

---

# Uma Introdução à Otimização sob Incerteza

Humberto José Bortolossi

Departamento de Matemática Aplicada  
Universidade Federal  
Fluminense

Bernardo Kulnig Pagnoncelli

Departamento de Matemática  
Pontifícia Universidade Católica  
do Rio de Janeiro

---



---

III BIENAL DA SOCIEDADE BRASILEIRA DE MATEMÁTICA  
UNIVERSIDADE FEDERAL DE GOIÁS  
6 A 10 DE NOVEMBRO DE 2006

---

# Sumário

<b>Prefácio</b>	<b>iii</b>
<b>1 O Problema do Fazendeiro</b>	<b>1</b>
1.1 Representando cenários . . . . .	4
1.2 EVPI e VSS . . . . .	7
<b>2 O Problema do Jornaleiro</b>	<b>10</b>
2.1 Resolução do Problema . . . . .	11
2.2 Um exemplo numérico . . . . .	13
2.3 Outras interpretações para o problema . . . . .	15
<b>3 Programação Linear com Coeficientes Aleatórios</b>	<b>17</b>
3.1 O problema da mistura . . . . .	18
3.2 O problema da produção . . . . .	27
<b>4 Modelos de Recurso</b>	<b>32</b>
4.1 Motivação: programação linear por metas . . . . .	32
4.2 Modelos de recurso em otimização estocástica . . . . .	36
4.3 Admissibilidade . . . . .	38
4.4 Propriedades das funções de recurso . . . . .	40
4.5 Casos especiais: recurso completo e simples . . . . .	41

---

4.6	Mínimos e esperanças . . . . .	42
4.7	Cotas para o valor ótimo . . . . .	42
4.8	O caso $\Omega$ finito . . . . .	45
<b>5</b>	<b>O método L-shaped</b>	<b>47</b>
5.1	A decomposição de Benders . . . . .	47
5.2	O algoritmo de Benders . . . . .	50
5.3	Um exemplo completo . . . . .	51
5.4	Decomposição de Benders em otimização estocástica: o método L-shaped . . . . .	53
<b>6</b>	<b>Métodos Amostrais</b>	<b>56</b>
6.1	Aproximação pela média amostral . . . . .	56
6.2	A decomposição estocástica . . . . .	61
<b>A</b>	<b>Probabilidade</b>	<b>69</b>
<b>B</b>	<b>Estatística</b>	<b>77</b>
<b>C</b>	<b>Convexidade</b>	<b>81</b>
<b>D</b>	<b>Programação Linear</b>	<b>86</b>
	<b>Bibliografia</b>	<b>91</b>

# Prefácio

A maioria dos problemas da vida real trazem, em si, incertezas: elas são inerentes em virtualmente todos os sistemas relacionados com atuária, economia, meteorologia, demografia, ecologia, etc. Nos dias de hoje, problemas envolvendo interações entre homem, natureza e tecnologia estão sujeitos a mudanças rápidas, o que aumenta a incerteza. Cada nova revolução tecnológica traz novos desafios para o conhecimento estabelecido até então. Mesmo no contexto determinístico, existem sistemas que são tão complexos, que eles não permitem uma medida precisa de seus parâmetros.

A área de *otimização estocástica* (também conhecida como *otimização sob incerteza*) estuda modelos e métodos para abordar tais situações: elas incorporam incertezas na modelagem através da inclusão de variáveis aleatórias com distribuição de probabilidade conhecida. O objetivo é, então, encontrar soluções que sejam admissíveis para todas as possíveis realizações das variáveis aleatórias que são parte da modelagem.

A inclusão de variáveis aleatórias em um modelo de otimização cria muitas dificuldades: O que é uma solução admissível? O que é uma solução ótima? Como resolver estes problemas? Apresentaremos neste texto algumas das abordagens que procuram responder (dar um sentido) a estas perguntas. Nos concentraremos em uma classe muito importante de problemas de otimização estocástica: os chamados *modelos de recurso em dois estágios*. Em linhas gerais, estes modelos permitem que se faça uma escolha inicial (dita de primeiro estágio) antes de se conhecer o valor dos parâmetros incertos. Após o conhecimento dos valores dos mesmos, o agente de decisão faz novas escolhas (ditas de segundo estágio) que visam corrigir possíveis efeitos negativos gerados pela decisão de primeiro estágio (por este motivo, as decisões de segundo estágio também são chamadas de *ações corretivas*).

A solução obtida através da resolução de um problema de otimização estocástica é balanceada para todos os possíveis *cenários*, ou seja, é a melhor solução que leva em contas todos os possíveis valores que os parâmetros aleatórios podem assumir. Não fixamos simplesmente cada cenário e resolve-

mos vários problemas de otimização: estamos incorporando todos os cenários em um mesmo de problema e nos perguntando qual é a melhor decisão a se tomar levando em conta todas as situações que podem ocorrer.

É um fato geral que muitas aplicações de otimização estocástica dão origem a problemas de otimização determinística de grande porte, que são intratáveis mesmo para os computadores mais modernos. Uma área de pesquisa bastante ativa atualmente está voltada para o desenvolvimento de algoritmos que aproximam as soluções de problemas de grande porte. Neste texto apresentaremos dois deles, a *aproximação pela média amostral* e a *decomposição estocástica*.

Do ponto de vista pedagógico, a área de otimização estocástica é muito rica, por usar conceitos e resultados de programação linear, probabilidade e estatística.

## Agradecimentos

Este texto é fruto de um ciclo de seminários realizados na Pontifícia Universidade Católica do Rio de Janeiro desde o segundo semestre de 2005, como parte do programa de pós-graduação em atuária. Além de vários artigos da área, os livros [03, 09, 10, 13, 14, 17] foram muito inspiradores!

Gostaríamos de agradecer a todos que participaram dos seminários: Derek Hacon, Jessica Kubrusly, Marina Sequeiros Dias, Débora Freire Mondaini, Eduardo Teles da Silva, Niko A. Iliadis, Raphael M. Chabar e, em especial, ao professor Carlos Tomei, organizador dos seminários e co-autor de direito deste texto!

Humberto José Bortolossi

([hjbortol@vm.uff.br](mailto:hjbortol@vm.uff.br))

Departamento de Matemática Aplicada

Universidade Federal

Fluminense

Bernardo Kulnig Pagnoncelli

([bernardo@mat.puc-rio.br](mailto:bernardo@mat.puc-rio.br))

Departamento de Matemática

Pontifícia Universidade Católica

do Rio de Janeiro

# Capítulo 1

## O Problema do Fazendeiro

Vamos começar nosso estudo de otimização estocástica pelo *problema do fazendeiro* [03]. João é um fazendeiro que possui de 500 hectares (ha) de terra disponíveis para cultivo. Aliás, lembre-se que 500 ha equivalem a 5 000 000 m<sup>2</sup>. Ele é especialista em três cultivos: trigo, milho e cana-de-açúcar. Durante o inverno, ele tem que decidir quanto de terra será dedicada a cada uma das três culturas. A Figura 1.1 mostra duas possibilidades de divisão.

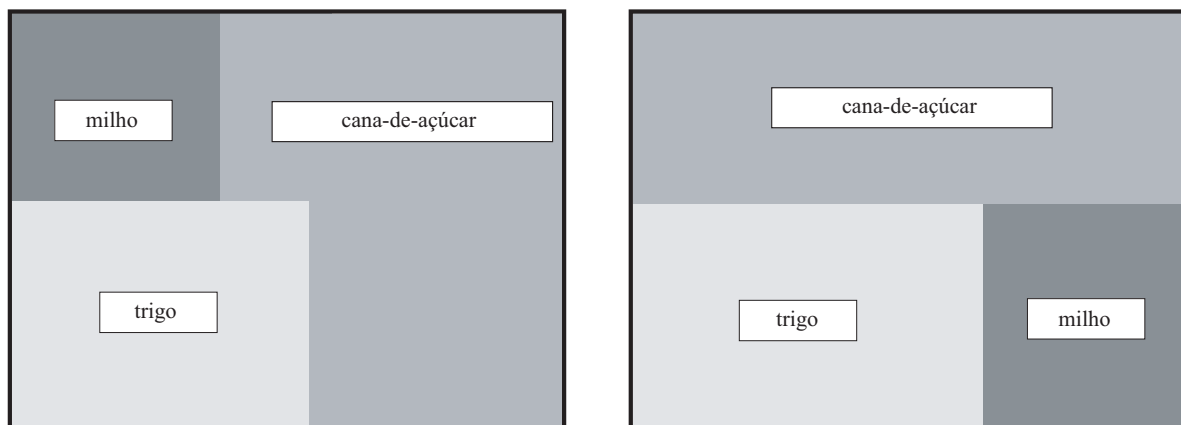


Figura 1.1: Duas divisões possíveis da terra.

Além do tamanho de sua propriedade, João possui outras restrições a serem consideradas. Ele também é proprietário de gado, que precisa ser alimentado. Seu gado precisa de pelo menos 200 toneladas (T) de trigo e 240 T de milho para a ração. Além do trigo e milho produzidos em suas terras, ele pode comprar esses produtos de outros produtores, no mercado local. Seu excesso de produção pode ser vendido para atacadistas, porém o preço é bem menor devido a margem de lucro destes comerciantes.

A cana-de-açúcar é um cultivo exclusivamente para dar lucro: toda sua produção é vendida para atacadistas a 36 reais por tonelada (R\$/T). No entanto, o governo impõe uma cota de produção de 6 000 T: qualquer quantidade produzida acima desse valor deve ser vendida por apenas 10R\$/T.

Baseado em informações de anos anteriores, João sabe que o rendimento médio de suas lavouras é 2.5, 3.0 e 20 toneladas por hectare (T/ha). Além disso, existe um custo de produção específico de cada lavoura, que é dado em R\$/ha. Os dados completos do modelo estão representados na Tabela 1.1 a seguir:

	Trigo	Milho	Cana-de-açúcar
Rendimento (T/ha)	2.5	3.0	20
Custo de produção (\$/ha)	150	230	260
Preço de venda (\$/T)	170	150	36( $\leq 6\,000$ T) 10( $> 6\,000$ T)
Preço de compra (\$/T)	238	210	–
Requerimento mínimo para o gado (T)	200	240	–
Total de terra disponível: 500 ha			

Tabela 1.1: Dados para o problema do fazendeiro.

Para ajudar João a decidir sobre como dividir suas terras de forma a maximizar seus lucros, vamos formular um problema de otimização linear que descreve essa situação. Defina

- $x_1$  = hectares dedicados ao trigo,
- $x_2$  = hectares dedicados ao milho,
- $x_3$  = hectares dedicados a cana-de-açúcar,
- $w_1$  = toneladas de trigo vendidas,
- $y_1$  = toneladas de trigo compradas,
- $w_2$  = toneladas de milho vendidas,
- $y_2$  = toneladas de milho compradas,
- $w_3$  = toneladas de cana-de-açúcar e
- $w_4$  = toneladas de cana-de-açúcar.

