c \\ \Phi_{t, \hat{p}}^\top \pi_a \end{array} \right) \left| \begin{array}{l} \pi_v, \pi_a \text{ e } \pi_c \text{ são as soluções duais} \\ \text{ótimas da restrição hidráulica,} \\ \text{da equação do modelo PVAR e da} \\ \text{inequação dos planos de cortes do} \\ \text{problema (A.58), respectivamente.} \end{array} \right. \end{array} \right\}.$$

Cálculo de subgradientes: caso PVAR_m(p)

Como enunciado em (A.56), o problema de planejamento de longo prazo da operação que considera a equação do modelo PVAR como uma restrição é dado por:

$$\begin{aligned} Q_t([v_t, \mathbf{a}_{[t-\hat{p}, t-1]}], \boldsymbol{\eta}_t) = & \min \quad c^\top g_t + \rho_t^\top u_t + \beta \bar{Q}_{t+1}([v_{t+1}, \mathbf{a}_{[t+1-\hat{p}, t]}]) \\ \text{s.a.} \quad & v_{t+1} = v_t + \mathbf{a}_t - q_t - s_t \\ & \mathbf{a}_t = \text{diag}(\boldsymbol{\eta}_t) \left(\zeta_t + \sum_{\nu=1}^{\hat{p}} \Phi_{t, \nu} \mathbf{a}_{t-\nu} \right) \\ & q_t + M_I g_t + M_D f_t = d_t \\ & 0 \leq v_{t+1} \leq \bar{v}, \quad 0 \leq q_t \leq \bar{q}, \quad 0 \leq s_t, \quad 0 \leq u_t, \\ & q_t + s_t + u_t \geq \underline{q}, \quad \underline{g} \leq g_t \leq \bar{g}, \quad \underline{f} \leq f_t \leq \bar{f} \end{aligned} \quad (\text{A.65})$$

$$\bar{Q}_{t+1}([v_{t+1}, \mathbf{a}_{[t+1-\hat{p}, t]}]) = \begin{cases} \mathbb{E} [Q_{t+1}([v_{t+1}, \mathbf{a}_{[t+1-\hat{p}, t]}], \boldsymbol{\eta}_{t+1})] & , t \in \{1, \dots, T-1\} \\ 0 & , t = T \end{cases}.$$

Numa iteração qualquer do SDDP, a aproximação da função de custo futuro é dada por

$$\bar{\mathfrak{Q}}_{t+1}([v_{t+1}, \mathbf{a}_{[t+1-\hat{p}, t]}]) = \max_{l \in L_t} (r_{t,l} v_{t+1} + \mathbb{A}_{t,l} \mathbf{a}_{[t+1-\hat{p}, t-1]} + \mathfrak{K}_{t,l} \mathbf{a}_t + \gamma_{t,l}) \quad (\text{A.66})$$

A partir da aproximação da função de custo futuro (A.66), é possível construir um problema de minimização capaz representá-la de maneira equivalente:

$$\begin{aligned} \bar{\mathfrak{Q}}_{t+1}([v_{t+1}, \mathbf{a}_{[t+1-\hat{p}, t]}]) = & \min \quad \bar{\mathfrak{Q}}_{t+1} \\ \text{s.a.} \quad & r_{t,l} v_{t+1} + \mathbb{A}_{t,l} \mathbf{a}_{[t+1-\hat{p}, t-1]} + \mathfrak{K}_{t,l} \mathbf{a}_t + \gamma_{t,l} \leq \bar{\mathfrak{Q}}_{t+1} \\ & \forall l \in L_t \end{aligned}$$

Este mesmo problema pode ser escrito em uma forma vetorial:

$$\begin{aligned} \bar{\mathfrak{Q}}_{t+1}([v_{t+1}, \mathbf{a}_{[t+1-\hat{p}, t]}]) &= \min \bar{\mathfrak{Q}}_{t+1} \\ \text{s.a.} \quad R_t v_{t+1} + \mathbb{A}_t \mathbf{a}_{[t+1-\hat{p}, t-1]} + \mathbb{K}_t \mathbf{a}_t + \gamma_t &\leq \bar{\mathfrak{Q}}_{t+1} \cdot \mathbf{1} \end{aligned} \quad (\text{A.67})$$