Queremos modelar essa situação como um problema de minimização ao invés de um de maximização, por razões que ficaram claras um pouco mais à frente no texto. Logicamente, o valor da função objetivo deve ser interpre-

tado com o sinal oposto. Dessa forma o problema fica

$$\begin{array}{ll}
 \text{minimizar} & 150 x_1 + 230 x_2 + 260 x_3 + \\
 & 238 y_1 - 170 w_1 + 210 y_2 - 150 w_2 - 36 w_3 - 10 w_4 \\
 \text{sujeito a} & x_1 + x_2 + x_3 \leq 500, \\
 & 2.5 x_1 + y_1 - w_1 \geq 200, \\
 & 3 x_2 + y_2 - w_2 \geq 240, \\
 & w_3 + w_4 \leq 20 x_3, \\
 & w_3 \leq 6\,000, \\
 & x_1, x_2, x_3, y_1, y_2, w_1, w_2, w_3, w_4 \geq 0.
 \end{array} \tag{1.1}$$

Esse é um problema de otimização linear e existem diversos programas disponíveis na internet que resolvem esse problema rapidamente. No entanto, é necessário escrevê-lo em uma linguagem especial, chamada AMPL ([08]). Essa linguagem é própria para problemas de otimização e é muito simples de aprender, pois seu código é muito semelhante a maneira como escrevemos um problema de otimização. Mais detalhes em <http://www.ampl.com/>. O mais popular resolvidor de problemas de otimização é o CPLEX, que apesar de não ser gratuito possui uma versão para estudantes gratuita, disponível em <http://www.ampl.com/DOWNLOADS/cplex80.html>. A solução do problema está descrita na Tabela 1.2.

	Trigo	Milho	Cana-de-açúcar
Área (ha)	120	80	300
Total produzido	300	240	6 000
Total vendido	100	–	6 000
Total comprado	–	–	–
Lucro total: R \$118 600			

Tabela 1.2: Solução do problema.

Pronto, o problema de João está resolvido: basta dividir as terras de acordo com a Tabela 1.2 para que ele maximize seus lucros. No entanto, João fica desconfiado com a solução. E se sua experiência em relação ao rendimento médio das culturas não for tão precisa quanto ele pensa? E se o ano em questão tiver um clima particularmente desfavorável e sua lavoura render menos do que o esperado? Será que a mesma divisão de terras é a melhor possível? Vamos estudar essas questões na próxima seção.



## 1.1 Representando cenários

Vamos supor que num ano particularmente favorável os rendimentos sejam 20% maiores que os rendimentos médios sugeridos por João. Alterando esse dados e resolvendo o problema para esses rendimentos obtemos a solução descrita na Tabela 1.3. Por outro lado podemos ter um ano desfavorável no qual os rendimentos fiquem 20% abaixo da média. Nesse caso a solução é dada pela Tabela 1.4.

	Trigo	Milho	Cana-de-açúcar
Área (ha)	183.33	66.67	250
Total produzido	550	240	6 000
Total vendido	350	-	6 000
Total comprado	-	-	-
Lucro total: R\$ 167 600			

Tabela 1.3: Solução ótima com rendimentos 20% acima da média.

	Trigo	Milho	Cana-de-açúcar
Área (ha)	100	25	375
Total produzido	200	60	6 000
Total vendido	-	-	6 000
Total comprado	-	180	-
Lucro total: R\$ 59 950			

Tabela 1.4: Solução ótima com rendimentos 20% abaixo da média.

Esses resultados são alarmantes para as finanças de João: mudanças de 20% nos rendimentos das culturas em relação ao rendimento médio fazem o seu lucro variar de R\$ 59 950 a R\$ 167 667! Pensando na cana-da-açúcar, João tem o seguinte dilema: se reservar uma área muito grande para esse cultivo e os rendimentos foram acima da média, então ele terá que vender uma quantidade da produção a um preço desfavorável por causa da cota. Por outro lado, se ele reservar um área muito pequena e os rendimentos forem abaixo da média, então ele vai perder a oportunidade de vender cana-de-açúcar a um preço favorável.

João conclui que não existe uma solução que seja ótima para todos os casos. No entanto, ele se questiona se existe uma solução que seja satisfatória para todos os tipos de rendimentos possíveis. A resposta para essa pergunta

virá com a primeira formulação de otimização estocástica, que estudaremos a seguir.

Vamos introduzir um pouco de nomenclatura: os cenários 20% acima da média, na média e 20% abaixo da média serão indexados por  $s = 1, 2, 3$  respectivamente. As variáveis  $y$  e  $w$  terão o mesmo significado da formulação (1.1), mas serão indexadas por  $w_{is}$ ,  $i = 1, 2, 3, 4$ ,  $s = 1, 2, 3$  e  $y_{js}$ ,  $j = 1, 2$ ,  $s = 1, 2, 3$ . Por exemplo,  $y_{23}$  representa a quantidade de milho vendida no caso de preços abaixo da média. Vamos assumir que os cenários são equiprováveis, ou seja, que cada um ocorre com probabilidade  $1/3$ . Além disso, supondo que João quer maximizar seus ganhos a longo prazo, é razoável supor que ele procura uma solução que maximize seu *lucro esperado*. Nesse caso o problema fica

$$\begin{array}{c}
 \text{minimizar} \\
 150 x_1 + 230 x_2 + 260 x_3 \\
 -\frac{1}{3}(170 w_{11} - 238 y_{11} + 150 w_{21} - 210 y_{21} + 36 w_{31} + 10 w_{41}) \\
 -\frac{1}{3}(170 w_{12} - 238 y_{12} + 150 w_{22} - 210 y_{22} + 36 w_{32} + 10 w_{42}) \\
 -\frac{1}{3}(170 w_{13} - 238 y_{13} + 150 w_{23} - 210 y_{23} + 36 w_{33} + 10 w_{43}) \\
 \\
 \text{sujeito a} \\
 x_1 + x_2 + x_3 \leq 500 \\
 3 x_1 + y_{11} - w_{11} \geq 200, \quad 3.6 x_2 + y_{21} - w_{21} \geq 240, \\
 w_{31} + w_{41} \leq 24 x_3, \quad w_{31} \leq 6\,000, \\
 2.5 x_1 + y_{12} - w_{12} \geq 200, \quad 3 x_2 + y_{22} - w_{22} \geq 240, \\
 w_{32} + w_{42} \leq 20 x_3, \quad w_{32} \leq 6\,000, \\
 2 x_1 + y_{13} - w_{13} \geq 200, \quad 2.4 x_2 + y_{23} - w_{23} \geq 240, \\
 w_{33} + w_{43} \leq 16 x_3, \quad w_{33} \leq 6\,000 \\
 x_1, x_2, x_3 \geq 0, \\
 y_{11}, y_{21}, y_{12}, y_{22}, y_{13}, y_{23} \geq 0, \\
 w_{11}, w_{21}, w_{31}, w_{41}, w_{12}, w_{22}, w_{32}, w_{42}, w_{13}, w_{23}, w_{33}, w_{43} \geq 0
 \end{array} \tag{1.2}$$

Esta é a chamada *forma extensa* de um problema de otimização es-

tocástica. Essa denominação vem do fato que todas as variáveis que dependem de cenários estão explicitamente descritas no modelo. As variáveis  $x$  são chamadas variáveis de *primeiro estágio*, pois seu valor tem que ser definido antes de se conhecer o clima e, conseqüentemente, o rendimento das culturas. As variáveis  $y_{is}$  e  $w_{is}$  são variáveis de *segundo estágio*. São variáveis que são escolhidas após o conhecimento do rendimento das lavouras. Elas servem para *corrigir* uma possível situação de déficit nas necessidades alimentares do gado resultante da escolha  $x$  de primeiro estágio. O problema do fazendeiro é um exemplo de *problema de recurso com dois estágios*, que será estudado em detalhe mais adiante no texto.

Note que o problema 1.2 é linear e pode ser resolvido da mesma forma que os anteriores. Exibimos na Tabela 1.5 a solução ótima, bem como as quantidades produzidas em cada cenário e os valores de compra e venda das culturas.

		Trigo	Milho	Cana-de-açúcar
Primeiro estágio	Área (ha)	170	80	250
$s = 1$ (Acima)	Rendimento (T)	510	288	6 000
	Venda (T)	310	48	6 000
	Compra(T)	–	–	(preço favorável) – –
$s = 2$ (Média)	Rendimento (T)	425	240	5 000
	Venda (T)	225	–	5 000
	Compra(T)	–	–	(preço favorável) –
$s = 3$ (Abaixo)	Rendimento (T)	340	192	4 000
	Venda (T)	140	–	4 000
	Compra(T)	–	48	(preço favorável) –
Lucro total: R\$ 108 390				

Tabela 1.5: Solução ótima do modelo estocástico.

A primeira linha da Tabela 1.5 nos dá a solução de primeiro estágio enquanto que as outras descrevem a solução de segundo estágio para cada cenário. O aspecto mais interessante da solução estocástica é que ela deixa claro ser impossível escolher uma solução que seja ótima para todos os cenários. No caso  $s = 3$  por exemplo, onde os rendimentos são 20% abaixo da média, temos a compra de 48 toneladas de milho para suprir as necessidades do gado. É claro que se soubéssemos que os rendimentos seriam abaixo

da média teríamos reservado mais área para o plantio de milho para evitar que este produto fosse comprado de outros comerciantes.

Dessa forma, a solução de primeiro estágio  $(x_1, x_2, x_3) = (170, 80, 250)$  do problema (1.2) representa o melhor que se pode fazer diante dos diferentes cenários que podem ocorrer. Na próxima seção vamos tentar mensurar o ganho de João por considerar o problema estocástico bem como a quantidade de dinheiro perdida por não conhecer com exatidão o futuro.

## 1.2 EVPI e VSS

Imagine que João tenha uma bola de cristal e consiga prever o clima no futuro. Sob essa hipótese, ele não precisa do modelo estocástico (1.2): sempre que ele antevê um rendimento 20% abaixo da média (respectivamente 20% acima da média) ele escolhe a solução dada na Tabela 1.4 (resp. Tabela 1.3). Se os rendimentos forem na média, ele se baseia na Tabela 1.2.

Se esperarmos um número grande de anos, então o rendimento médio de João sob informação perfeita (WS) será

$$WS = \frac{\text{R\$ } 59\,950 + \text{R\$ } 167\,667 + \text{R\$ } 118\,600}{3} = \text{R\$ } 115\,406. \quad (1.3)$$

Note que estamos assumindo que os diferentes cenários ocorrem ao acaso com probabilidade 1/3 cada. Essa rendimento médio corresponde à situação sob informação perfeita, ou seja, à situação onde João sabe com precisão que cenário ocorrerá no futuro.

Infelizmente, nós e os meteorologistas sabemos que tal hipótese não é realista. Assim, ao longo de um período de, digamos, 20 anos, o melhor que João tem a fazer é utilizar a solução estocástica dada pela Tabela 1.5, obtendo um lucro esperado de R\$ 108 390. A diferença entre este valor e o lucro no caso sob informação perfeita (equação (1.3)) é o *valor esperado de informação perfeita*, ou EVPI:

$$EVPI = \text{R\$ } 115\,406 - \text{R\$ } 108\,390 = \text{R\$ } 7\,016. \quad (1.4)$$

Um outro conceito importante em otimização estocástica é o *valor da solução estocástica* (VSS). O VSS mede o ganho em considerar o modelo estocástico ao invés de simplesmente basear a decisão nos rendimentos médios.

Pense que João é um fazendeiro teimoso: mesmo sabendo que possíveis variações de rendimento podem ocorrer, ele insiste em dividir sua terra de acordo com a situação de rendimentos médios dado pela Tabela 1.2. O lucro obtido com essa política é chamado *Solução do Valor Esperado*, ou EEV.