Com o intuito de descrever o problema que define a aproximação da função de custo total de modo mais conveniente para o cálculo de subgradientes, é interessante representar a equação do modelo PVAR_m da seguinte forma:

$$\begin{aligned} \mathbf{a}_t &= \text{diag}(\boldsymbol{\eta}_t) \left(\zeta_t + \sum_{\nu=1}^{\hat{p}} \Phi_{t,\nu} \mathbf{a}_{t-\nu} \right) \\ &= \text{diag}(\boldsymbol{\eta}_t) \left(\zeta_t + \begin{bmatrix} \Phi_{t,1} & \cdots & \Phi_{t,\hat{p}-1} \end{bmatrix} \begin{bmatrix} \mathbf{a}_{t-1} \\ \cdots \\ \mathbf{a}_{t+1-\hat{p}} \end{bmatrix} + \Phi_{t,\hat{p}} \mathbf{a}_{t-\hat{p}} + \boldsymbol{\epsilon}_t \right) \\ &= \text{diag}(\boldsymbol{\eta}_t) \left(\zeta_t + \Phi_{t,[1,\hat{p}-1]} \mathbf{a}_{[t+1-\hat{p}, t-1]} + \Phi_{t,\hat{p}} \mathbf{a}_{t-\hat{p}} \right) \end{aligned} \quad (\text{A.68})$$

Reunindo as equações (A.67), (A.68) e apoiando-se na formulação (A.65) é possível descrever a aproximação da função de custo total para o problema em questão como:

$$\begin{aligned} \mathfrak{Q}_t([v_t, \mathbf{a}_{[t-\hat{p}, t-1]}], \boldsymbol{\eta}_t) &= \min \quad c^\top g_t + \rho_t^\top u_t + \beta \bar{\mathfrak{Q}}_{t+1} \\ \text{s.a.} \quad v_{t+1} &= v_t + \mathbf{a}_t - q_t - s_t \\ \mathbf{a}_t &= \text{diag}(\boldsymbol{\eta}_t) \left(\zeta_t + \Phi_{t,[1,\hat{p}-1]} \mathbf{a}_{[t+1-\hat{p}, t-1]} + \Phi_{t,\hat{p}} \mathbf{a}_{t-\hat{p}} \right) \\ q_t + M_I g_t + M_D f_t &= d_t \\ 0 \leq v_{t+1} \leq \bar{v}, \quad 0 \leq q_t \leq \bar{q}, \quad 0 \leq s_t, \quad 0 \leq u_t, \\ q_t + s_t + u_t &\geq \underline{q}, \quad \underline{g} \leq g_t \leq \bar{g}, \quad \underline{f} \leq f_t \leq \bar{f} \\ R_t v_{t+1} + \mathbb{A}_t \mathbf{a}_{[t+1-\hat{p}, t-1]} + \mathbb{K}_t \mathbf{a}_t + \gamma_t &\leq \bar{\mathfrak{Q}}_{t+1} \cdot \mathbf{1} \end{aligned} \quad (\text{A.69})$$

A seguir, relembra-se a fórmula de referência para o cálculo de subgradientes de funções que são valor ótimo de um problemas otimização linear:

$$\begin{aligned} V(x) &= \min_y \quad \tilde{c}^\top y \\ \text{s.a.} \quad A_1 y &= b_1 \quad (\pi_{b_1}) \\ A_2 y &\leq b_2 \quad (\pi_{b_2}) , \\ T_1 x + W_1 y &= h_1 \quad (\pi_{h_1}) \\ T_2 x + W_2 y &\leq h_2 \quad (\pi_{h_2}) \end{aligned} \quad (\text{A.70})$$

onde y é a variável de decisão de (A.70), $(\pi_{b_1}, \pi_{b_2}, \pi_{h_1}, \pi_{h_2})$ são as soluções duais ótimas associadas às restrições indicadas e x é a variável de estado, isto é, a variável de interesse que uma vez fixada define as restrições do problema (A.70). Abaixo relembra-se as propriedades da função $V(x)$.