Como calculá-lo? É simples: fixe a distribuição de terras do caso de rendimentos médios, ou seja, calcule a solução do problema (1.1) nas variáveis  $y_{is}$  e  $w_{is}$ , tomando  $x_1 = 120$ ,  $x_2 = 80$  e  $x_3 = 300$  e os rendimentos iguais a 3.0, 3.6 e 24 (para  $s = 1$ ) e depois 2, 2.4 e 16 (para  $s = 3$ ). As soluções são R \$55 120 e R\$ 148 000 respectivamente. Lembrando que a solução é R\$ 118 600 no caso de rendimentos médios e R\$ 108 390 no caso estocástico, temos

$$\text{EEV} = \frac{\text{R\$ } 55\,120 + \text{R\$ } 118\,600 + \text{R\$ } 148\,000}{3} = \text{R\$ } 107\,240,$$

$$\text{VSS} = \text{R\$ } 108\,390 - \text{R\$ } 107\,240 = \text{R\$ } 1\,150.$$

Os conceitos de EVPI e VSS são importantes pois eles quantificam o valor da informação e o ganho em se considerar a formulação estocástica. No caso do EVPI, ele diz o quanto vale a pena pagar para se obter informação perfeita. Já o VSS nos dá acesso ao quanto estamos ganhando em considerar o modelo estocástico ao invés de simplesmente supor que os rendimentos das culturas são dados pelos rendimentos médios.

## Exercícios

- [01] No problema do fazendeiro, suponha que quando os rendimentos são altos para um fazendeiro o mesmo ocorre para os fazendeiros vizinhos. Assim, o aumento na demanda reduz os preços. Considere por exemplo que os preços do milho e do trigo caem 10% quando os rendimentos são acima da média e sobem 10% quando são abaixo. Formule e resolva o problema nesse caso, supondo que as alterações de preço são verificadas para compra e para venda de milho e trigo e que a cana-de-açúcar não sofre mudanças.
- [02] Suponha agora que a propriedade do fazendeiro é dividida em quatro lotes, de tamanhos 185, 145, 105 e 65 hectares respectivamente. Por

razões de eficiência, o fazendeiro só pode cultivar um tipo de produto por lote. Formule e resolva o problema do fazendeiro nesse caso.

- [03] Imagine que as compras e vendas de trigo e milho só podem ser feitas em centenas de toneladas, ou seja, não é possível comprar nem vender esses produtos em quantidades diferentes de múltiplos de 100. Formule e resolva o problema do fazendeiro sob essas restrições.

## Capítulo 2

# O Problema do Jornaleiro

O segundo exemplo que vamos considerar é conhecido como problema do jornaleiro ou problema da árvore de natal. Este problema é um clássico na área de otimização, possuindo vasta literatura a respeito. Uma interessante aplicação do problema do jornaleiro é descrita em [01]. Nesse artigo, idéias do problema do jornaleiro são aplicadas à distribuição de revistas da empresa *Time inc.* e o processo desenvolvido pelos autores gerou uma economia de 3.5 milhões de dólares por ano. Vamos descrever o problema seguindo a formulação proposta por [03].

O fazendeiro João tem um irmão na cidade chamado José, que é jornaleiro. Toda manhã ele vai ao editor do jornal e compra uma quantidade  $x$  de jornais a um preço  $c$  por unidade. Essa quantidade  $x$  é limitada superiormente por um valor  $u$ , pois José tem um poder de compra finito. Ele vende seus jornais a um preço  $q$  por unidade. José possui um acordo com o editor do jornal: qualquer jornal não vendido pode ser devolvido ao editor, que paga um preço  $r < c$  por ele.

O dilema de José diz respeito a *demanda diária* por jornal, que é incerta. Se ele comprar um número muito grande de jornais corre o risco de não vendê-los e perder dinheiro com isso. Por outro lado, se comprar poucos José pode não atender a demanda e deixar de faturar dinheiro. Vamos supor que a demanda  $\xi$  é uma variável aleatória não-negativa com função densidade  $f$  e função distribuição  $F$ , que  $y$  é o número de jornais efetivamente vendidos e que  $w$  é o número de possíveis jornais devolvidos ao editor. A formulação

do problema do jornaleiro é

$$\min_{0 \leq x \leq u} \{cx + \mathcal{Q}(x)\}$$

onde

$$\mathcal{Q}(x) = \mathbb{E}_\xi[\mathcal{Q}(x, \xi)]$$

e

$$\begin{aligned} \mathcal{Q}(x, \xi) = \min & \quad -qy(\xi) - rw(\xi) \\ \text{sujeito a} & \quad y(\xi) \leq \xi, \\ & \quad y(\xi) + w(\xi) \leq x, \\ & \quad y(\xi), w(\xi) \geq 0. \end{aligned}$$

O símbolo  $\mathbb{E}_\xi$  representa a esperança com respeito a  $\xi$ . Para uma quantidade  $x$  de jornais comprados, a função  $-\mathcal{Q}(x, \xi)$  denota o lucro obtido com a venda destes jornais para um valor fixo de demanda  $\xi$ . O valor  $-\mathcal{Q}(x)$  é o lucro esperado calculado sobre todos os valores  $\xi$  possíveis.

Assim como no problema do fazendeiro, o problema do jornaleiro é estruturado em dois estágios: no primeiro estágio José decide quantos jornais vai comprar através da variável  $x$ . Após essa escolha, ele vai tentar vender esses jornais para uma demanda  $\xi$ . As variáveis de segundo estágio representam quanto ele conseguiu vender ( $y(\xi)$ ) e quanto ele devolveu ao editor ( $w(\xi)$ ). Observe que a dependência dessas variáveis em  $\xi$  deixa claro que elas são de segundo estágio, pois seu valor só é determinado após o conhecimento da demanda  $\xi$ .

José procura a quantidade certa de jornais a comprar de forma a maximizar seu *lucro esperado* sob incerteza de demanda. Note aqui a semelhança com o problema do fazendeiro: se a demanda fosse conhecida José simplesmente comprava  $\xi$  jornais e obteria o lucro máximo. No entanto, como no problema do fazendeiro, não é possível escolher um valor  $x$  que maximize seu lucro para todos os possíveis valores de demanda  $\xi$ . O que José busca então é uma escolha que, *em média*, lhe dê o maior lucro.

## 2.1 Resolução do Problema

O primeiro passo para encontrar uma solução explícita do problema do jornaleiro é resolver o problema de segundo estágio. Felizmente a solução



é imediata: se a demanda  $\xi$  for menor que o número de jornais comprados então  $y^*(\xi) = \xi$ . Se for maior então  $y^*(\xi) = x$ . Para encontrar  $w^*(\xi)$  basta observar que retornos de jornais ao editor só ocorrem se a demanda  $\xi$  for menor que  $x$ . Conclui-se então que

$$\begin{aligned} y^*(\xi) &= \min\{\xi, x\}, \\ w^*(\xi) &= \max\{x - \xi, 0\}. \end{aligned}$$

A resolução desse problema nos permite escrever  $\mathcal{Q}(x)$  explicitamente:

$$\mathcal{Q}(x) = \mathbb{E}_\xi [-q \min\{\xi, x\} - r \max\{x - \xi, 0\}].$$

Vamos ver posteriormente que a função  $\mathcal{Q}$  é convexa e derivável quando a variável aleatória  $\xi$  for contínua. Como estamos no intervalo  $[0, u]$  e a função  $\mathcal{Q}(x)$  é convexa, sabemos que se  $c + \mathcal{Q}'(0) > 0$ , então a derivada não troca de sinal no intervalo e a solução ótima é  $x = 0$ . De maneira análoga, se  $c + \mathcal{Q}'(u) < 0$ , então a solução ótima é  $x = u$ . Caso nenhuma dessas condições se verifique temos que encontrar o ponto crítico de  $c + \mathcal{Q}(x)$ .

Usando a definição A.15 dada no apêndice A, temos que

$$\mathcal{Q}(x) = \int_{-\infty}^x (-qt - r(x - t))f(t) dt + \int_x^{\infty} -qx f(t) dt.$$

Manipulando a expressão e usando a equação (A.2) do apêndice A obtemos que

$$\mathcal{Q}(x) = -(q - r) \int_{-\infty}^x t f(t) dt - rxF(x) - qx(1 - F(x)).$$

Usando integração por partes, podemos simplificar ainda mais a expressão:

$$\mathcal{Q}(x) = -qx + (q - r) \int_{-\infty}^x F(t) dt. \quad (2.1)$$

A partir desta expressão podemos concluir que

$$\mathcal{Q}'(x) = -q + (q - r)F(x). \quad (2.2)$$

Finalmente, a solução do problema é

$$\begin{cases} x^* = 0, & \text{se } \frac{q-c}{q-r} < F(0), \\ x^* = u, & \text{se } \frac{q-c}{q-r} > F(u), \\ x^* = F^{-1}\left(\frac{q-c}{q-r}\right), & \text{caso contrário.} \end{cases} \quad (2.3)$$

Qualquer modelagem razoável da demanda  $\xi$  admite que ela só assume valores positivos. Nesse caso  $F(0) = 0$  e, portanto, nunca temos  $x^* = 0$ .

O exemplo do jornaleiro é mais um exemplo de problema de recurso com dois estágios. Novamente o agente decisório, nesse caso José, tem que fazer uma escolha sob incerteza. Ele não conhece a demanda no momento que compra os jornais junto ao editor. Após a compra ele ajusta as variáveis de segundo estágio de acordo com o valor da demanda, agora conhecido. A solução do problema representa a política de compras que rende o maior *lucro esperado* para José.

## 2.2 Um exemplo numérico

Vamos apresentar um exemplo numérico do problema do jornaleiro. Suponha que o custo por jornal para o jornaleiro seja  $c = 10$ , que o preço de venda seja  $q = 25$ , que o preço de devolução ao editor seja de  $r = 5$  por jornal e que o poder de compra é  $u = 150$ . Além disso, considere que a demanda  $\xi$  é dada por uma variável aleatória uniforme contínua definida no intervalo  $[50, 150]$ . Na Tabela A.1 do apêndice A listamos a densidade, média e variância dessa variável aleatória.

Integrando-se a densidade de  $\xi$ , obtemos a função distribuição da demanda:

$$F(x) = \begin{cases} \frac{x-50}{100}, & \text{se } 50 \leq x \leq 150, \\ 1, & \text{se } x > 150, \\ 0, & \text{caso contrário.} \end{cases} \quad (2.4)$$

A inversa dessa função é  $F^{-1}(y) = 100y + 50$  no intervalo  $[50, 150]$ . Usando (2.3), temos que a solução do problema é

$$x^* = F^{-1}(3/4) = 125,$$

com lucro esperado de 1312.5. Assim, José deve comprar 125 jornais por dia para maximizar seu lucro esperado.

Podemos também calcular o valor da solução estocástica (VSS) para esse problema. Lembrando: temos que inicialmente calcular a solução ótima para o problema do jornaleiro para  $\xi = 100$ , ou seja, com demanda constante igual

a média de  $\xi$ , isto é, temos que resolver é

$$\min_{0 \leq x \leq 150} \{ cx - q \min\{100, x\} - r \max\{x - 100, 0\} \}.$$

Ao invés de obter o máximo usando cálculo, podemos ver imediatamente pela Figura 2.1 que a solução é  $x^* = 100$ .

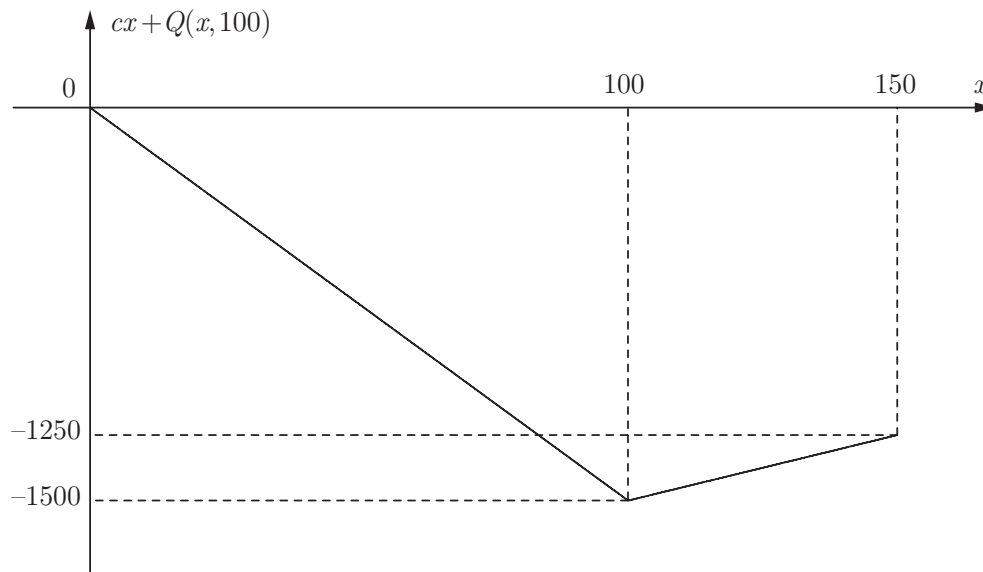


Figura 2.1: Gráfico de  $cx + Q(x, \xi)$  para  $\xi = 100$ .

Uma maneira ainda mais fácil de ver é que se sabemos que a demanda é 100, então devemos comprar  $x = 100$  jornais para maximizar o lucro!

Ainda falta calcular o valor de EEV, que é o valor esperado da solução supondo que o jornaleiro comprou 100 jornais. Para isso fazemos

$$\begin{aligned} \text{EEV} &= \mathbb{E}_{\xi}[10 \cdot 100 + Q(100, \xi)] = 100 \cdot 10 - 25 \cdot 100 + 20 \int_{50}^{100} \frac{\xi - 50}{100} d\xi \\ &= 1000 - 2500 + 20 \left( \frac{75}{2} - 25 \right) = -1250, \end{aligned}$$

que resulta num lucro de R\$ 1 250. Logo, temos que

$$\text{VSS} = 1312.5 - 1250 = 62.5.$$

Por fim, vamos ao cálculo do EVPI. Recordando: para obter o EVPI, supomos que se conhece o futuro, ou seja, que se sabe o valor que demanda

$\xi$ . O valor do EVPI é a esperança com relação a  $\xi$  de todas essas soluções. No problema do fazendeiro, a incerteza estava associada a apenas três tipos de acontecimentos. Aqui a demanda  $\xi$  pode assumir uma quantidade não enumerável de valores. Portanto, teremos que fazer uso da integral para obter o EVPI.

Dado um valor qualquer de demanda  $\xi$ , a solução ótima obviamente é  $x^* = \xi$ . Assim, temos

$$\text{WS} = \mathbb{E}_\xi[c\xi + -q\xi] = -15\mathbb{E}_\xi(\xi) = -1500. \quad (2.5)$$

Conseqüentemente, temos que

$$\text{EVPI} = 1500 - 1312.5 = 187.5.$$

### 2.3 Outras interpretações para o problema

Primeiramente vamos usar o conceito de ganho marginal para derivar a solução do problema por uma outra trilha. A expressão *ganho marginal* em economia se refere ao crescimento no lucro obtido quando se aumenta em uma unidade a quantidade vendida ou adquirida de um determinado bem. Vamos apresentar uma aplicação desse conceito ao problema do jornaleiro que nos permite chegar a resposta (2.3) do problema do jornaleiro de maneira elementar.

Suponha que jornaleiro comprou  $k$  jornais. Qual é o lucro esperado na venda do  $k$ -ésimo jornal? A resposta é

$$\text{lucro esperado} = \mathbb{P}(\xi < k)(r - c) + \mathbb{P}(\xi \geq k)(q - c), \quad (2.6)$$

onde  $\mathbb{P}(\xi < k)$  é probabilidade dele não vender o  $k$ -ésimo jornal e  $\mathbb{P}(\xi \geq k)$  é a probabilidade dele vender este  $k$ -ésimo jornal.

A situação ideal ocorre quando o lucro esperado com a venda do último jornal é zero: se fosse negativo a demanda seria menor que  $k$  (jornal “en-calhado”) e se fosse positivo a demanda seria maior que  $k$  (falta de jornal). Igualando-se a equação (2.6) a zero, temos

$$\begin{aligned} \text{lucro esperado} &= 0 \\ &= \mathbb{P}(\xi < k)(r - c) + \mathbb{P}(\xi \geq k)(q - c) \\ &= F(k)(r - c) + (1 - F(k))(q - c). \end{aligned}$$

Desta maneira,  $F(k) = (q - c)/(q - r)$  e, portanto,

$$k = F^{-1} \left( \frac{q - c}{q - r} \right). \quad (2.7)$$

Assim, o número de jornais a ser comprado para que *em média* todos sejam vendidos é  $k^* = F^{-1}(\frac{q-c}{q-r})$ , a mesma solução encontrada anteriormente.

Uma outra interpretação interessante do problema do jornaleiro, mencionada em [02], surge quando nos perguntamos sobre a probabilidade de se vender todos os jornais para um dada escolha de  $x$ . Esse valor é igual a

$$\mathbb{P}(\{\text{vender tudo}\}) = \mathbb{P}(\xi \geq x) = 1 - F(x).$$

Vamos ver qual é a probabilidade de se vender tudo se comprarmos  $x^*$  jornais:

$$\mathbb{P}(\{\text{vender tudo}\}) = 1 - F(x^*) = 1 - \frac{q - c}{q - r} = \frac{c - r}{q - r}.$$

É comum encontrar na literatura artigos que não permitem que um jornal não vendido seja devolvido ao editor, ou seja,  $r = 0$ . Nesse caso, temos que a quantidade de jornal a ser comprada deve ser escolhida de maneira que a probabilidade de se vender todos os jornais seja igual a razão custo unitário  $c$  do jornal dividido pelo seu preço unitário  $q$ .

## Capítulo 3

# Programação Linear com Coeficientes Aleatórios

Neste capítulo apresentaremos as abordagens clássicas usadas na modelagem e solução de problemas de programação linear onde um ou mais coeficientes são aleatórios (otimização estocástica linear).

Tradicionalmente, são propostos dois tipos de modelos clássicos para se tratar problemas de otimização com coeficientes aleatórios: a abordagem “espere e veja” (em inglês, “wait and see”) e a abordagem “aqui e agora” (em inglês, “here and now”). Em “espere e veja”, o agente de decisão pode esperar por uma realização dos coeficientes aleatórios para tomar a sua decisão. Já em “aqui e agora”, o agente de decisão deve fazer suas escolhas *antes* ou *sem o conhecimento* das realizações dos coeficientes aleatórios. Neste segundo caso, uma dificuldade adicional aparece: sem se conhecer os coeficientes, as definições habituais de admissibilidade e otimalidade não se aplicam e especificações adicionais são necessárias.

A teoria pressupõe que seja dada (conhecida) a distribuição conjunta dos coeficientes. Poder-se-ia argumentar que esta hipótese é restritiva, visto que dificilmente existem dados suficientes para a construção de uma estimativa confiável. Como consequência, é o modelador do problema que acaba fazendo a escolha da distribuição conjunta. Note, contudo, que este tipo de arbitrariedade não é diferente da que uma abordagem determinística faria ao escolher uma realização particular dos coeficientes aleatórios.

### 3.1 O problema da mistura

Vamos começar com um exemplo onde a aleatoriedade se manifesta apenas em alguns dos coeficientes das restrições em desigualdades. Para isto, considere a seguinte situação: um fazendeiro consultou um engenheiro agrônomo que recomendou 7 g de um nutriente A e 4 g de um nutriente B para cada 100 m<sup>2</sup> de terra. O fazendeiro dispõe de dois tipos de adubo. Cada kg do primeiro adubo possui  $\omega_1$  g do nutriente A e  $\omega_2$  g de um nutriente B. Cada kg do segundo adubo, por sua vez, possui 1 g de cada nutriente. Os custos de compra dos dois adubos são iguais: uma unidade monetária por kg. As quantidades  $\omega_1$  e  $\omega_2$  são incertas: o fabricante dos adubos garante que elas são variáveis aleatórias independentes, uniformemente distribuídas e com suportes nos intervalos  $[1, 4]$  e  $[1/3, 1]$ , respectivamente. O problema (da mistura) é então decidir o quanto comprar de cada adubo para atender a necessidade de nutrientes em 100 m<sup>2</sup> de terra minimizando o custo de compra:

$$\begin{array}{ll}
 \text{minimizar} & f(x_1, x_2) = x_1 + x_2 \\
 \text{sujeito a} & \omega_1 x_1 + x_2 \geq 7, \\
 & \omega_2 x_1 + x_2 \geq 4, \\
 & x_1 \geq 0, \\
 & x_2 \geq 0.
 \end{array} \tag{3.1}$$

Note que o conjunto admissível deste programa linear depende dos valores dos coeficientes  $\omega_1$  e  $\omega_2$ .

#### Abordagem “Espere e Veja”

Nesta abordagem, supõe-se que o agente de decisão possa fazer a escolha dos valores de  $\mathbf{x} = (x_1, x_2)$  *depois* da realização de  $\boldsymbol{\omega} = (\omega_1, \omega_2)$ . Desta maneira, o problema (3.1) pode ser considerado um programa linear paramétrico<sup>1</sup>: as soluções ótimas e o valor ótimo são calculados em função de  $\boldsymbol{\omega}$ . Por exemplo:

- (a) Para  $\boldsymbol{\omega} = (\omega_1, \omega_2) = (1, 1/3)$ , o conjunto admissível correspondente é o apresentado na Figura 3.1, a solução ótima é

<sup>1</sup>No endereço <http://www.professores.uff.br/hjbortol/car/activities/problema-da-mistura-01.html> você encontrará um applet JAVA interativo que desenha o conjunto admissível e calcula a solução ótima do problema (3.1) para diferentes valores de  $\omega_1$  e  $\omega_2$ .

$$\mathbf{x}^*(\boldsymbol{\omega}) = \mathbf{x}^*(1, 1/3) = (x_1^*(1, 1/3), x_2^*(1, 1/3)) = (9/2, 5/2)$$

e o valor ótimo é  $v^*(1, 1/3) = x_1^*(1, 1/3) + x_2^*(1, 1/3) = 7$ .