Fato 5: $V(\cdot)$ é uma função poliedral. Quando $V(x)$ é finito, o subdiferencial de $V(\cdot)$ no ponto x é:

$$\begin{aligned} \partial V(x) &= - [0 \ 0 \ T_1^\top \ T_2^\top] \widehat{\mathfrak{D}}(x), \\ &= \left\{ -T_1^\top \pi_{h_1} - T_2^\top \pi_{h_2} \left| \begin{array}{l} \pi_{h_1} \text{ e } \pi_{h_2} \text{ são as soluções duais ótimas} \\ \text{de } T_1 x + W_1 y = h_1 \text{ e } T_2 x + W_2 y \leq h_2 \end{array} \right. \right\}, \end{aligned} \quad (\text{A.71})$$

onde $\widehat{\mathfrak{D}}(x)$ é o conjunto das soluções duais ótimas $(\pi_{b_1}, \pi_{b_2}, \pi_{h_1}, \pi_{h_2})$ associadas á (A.70).

Dessa forma, pode-se obter a expressão do subdiferencial de $\mathfrak{Q}_t([v_t, \mathbf{a}_{[t-\hat{p}, t-1]}], \epsilon_t)$ associando as variáveis e restrições do problema (A.69) do mesmo modo que o problema (A.70). Neste sentido, observa-se que

- a variável de decisão y é igual a $(v_{t+1}, q_t, s_t, g_t, f_t, u_t, \mathbf{a}_t, \bar{\mathfrak{Q}}_{t+1})$;
- a variável de estado x é igual a $(v_t, \mathbf{a}_{[t-p, t-1]})$;
- o custo associado à decisão y é:

$$\tilde{c} = \begin{pmatrix} v_{t+1} & q_t & s_t & g_t & f_t & u_t & \mathbf{a}_t & \bar{\mathfrak{Q}}_{t+1} \\ 0 & 0 & 0 & c & 0 & \rho_t & 0 & \beta \end{pmatrix};$$

- o bloco referente à restrição de igualdade $A_1 y = b_1$ é

$$A_1 = \begin{pmatrix} v_{t+1} & q_t & s_t & g_t & f_t & u_t & \mathbf{a}_t & \bar{\mathfrak{Q}}_{t+1} \\ 0 & I & 0 & M_I & M_D & 0 & 0 & 0 \end{pmatrix} \quad b_1 = (d_t);$$

- o bloco referente à restrição de desigualdade $A_2 y \leq b_2$ é

$$A_2 = \begin{pmatrix} v_{t+1} & q_t & s_t & g_t & f_t & u_t & \mathbf{a}_t & \bar{\Sigma}_{t+1} \\ I & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I & 0 & 0 & 0 \\ -I & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -I & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -I & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -I & 0 & 0 \\ 0 & -I & -I & 0 & 0 & -I & 0 & 0 \\ 0 & 0 & 0 & -I & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -I & 0 & 0 & 0 \end{pmatrix}, \quad b_2 = \begin{pmatrix} \bar{v} \\ \bar{q} \\ \bar{g} \\ \bar{f} \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ -\underline{q} \\ -\underline{g} \\ -\underline{f} \end{pmatrix},$$

- o bloco referente à restrição de igualdade $T_1 x + W_1 y = h_1$ é

$$T_1 = \begin{pmatrix} v_t & \mathbf{a}_{[t-\hat{p}+1, t-1]} & \mathbf{a}_{t-\hat{p}} \\ -I & 0 & 0 \\ 0 & -\text{diag}(\boldsymbol{\eta}_t)\Phi_{t, [1, \hat{p}-1]} & -\text{diag}(\boldsymbol{\eta}_t)\Phi_{t, \hat{p}} \end{pmatrix}$$

$$W_1 = \begin{pmatrix} v_{t+1} & q_t & s_t & g_t & f_t & u_t & \mathbf{a}_t & \bar{\Sigma}_{t+1} \\ I & I & I & 0 & 0 & 0 & -I & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & I & 0 \end{pmatrix}$$

$$h_1 = \begin{pmatrix} 0 \\ \text{diag}(\boldsymbol{\eta}_t)\zeta_t \end{pmatrix};$$