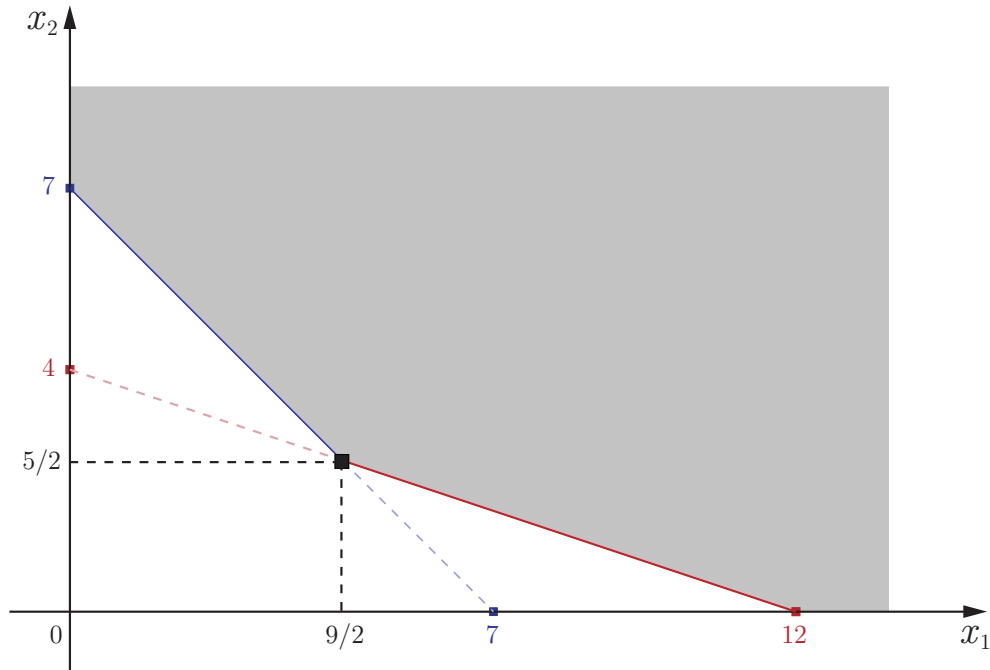


Figura 3.1: Conjunto admissível do problema da mistura para  $\boldsymbol{\omega} = (\omega_1, \omega_2) = (1, 1/3)$ .

- (b) Para  $\boldsymbol{\omega} = (\omega_1, \omega_2) = (5/2, 2/3)$ , o conjunto admissível correspondente é o apresentado na Figura 3.2, a solução ótima é

$$\mathbf{x}^*(\boldsymbol{\omega}) = \mathbf{x}^*(5/2, 2/3) = (x_1^*(5/2, 2/3), x_2^*(5/2, 2/3)) = (18/11, 32/11)$$

e o valor ótimo é  $v^*(5/2, 2/3) = x_1^*(5/2, 2/3) + x_2^*(5/2, 2/3) = 50/11 = 4.\overline{54}$ .

- (c) Para  $\boldsymbol{\omega} = (\omega_1, \omega_2) = (4, 1)$ , o conjunto admissível correspondente é o apresentado na Figura 3.3, a solução ótima é

$$\mathbf{x}^*(\boldsymbol{\omega}) = \mathbf{x}^*(4, 1) = (x_1^*(4, 1), x_2^*(4, 1)) = (1, 3)$$

e o valor ótimo é  $v^*(4, 1) = x_1^*(4, 1) + x_2^*(4, 1) = 4$ .



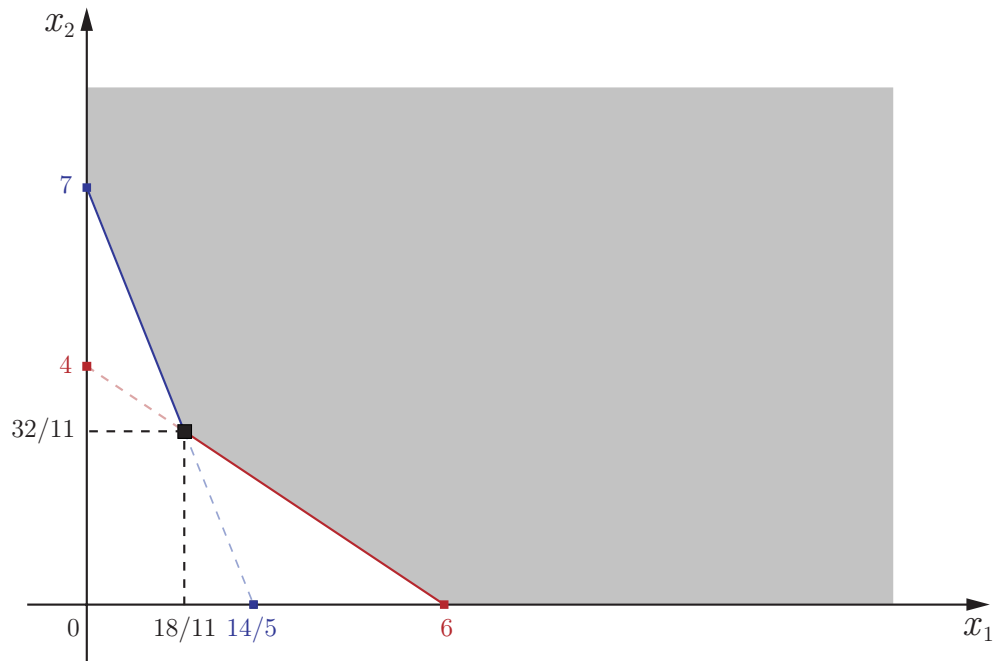


Figura 3.2: Conjunto admissível do problema da mistura para  $\omega = (\omega_1, \omega_2) = (5/2, 2/3)$ .

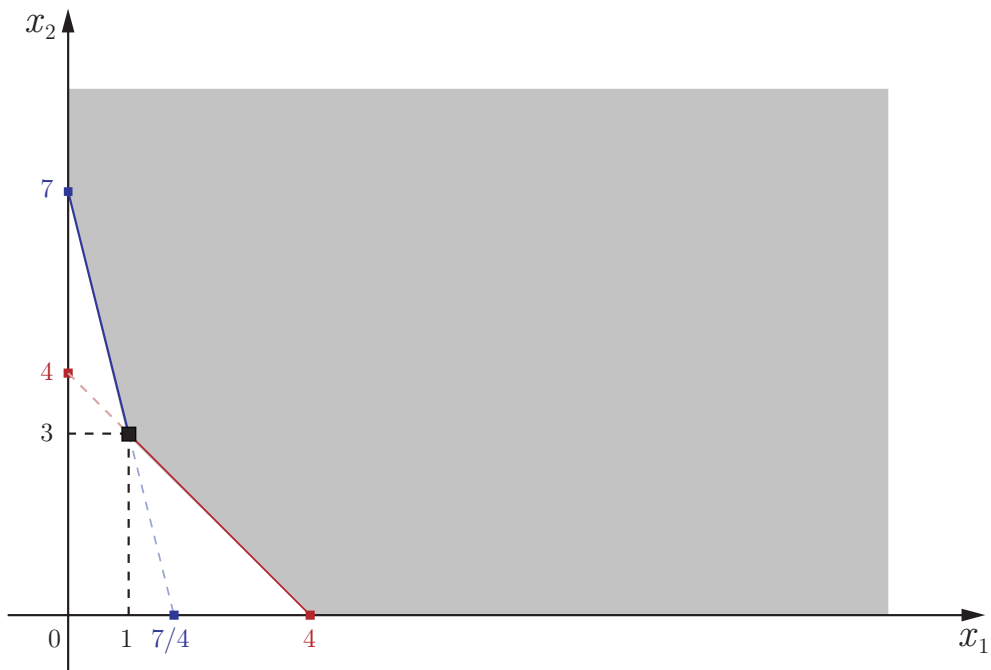


Figura 3.3: Conjunto admissível do problema da mistura para  $\omega = (\omega_1, \omega_2) = (4, 1)$ .

De fato, é possível mostrar (exercício) que a solução ótima do problema (3.1) para  $(\omega_1, \omega_2) \in \Omega = [1, 4] \times [1/3, 1]$  é dada por

$$(x_1^*(\omega_1, \omega_2), x_2^*(\omega_1, \omega_2)) = \begin{cases} \left( \frac{3}{\omega_1 - \omega_2}, \frac{4\omega_1 - 7\omega_2}{\omega_1 - \omega_2} \right), & \text{se } \frac{7}{\omega_1} \leq \frac{4}{\omega_2}, \\ \left( \frac{7}{\omega_1}, 0 \right), & \text{caso contrário,} \end{cases}$$

e que o valor ótimo associado é dado por

$$v^*(x_1^*(\omega_1, \omega_2), x_2^*(\omega_1, \omega_2)) = \begin{cases} \frac{3 + 4\omega_1 - 7\omega_2}{\omega_1 - \omega_2}, & \text{se } \frac{7}{\omega_1} \leq \frac{4}{\omega_2}, \\ \frac{7}{\omega_1}, & \text{caso contrário.} \end{cases}$$

A partir destas expressões, o agente de decisão pode então calcular as distribuições de  $\mathbf{x}^* = (x_1^*(\omega_1, \omega_2), x_2^*(\omega_1, \omega_2))$  e  $v^*(x_1^*(\omega_1, \omega_2), x_2^*(\omega_1, \omega_2))$  e suas características como média, variância, etc. (veja o exercício [03]).

## Abordagem “Aqui e Agora”

Nesta abordagem, o agente de decisão deve fazer a escolha de  $\mathbf{x} = (x_1, x_2)$  *sem conhecer* os valores de  $\boldsymbol{\omega} = (\omega_1, \omega_2)$  (mas sabendo a função distribuição de  $\boldsymbol{\omega}$ ). Sem se conhecer os coeficientes, as definições habituais de admissibilidade e otimalidade não se aplicam e especificações adicionais de modelagem são necessárias. Apresentaremos agora os tipos de especificações mais tradicionais.

### 1. *Abolir incertezas*

O agente de decisão simplesmente faz uma escolha apropriada para  $\boldsymbol{\omega}$  e, então, ele resolve o problema determinístico correspondente.

(a) Escolha “pessimista”:  $\hat{\boldsymbol{\omega}} = (1, 1/3)$ .

Neste caso, o conjunto admissível é o representado na Figura 3.1 e o valor ótimo correspondente é  $\hat{v} = 7$ .

(b) Escolha “neutra”:  $\hat{\omega} = (5/2, 2/2) = \mathbb{E}[(\omega_1, \omega_2)]$ .

Neste caso, o conjunto admissível é o representado na Figura 3.2 e o valor ótimo correspondente é  $\hat{v} = 50/11 = 4.\overline{54}$ .

(c) Escolha “otimista”:  $\hat{\omega} = (4, 1)$ .

Neste caso, o conjunto admissível é o representado na Figura 3.3 e o valor ótimo correspondente é  $\hat{v} = 4$ .

Vantagem da especificação: o problema reformulado é fácil de se resolver, pois ele é um programa linear determinístico. Desvantagem: a solução ótima  $\mathbf{x}^* = (x_1^*, x_2^*)$ , quando implementada, pode não ser admissível.

## 2. Incorporar riscos nas restrições (*chance constraints*)

O agente de decisão descreve uma “medida de risco”, faz uma escolha do “nível máximo de risco aceitável” e, então, ele incorpora estes elementos nas restrições do programa linear. Aqui, o agente de decisão pode ainda escolher entre níveis de confiabilidade individuais ou um nível de confiabilidade conjunto.

(a) Níveis de confiabilidade individuais.

O agente de decisão escolhe dois níveis de confiabilidade individuais  $\alpha_1, \alpha_2 \in [0, 1]$  e ele *decreta* que  $\mathbf{x} = (x_1, x_2) \in [0, +\infty) \times [0, +\infty)$  é admissível se, e somente se,

$$\begin{cases} \mathbb{P}(\omega_1 x_1 + x_2 \geq 7) \geq \alpha_1 \\ \mathbb{P}(\omega_2 x_1 + x_2 \geq 4) \geq \alpha_2 \end{cases}. \quad (3.2)$$

Restrições deste tipo são denominadas *restrições probabilísticas individuais (separadas)* (em inglês, *individual (separate) chance constraints*). Os riscos são definidos em termos da probabilidade de inadmissibilidade, isto é,

$$\begin{cases} \text{risco}_1 := \mathbb{P}(\omega_1 x_1 + x_2 < 7) \\ \text{risco}_2 := \mathbb{P}(\omega_2 x_1 + x_2 < 4) \end{cases}. \quad (3.3)$$

Podemos reescrever as condições (3.2) de forma mais explícita usando as funções distribuição<sup>2</sup>  $F_1$  e  $F_2$  das variáveis  $\omega_1$  e  $\omega_2$ . De fato, é possível

---

<sup>2</sup>No apêndice A você encontrará, entre outros conceitos de probabilidade, a definição de função distribuição de uma variável aleatória.

mostrar (exercício) que se  $0 \leq \alpha_1 < 1$  e  $0 \leq \alpha_2 < 1$ , então

$$\begin{cases} \mathbb{P}(\omega_1 x_1 + x_2 \geq 7) \geq \alpha_1 \\ \mathbb{P}(\omega_2 x_1 + x_2 \geq 4) \geq \alpha_2 \end{cases} \iff \begin{cases} F_1^{-1}(1 - \alpha_1) x_1 + x_2 \geq 7 \\ F_2^{-1}(1 - \alpha_2) x_1 + x_2 \geq 4 \end{cases} \quad (3.4)$$

onde

$$F_i^{-1}(\alpha) := \min_{t \in [-\infty, +\infty)} \{t \mid F_i(t) \geq \alpha\} \quad (3.5)$$

é o  $\alpha$ -ésimo quantil de  $\omega_i$ . Se definirmos  $F_1^{-1}(0) := 1$  e  $F_2^{-1}(0) := 1/3$ , então a equivalência (3.4) é válida mesmo para  $\alpha_1 = 1$  e  $\alpha_2 = 1$ . As desigualdades em (3.4) que usam  $F_i^{-1}$  são denominadas *formas reduzidas* das respectivas restrições probabilísticas individuais.

Com esta abordagem, o problema da mistura (3.1) fica modelado assim:

minimizar $f(x_1, x_2) = x_1 + x_2$ sujeito a $F_1^{-1}(1 - \alpha_1) x_1 + x_2 \geq 7,$ $F_2^{-1}(1 - \alpha_2) x_1 + x_2 \geq 4,$ $x_1 \geq 0,$ $x_2 \geq 0,$	(3.6)
---	-------

isto é, *como um programa linear!* Por exemplo, para os níveis de confiabilidade individuais  $\alpha_1 = \alpha_2 = 2/3$ , verifica-se que

$$F_1^{-1}(1 - \alpha_1) = F_1^{-1}(1/3) = 2 \quad \text{e} \quad F_2^{-1}(1 - \alpha_2) = F_2^{-1}(1/3) = 5/9.$$

Para estes valores<sup>3</sup>, o problema (3.6) se escreve como

minimizar $f(x_1, x_2) = x_1 + x_2$ sujeito a $2x_1 + x_2 \geq 7,$ $5x_1/9 + x_2 \geq 4,$ $x_1 \geq 0,$ $x_2 \geq 0,$	(3.7)
---	-------

que alcança o valor ótimo  $v^* = 64/13 = 4.\overline{923076}$  no ponto (ótimo)  $\mathbf{x}^* = (x_1^*, x_2^*) = (27/13, 37/13) = (2.\overline{076923}, 2.\overline{846153})$ .

---

<sup>3</sup>No endereço <http://www.professores.uff.br/hjbortol/car/activities/problema-da-mistura-02.html>, você encontrará um applet JAVA interativo que desenha o conjunto admissível e calcula a solução ótima do problema (3.6) para diferentes valores dos níveis de confiabilidade individuais  $\alpha_1$  e  $\alpha_2$ .

(b) Nível de confiabilidade conjunto.

O agente de decisão escolhe um nível de confiabilidade conjunto  $\alpha \in [0, 1]$  e ele *decreta* que  $\mathbf{x} = (x_1, x_2) \in [0, +\infty) \times [0, +\infty)$  é admissível se, e somente se,

$$\mathbb{P}(\omega_1 x_1 + x_2 \geq 7 \text{ e } \omega_2 x_1 + x_2 \geq 4) \geq \alpha. \quad (3.8)$$

Restrições deste tipo são denominadas *restrições probabilísticas conjuntas* (em inglês, *joint chance constraints*). O risco é definido como a probabilidade de inadmissibilidade do sistema de restrições do programa linear, isto é, como o número

$$\text{risco} := \mathbb{P}(\omega_1 x_1 + x_2 < 7 \text{ ou } \omega_2 x_1 + x_2 < 4). \quad (3.9)$$

Para obter uma forma reduzida da restrição probabilística conjunta (3.8), note que

$$\begin{aligned} & \mathbb{P}(\omega_1 x_1 + x_2 \geq 7 \text{ e } \omega_2 x_1 + x_2 \geq 4) \\ & \quad \parallel \\ & \mathbb{P}(\omega_1 x_1 + x_2 \geq 7) \cdot \mathbb{P}(\omega_2 x_1 + x_2 \geq 4) \\ & \quad \parallel \\ & \begin{cases} \left(1 - F_1\left(\frac{7-x_2}{x_1}\right)\right) \cdot \left(1 - F_2\left(\frac{4-x_2}{x_1}\right)\right), & \text{se } x_1 > 0, \\ 1, & \text{se } x_1 = 0 \text{ e } x_2 \geq 7, \\ 0, & \text{se } x_1 = 0 \text{ e } 0 \leq x_2 < 7, \end{cases} \end{aligned}$$

onde, na primeira igualdade, usamos o fato de que  $\omega_1$  e  $\omega_2$  são variáveis aleatórias independentes.

Observe que a forma reduzida de uma restrição probabilística conjunta pode ser *não-linear*. Por exemplo, para o nível de confiabilidade conjunto  $\alpha = 2/3$ , pode-se mostrar (exercício) que a restrição

$$\mathbb{P}(\omega_1 x_1 + x_2 \geq 7 \text{ e } \omega_2 x_1 + x_2 \geq 4) \geq \frac{2}{3} \quad (3.10)$$

é equivalente a

$$x_2 \geq \max \left\{ -2x_1 + 7, \frac{11 - 5x_1 + \sqrt{9 - 18x_1 + \frac{43}{3}x_1^2}}{2}, -\frac{5x_1}{9} + 4 \right\} \quad (3.11)$$

Com esta abordagem e para este valor de  $\alpha$ , o problema da mistura (3.1) fica modelado assim:

$$\begin{array}{c}
 \text{minimizar} \\
 f(x_1, x_2) = x_1 + x_2 \\
 \\
 \text{sujeito a} \\
 x_2 \geq \max \left\{ -2x_1 + 7, \frac{11 - 5x_1 + \sqrt{9 - 18x_1 + \frac{43}{3}x_1^2}}{2}, -\frac{5x_1}{9} + 4 \right\}, \\
 x_1 \geq 0, x_2 \geq 0.
 \end{array}
 \tag{3.12}$$

Este problema de otimização não-linear assume o valor mínimo

$$v^* = 220/43 = 5.1162790 \dots$$

no ponto ótimo  $\mathbf{x}^* = (x_1^*, x_2^*) = (54/43, 166/43) = (1.25 \dots, 3.86 \dots)$ . O conjunto admissível<sup>4</sup> de (3.12) é apresentado na Figura 3.4.

### 3. Aceitar inadmissibilidade, penalizando déficits esperados

A idéia aqui é acrescentar à função objetivo parcelas que penalizam inadmissibilidade. Vamos primeiro estabelecer algumas notações. Note que a restrição  $\omega_1 x_1 + x_2 \geq 7$  não é satisfeita se, e somente se,  $\omega_1 x_1 + x_2 - 7 < 0$ . Usando-se a (conveniente) notação

$$z^- = \begin{cases} 0, & \text{se } z \geq 0, \\ -z, & \text{se } z < 0, \end{cases}$$

vemos que uma realização de  $\omega_1$  e escolhas de  $x_1$  e  $x_2$  não satisfazem a restrição  $\omega_1 x_1 + x_2 \geq 7$  se, e somente se,  $(\omega_1 x_1 + x_2 - 7)^- > 0$  (podemos então pensar em  $(\omega_1 x_1 + x_2 - 7)^- > 0$  como uma “medida de inadmissibilidade” para a restrição  $\omega_1 x_1 + x_2 \geq 7$ ). Analogamente,  $\omega_2 x_1 + x_2 \geq 4$  não é satisfeita se, e somente se,  $(\omega_2 x_1 + x_2 - 4)^- > 0$ . Escolhendo-se custos de penalidade unitários  $q_1 > 0$  e  $q_2 > 0$ , as expressões

$$q_1 \mathbb{E}_{\omega_1} [(\omega_1 x_1 + x_2 - 7)^-] \quad \text{e} \quad q_2 \mathbb{E}_{\omega_2} [(\omega_2 x_1 + x_2 - 4)^-]$$

<sup>4</sup>No endereço <http://www.professores.uff.br/hjbortol/car/activities/problema-da-mistura-03.html>, você encontrará um applet JAVA interativo que desenha o conjunto admissível e calcula a solução ótima do problema da mistura usando restrições probabilísticas para diferentes valores dos nível de confiabilidade conjunto  $\alpha$ .

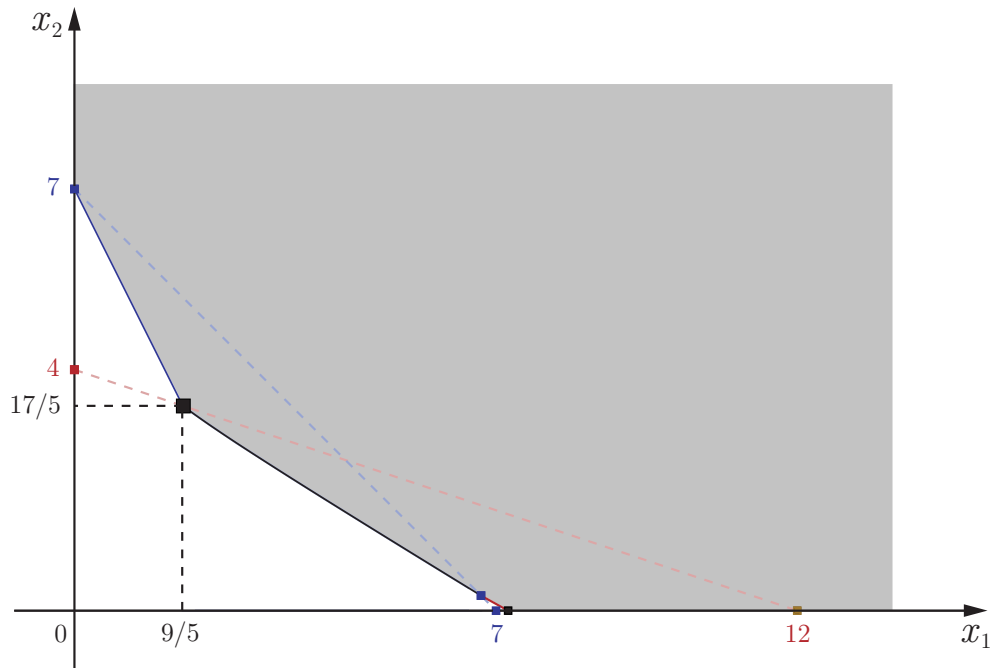


Figura 3.4: Conjunto admissível do problema da mistura usando restrições probabilísticas para o nível de confiabilidade conjunto  $\alpha = 2/3$ .

representam então, respectivamente, os custos médios para inadmissibilidade nas restrições  $\omega_1 x_1 + x_2 \geq 7$  e  $\omega_2 x_1 + x_2 \geq 4$ . Nesta abordagem, o agente de decisão substitui o problema da mistura (3.1) original pelo problema

$$\begin{aligned} & \text{minimizar} \\ & g(x_1, x_2) = x_1 + x_2 + q_1 \mathbb{E}_{\omega_1} [(\omega_1 x_1 + x_2 - 7)^-] + q_2 \mathbb{E}_{\omega_2} [(\omega_2 x_1 + x_2 - 4)^-] \\ & \text{sujeito a } x_1 \geq 0, x_2 \geq 0. \end{aligned}$$

(3.13)

Várias questões surgem neste momento: como calcular as médias (esperanças) envolvidas e como resolver o problema de otimização? Como veremos, o cálculo das esperanças é elementar, mas não-trivial. Apesar de ser não-linear, a função objetivo de (3.13) possui propriedades desejáveis para os algoritmos numéricos em otimização: ela é *convexa* e subdiferenciável. Se os coeficientes aleatórios têm distribuição contínua (como no problema da mistura), o cálculo da esperança é especialmente difícil. Nestes casos, uma prática comum é substituir a distribuição contínua por uma aproximação discreta.