- o bloco referente à restrição de desigualdade $T_2 x + W_2 y \leq h_2$ é

$$T_2 = \begin{pmatrix} v_t & \mathbf{a}_{[t-\hat{p}+1, t-1]} & \mathbf{a}_{t-\hat{p}} \\ 0 & \mathbb{A}_t & 0 \end{pmatrix}$$

$$W_2 = \begin{pmatrix} v_{t+1} & q_t & s_t & g_t & f_t & u_t & \mathbf{a}_t & \bar{\Sigma}_{t+1} \\ R_t & 0 & 0 & 0 & 0 & 0 & \mathbb{K}_t & -\mathbf{1} \end{pmatrix}$$

$$h_2 = (-\gamma),$$

Como resultado do Fato 5), a expressão do subgradiente da aproximação da função

de custo total (A.69) é:

$$\begin{aligned}
-T_1^\top \begin{pmatrix} \pi_a \\ \pi_v \end{pmatrix} - T_2^\top (\pi_c) &= \begin{pmatrix} I & 0 \\ 0 & \Phi_{t,[1,\hat{p}-1]}^\top \text{diag}(\boldsymbol{\eta}_t) \\ 0 & \Phi_{t,\hat{p}}^\top \text{diag}(\boldsymbol{\eta}_t) \end{pmatrix} \begin{pmatrix} \pi_v \\ \pi_a \end{pmatrix} + \begin{pmatrix} 0 \\ -\mathbb{D}_t^\top \\ 0 \end{pmatrix} (\pi_c) \\
&= \begin{pmatrix} \pi_v \\ \Phi_{t,[1,\hat{p}-1]}^\top \text{diag}(\boldsymbol{\eta}_t) \pi_a - \mathbb{D}_t^\top \pi_c \\ \Phi_{t,\hat{p}}^\top \text{diag}(\boldsymbol{\eta}_t) \pi_a \end{pmatrix} \\
&= \begin{pmatrix} \pi_v \\ \Phi_{t,[1,\hat{p}-1]}^\top (\boldsymbol{\eta}_t \bullet \pi_a) - \mathbb{D}_t^\top \pi_c \\ \Phi_{t,\hat{p}}^\top (\boldsymbol{\eta}_t \bullet \pi_a) \end{pmatrix}
\end{aligned}$$

Portanto, a expressão do subdiferencial de $\boldsymbol{\Omega}_t([v_t, \mathbf{a}_{[t-\hat{p},t-1]}], \boldsymbol{\eta}_t)$ é dada por:

$$\begin{aligned}
\partial \boldsymbol{\Omega}_t([v_t, \mathbf{a}_{[t-\hat{p},t-1]}], \boldsymbol{\eta}_t) &= - [0 \ 0 \ T_1^\top \ T_2^\top] \widehat{\mathcal{D}}(x) = \\
&= \left\{ \begin{array}{l} \left(\begin{array}{c} \pi_v \\ \Phi_{t,[1,\hat{p}-1]}^\top (\boldsymbol{\eta}_t \bullet \pi_a) - \mathbb{D}_t^\top \pi_c \\ \Phi_{t,\hat{p}}^\top (\boldsymbol{\eta}_t \bullet \pi_a) \end{array} \right) \left| \begin{array}{l} \pi_v, \pi_a \text{ e } \pi_c \text{ são as soluções duais} \\ \text{ótimas da restrição hidráulica,} \\ \text{da equação do modelo PVAR e da} \\ \text{inequação dos planos de cortes do} \\ \text{problema (A.65), respectivamente.} \end{array} \right. \end{array} \right\},
\end{aligned}$$