## 3.2 O problema da produção

Vamos estudar agora um programa linear onde a aleatoriedade aparece em uma restrição em igualdade. Mais precisamente, considere o problema (da produção):

$$\begin{array}{ll} \text{minimizar} & f(x) = c x \\ \text{sujeito a} & x = \omega, \\ & x \geq 0, \end{array} \quad (3.14)$$

onde  $c > 0$  é o custo unitário de produção. Este problema de otimização simples modela o processo de minimização do custo de produção  $c x$  sob a restrição de que a produção  $x$  atenda à demanda  $\omega$ . Aqui, vamos supor que  $\omega$  é uma variável aleatória contínua não-negativa com média  $\mu = \mathbb{E}[\omega]$ , variância  $\sigma^2 = \mathbb{E}[(\omega - \mathbb{E}[\omega])^2]$  e função distribuição  $F(t) = \mathbb{P}(\omega \leq t)$ , com  $t \in \mathbb{R}$ .

### Abordagem “Espere e Veja”

Se o agente de decisão pode esperar pela realização da demanda  $\omega$  antes de escolher o valor da produção  $x$ , então o problema é fácil se resolver:  $x^*(\omega) = \omega$  e  $v^*(\omega) = c x^*(\omega) = c \omega$ .

### Abordagem “Aqui e Agora”

#### 1. *Abolir incertezas*

Nesta abordagem, o agente de decisão pode, por exemplo, substituir o valor de  $\omega$  por  $\hat{\omega} = \mu$  ou  $\hat{\omega} = \mu + \Delta$ , onde  $\Delta$  é um “estoque reserva” (por exemplo,  $\Delta = \sigma$  ou  $\Delta = 2\sigma$ ). A probabilidade de que a demanda seja satisfeita (*o nível de serviço da produção*) é então dada por  $\mathbb{P}(\omega \leq \mu + \Delta) = F(\mu + \Delta)$ .

#### 2. *Incorporar riscos nas restrições (chance constraints)*

Construir uma restrição probabilística  $\mathbb{P}(x = \omega) \geq \alpha$  baseada em uma restrição em igualdade ( $x = \omega$ ) é inútil. De fato: se  $\omega$  tem distribuição contínua, então  $\mathbb{P}(x = \omega) = 0$ . Se  $\omega$  tem uma distribuição discreta finita, digamos  $\mathbb{P}(\omega = \omega_i) = p_i$  (com  $p_i \geq 0$  e  $p_1 + \dots + p_n = 1$ ), então  $\mathbb{P}(\omega = x) = 0$  para todo  $x \notin \{\omega_1, \dots, \omega_n\}$ .



Para valores adequados de  $\alpha_1$  e  $\alpha_2$ , também é inútil construir restrições probabilísticas  $\mathbb{P}(x \geq \omega) \geq \alpha_1$  e  $\mathbb{P}(x \leq \omega) \geq \alpha_2$  combinadas, pois não existe produção  $x$  que satisfaça as condições

$$\begin{cases} \mathbb{P}(x \geq \omega) \geq \alpha_1 \\ \mathbb{P}(x \leq \omega) \geq \alpha_2 \end{cases} \iff \begin{cases} F(x) \geq \alpha_1 \\ 1 - F(x) \geq \alpha_2 \end{cases} \iff \alpha_1 \leq F(x) \leq 1 - \alpha_2$$

se, por exemplo,  $\alpha_1 = \alpha_2 = 3/4$ , uma vez que a função distribuição  $F$  é não-decrescente.

Desta maneira, é preciso estabelecer prioridades. Podemos, por exemplo, especificar um nível de confiabilidade mínimo  $\alpha \in (1/2, 1)$  e modelar o problema (3.14) na forma

$$\min_{x \geq 0} \{cx \mid \mathbb{P}(x \geq \omega) \geq \alpha\} = \min_{x \geq 0} \{cx \mid x \geq F^{-1}(\alpha)\}$$

cuja solução é, evidentemente,  $x^* = F^{-1}(\alpha)$ .

### 3. Aceitar inadmissibilidade, penalizando desvios esperados

Aqui devemos penalizar tanto déficits quanto superávits na produção: usando-se as notações

$$z^- = \begin{cases} 0, & \text{se } z \geq 0, \\ -z, & \text{se } z < 0, \end{cases} \quad \text{e} \quad z^+ = \begin{cases} z, & \text{se } z \geq 0, \\ 0, & \text{se } z < 0, \end{cases} \quad (3.15)$$

isto é feito considerando-se o seguinte problema de otimização

$$\boxed{\begin{array}{ll} \text{minimizar} & f(x) = cx + \mathcal{Q}(x) \\ \text{sujeito a} & x \geq 0, \end{array}} \quad (3.16)$$

onde

$$\mathcal{Q}(x) = \mathbb{E} [h \cdot (\omega - x)^- + q \cdot (\omega - x)^+],$$

com  $h$  e  $q$  custos unitários de superávit e déficit na produção, respectivamente ( $h < c < q$ ).

Não é difícil de se mostrar (exercício) que  $\mathcal{Q}$  é uma função convexa e diferenciável. De fato,

$$\mathcal{Q}'(x) = -q + (q + h)F(x), \quad \text{para } x \in \mathbb{R}.$$

Desta maneira, a solução ótima de (3.16) é obtida resolvendo-se a equação  $f'(x) = c + \mathcal{Q}'(x) = c - q + (q + h)F(x) = 0$ :

$$x^* = F^{-1} \left( \frac{q - c}{q + h} \right).$$

Note que esta solução tem a mesma forma da solução obtida via restrições probabilísticas. De fato, se  $h = 0$ , a mesma solução é obtida se  $q/c = 1/(1 - \alpha)$ :

$\alpha$ (nível de confiabilidade)	$q/c$ (custo de déficit/custo de produção)
0.990	100
0.975	40
0.950	20
0.900	10
0.800	5
0.500	2

Esta tabela é interessante: ela nos dá uma idéia de que valores escolher para o custo  $q$  em termos do nível de confiabilidade  $\alpha$ .

## Exercícios

- [01] Deduza as equações para a solução ótima  $(x_1^*, x_2^*)$  do problema da mistura apresentadas na página 19.
- [02] Mostre  $4 \leq v^*(x_1^*(\omega_1, \omega_2), x_2^*(\omega_1, \omega_2)) \leq 7$  para todo  $(\omega_1, \omega_2)$  no conjunto  $\Omega = [1, 4] \times [1/3, 1]$ , onde  $v^* = v^*(x_1^*(\omega_1, \omega_2), x_2^*(\omega_1, \omega_2))$  é o valor ótimo do problema da mistura.
- [03] Mostre que a função distribuição  $F$  do valor ótimo  $v^*$  do problema da mistura é dada por

$$F(t) = \begin{cases} 0, & \text{se } 0 \leq t \leq 4, \\ \frac{8t^3 - 18t^2 - 105t + 196}{4t^2(7-t)}, & \text{se } 4 \leq t \leq 50/11, \\ \frac{49t^3 - 307t^2 + 648t - 1008}{36t^2(t-4)}, & \text{se } 50/11 \leq t \leq 7, \\ 1, & \text{se } t \geq 7. \end{cases}$$

Note que, com esta expressão para  $F$  em mãos, podemos então calcular a função de densidade de  $v^*$  (derivando-se  $F$ ):

$$f(t) = \begin{cases} 0, & \text{se } 0 < t < 4, \\ \frac{38t^3 - 210t^2 + 1323t - 2744}{4t^3(t-7)^2}, & \text{se } 4 < t < 50/11, \\ \frac{37t^3 - 432t^2 + 1872t - 2688}{12t^3(t-4)^2}, & \text{se } 50/11 < t < 7, \\ 0, & \text{se } t > 7, \end{cases}$$

e a sua média:

$$\begin{aligned} \mathbb{E}[v^*(\omega)] &= \int_{-\infty}^{+\infty} t f(t) dt = 4 + \frac{27}{2} \ln 3 - \frac{35}{6} \ln 2 - \frac{77}{12} \ln 11 + \frac{11}{4} \ln 7 \\ &= 4.7526655 \dots \end{aligned}$$

[04] Mostre a equivalência entre (3.10) e (3.11) para  $x_1 \geq 0$  e  $x_2 \geq 0$ .

[05] Mostre que o problema (3.13) pode ser reescrito na forma

$$\boxed{\begin{array}{l} \text{minimizar } g(x_1, x_2) = x_1 + x_2 + Q(x_1, x_2) \\ \text{sujeito a } \quad \quad \quad x_1 \geq 0, \quad x_2 \geq 0, \end{array}}$$

onde

$$Q(x_1, x_2) = \mathbb{E} \left[ \min_{y_1 \geq 0, y_2 \geq 0} \left\{ q_1 y_1 + q_2 y_2 \mid \begin{array}{l} \omega_1 x_1 + x_2 + y_1 \geq 7 \\ \omega_2 x_1 + x_2 + y_2 \geq 4 \end{array} \right\} \right].$$

[06] Prove as seguintes relações para os operadores  $z \mapsto z^+$  e  $z \mapsto z^-$  definidos em (3.15):

$$\begin{aligned} z^+ &= \max\{0, z\}, & z^- &= \max\{0, -z\} = -\min\{0, z\}, \\ (-z)^- &= z^+, & (cz)^+ &= cz^+ \text{ para todo } c \geq 0, & z^+ + z^- &= |z|, \\ z^+ - z^- &= z, & z^+ &= (|z| + z)/2, & z^- &= (|z| - z)/2, \\ z^+ \cdot z^- &= 0, & \max\{z^+, z^-\} &= |z|, & \min\{z^+, z^-\} &= 0, \\ (x^+ + y^+)^+ &= \max\{x + y, y, 0\}, & x^+ + y^+ &= \max\{x + y, x, y, 0\}, \\ x^+ - |y| &\leq (x - y)^+ \leq x^+ + |y|, & x^- - |y| &\leq (x - y)^- \leq x^- + |y|. \end{aligned}$$

[07] Mostre que (3.16) pode ser reformulado como um modelo de recurso:

$$\min_{x \geq 0} \left\{ cx + \mathbb{E} \left[ \min_{y_1 \geq 0, y_2 \geq 0} \left\{ qy_1 + hy_2 \mid x + y_1 - y_2 = \omega \right\} \right] \right\}.$$

# Capítulo 4

## Modelos de Recurso

Neste capítulo nos concentraremos na abordagem “aqui e agora” que aceita inadmissibilidade penalizando desvios médios. De fato, veremos que esta abordagem motiva uma classe importante de modelos em otimização estocástica: os modelos de recurso.

### 4.1 Motivação: programação linear por metas

Em problemas determinísticos, a técnica de programação linear por metas (em inglês, *goal programming*) consiste em classificar (separar) as restrições do problema em dois tipos: as restrições rígidas (*hard constraints*) que não podem ser violadas de maneira alguma e as restrições flexíveis (*soft constraints*) que podem ser violadas, mas *não a qualquer preço*. Mais precisamente, considere o programa linear determinístico:

$$\min_{\mathbf{x} \in \mathbf{X}} \{ \mathbf{c}\mathbf{x} \mid \mathbf{A}\mathbf{x} = \mathbf{b} \text{ e } \mathbf{T}\mathbf{x} \sim \mathbf{h} \}, \quad (4.1)$$

onde

- $\mathbf{X} = \{ \mathbf{x} \in \mathbb{R}^n \mid \underline{\mathbf{x}} \leq \mathbf{x} \leq \bar{\mathbf{x}} \}$  ou  $\mathbf{X} = \{ \mathbf{x} \in \mathbb{R}^n \mid \mathbf{0} \leq \mathbf{x} < +\infty \}$  (as desigualdades entre vetores devem ser interpretadas componente a componente),
- $\mathbf{c} \in \mathbb{R}^n$ ,  $\mathbf{A}$  é uma matriz  $\tilde{m} \times n$ ,  $\mathbf{b} \in \mathbb{R}^{\tilde{m}}$ ,  $\mathbf{T}$  é uma matriz  $m \times n$ ,  $\mathbf{h} \in \mathbb{R}^m$ ,
- $\mathbf{c}\mathbf{x} = \sum_{i=1}^n c_i \cdot x_i$  e

- o símbolo  $\sim$  representa uma das relações  $=$ ,  $\geq$  e  $\leq$  (componente a componente).

Neste programa linear, consideraremos  $\mathbf{Ax} = \mathbf{b}$  como restrições rígidas e  $\mathbf{T}\mathbf{x} \sim \mathbf{h}$  como restrições flexíveis. Como antes, a idéia é penalizar os *desvios de meta*  $\mathbf{z} = \mathbf{h} - \mathbf{T}\mathbf{x}$  das restrições flexíveis através de uma *função de penalidade*  $\mathbf{z} \mapsto v(\mathbf{z})$  que é incorporada à função objetivo do problema de otimização original:

$$\min_{\mathbf{x} \in \mathbf{X}} \{ \mathbf{c}\mathbf{x} + v(\mathbf{h} - \mathbf{T}\mathbf{x}) \mid \mathbf{Ax} = \mathbf{b} \} = \min_{\mathbf{x} \in \mathbf{X}} \{ \mathbf{c}\mathbf{x} + v(\mathbf{z}) \mid \mathbf{Ax} = \mathbf{b} \text{ e } \mathbf{T}\mathbf{x} + \mathbf{z} = \mathbf{h} \}. \quad (4.2)$$

A função de penalidade fornece uma medida do quanto se deve pagar pela violação das metas (restrições)  $\mathbf{z} \sim \mathbf{0}$  frente ao custo original  $\mathbf{c}\mathbf{x}$ . Existem várias maneiras de se especificar a função de penalidade  $v$ , o que torna o método flexível. Vamos ver algumas delas agora.

### 1. Função de penalidade com custos individuais

Escrevendo-se

$$\mathbf{T} = \begin{bmatrix} \boxed{\mathbf{t}_1} \\ \boxed{\mathbf{t}_2} \\ \vdots \\ \boxed{\mathbf{t}_m} \end{bmatrix}, \quad \mathbf{h} = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_m \end{bmatrix} \quad \text{e} \quad \mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{bmatrix},$$

vemos que a notação vetorial  $\mathbf{T}\mathbf{x} \sim \mathbf{h}$  (respectivamente,  $\mathbf{z} \sim \mathbf{0}$ ) é uma maneira compacta e conveniente de se representar as  $m$  restrições escalares:  $\mathbf{t}_i\mathbf{x} \sim h_i$  (respectivamente,  $z_i \sim 0$ ) para  $i = 1, \dots, m$ . Aqui

$$\mathbf{t}_i = (T_{i1}, T_{i2}, \dots, T_{in})$$

representa a  $i$ -ésima linha da matriz  $\mathbf{T}$  e o produto  $\mathbf{t}_i\mathbf{x}$  deve ser entendido como o produto escalar  $\sum_{j=1}^n T_{ij} x_j$ .

Podemos então construir uma função de penalidade  $v$  que é caracterizada por custos de penalidade individuais, que podem ser diferentes para superávits e déficits:

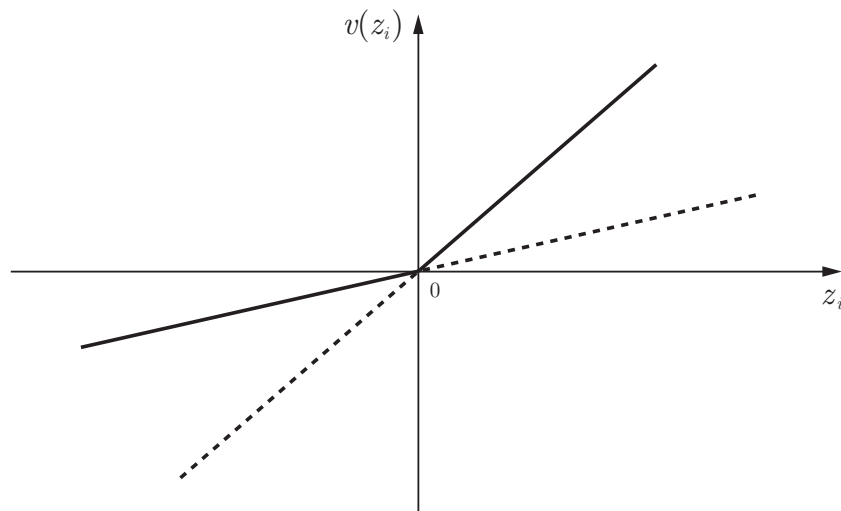
$$v(\mathbf{z}) = \sum_{i=1}^m v_i(z_i) = \sum_{i=1}^m \underbrace{(\bar{q}_i z_i^+ + \underline{q}_i z_i^-)}_{v_i(z_i)}. \quad (4.3)$$

A especificação dos custos de penalidade unitários  $\bar{q}_i$  e  $\underline{q}_i$  podem seguir as seguintes diretivas:

- Se a restrição é do tipo  $\mathbf{t}_i \mathbf{x} = h_i$ , isto é,  $z_i = 0$ , penalizamos superávits e déficits escolhendo  $\bar{q}_i > 0$  e  $\underline{q}_i > 0$ . Note que, neste caso, a função  $v_i$  é convexa como soma de funções convexas.
- Se a restrição é do tipo  $\mathbf{t}_i \mathbf{x} \geq h_i$ , isto é,  $z_i \leq 0$ , penalizamos superávits escolhendo  $\bar{q}_i > 0$  e premiamos déficits escolhendo  $\underline{q}_i \leq 0$ . Para obter convexidade, as escolhas de  $\bar{q}_i$  e  $\underline{q}_i$  devem ser tais que  $\underline{q}_i + \bar{q}_i \geq 0$ . De fato: observe que

$$v_i(z_i) = \bar{q}_i z_i^+ + \underline{q}_i z_i^- = \begin{cases} +\bar{q}_i z_i, & \text{se } z_i \geq 0, \\ -\underline{q}_i z_i, & \text{se } z_i < 0. \end{cases}$$

Desta maneira,  $v_i$  é convexa se a inclinação da reta  $z_i \mapsto -\underline{q}_i z_i$  é menor ou igual do que a inclinação da reta  $z_i \mapsto +\bar{q}_i z_i$ , isto é, se  $-\underline{q}_i \leq +\bar{q}_i$  ou, ainda,  $\underline{q}_i + \bar{q}_i \geq 0$ .



- Se a restrição é do tipo  $\mathbf{t}_i \mathbf{x} \leq h_i$ , isto é,  $z_i \geq 0$ , penalizamos déficits escolhendo  $\underline{q}_i > 0$  e premiamos superávits escolhendo  $\bar{q}_i \leq 0$ . Novamente, para que a função  $v_i$  seja convexa, é necessário que  $\underline{q}_i + \bar{q}_i \geq 0$ .

## 2. Função de penalidade com custos individuais refinados

Considere o caso de restrições do tipo  $\mathbf{t}_i \mathbf{x} = h_i$ , isto é,  $z_i = 0$ . Como antes, a função  $v$  é construída usando-se custos de penalidade individuais mas, agora,

desvios muito grandes, digamos, fora de um intervalo  $[-l_i, +u_i]$  contendo 0, receberão uma penalidade extra:

$$v_i(z_i) = \bar{q}_i^{(1)} z_i^+ + \left( \bar{q}_i^{(2)} - \bar{q}_i^{(1)} \right) (z_i - u_i)^+ + \underline{q}_i^{(1)} z_i^- + \left( \underline{q}_i^{(2)} - \underline{q}_i^{(1)} \right) (z_i + l_i)^-. \quad (4.4)$$

Note que esta função de penalidade também pode ser usada para se modelar restrições do tipo  $z_i \in [-l_i, +u_i]$  bastando, para isto, tomar  $\underline{q}_i^{(1)} = \bar{q}_i^{(1)} = 0$ .

### 3. Função de penalidade com custo conjunto

A seguinte função de penalidade pode ser usada se as restrições do programa linear são do tipo  $z_i \geq 0$  e se o desvio máximo é mais importante do que a soma ponderada dos desvios individuais:

$$v(\mathbf{z}) = v(z_1, z_2, \dots, z_m) = q_0 \max\{z_1^-, z_2^-, \dots, z_m^-\}. \quad (4.5)$$

Como o máximo de funções convexas resulta em uma função convexa, segue-se que  $v$  é uma função convexa de  $q_0 > 0$ .

### 4. Função de penalidade via ações de recurso

Este quarto exemplo de função de penalidade é motivado pela idéia de correções ( $\mathbf{y}$ ) para compensar desvios ( $\mathbf{z}$ ) gerados por decisões ( $\mathbf{x}$ ) tomadas a priori. Para construir a função de penalidade são necessários os seguintes ingredientes:

1. Uma *estrutura de recurso*  $(\mathbf{q}, \mathbf{W})$ .

Aqui  $\mathbf{q} \in \mathbb{R}^p$  e  $\mathbf{W}$  é uma matriz  $m \times p$ . A matriz  $\mathbf{W}$  é denominada *matriz de recurso* (ou *matriz de tecnologia*) e o vetor  $\mathbf{q}$  especifica os coeficientes do custo da ação de recurso.

2. Um conjunto  $\mathbf{Y}$  de variáveis de recurso.

Em geral,  $\mathbf{Y} = \{\mathbf{y} \in \mathbb{R}^p \mid \underline{\mathbf{y}} \leq \mathbf{y} \leq \bar{\mathbf{y}}\}$  ou  $\mathbf{Y} = \{\mathbf{y} \in \mathbb{R}^p \mid \mathbf{0} \leq \mathbf{y} \leq +\infty\} = \mathbb{R}_+^p$ .

A *função de recurso*  $v$  dá o custo mínimo das ações de recurso necessárias para compensar o desvio  $\mathbf{z} \in \mathbb{R}^m$  nas restrições  $\mathbf{T}\mathbf{x} \sim \mathbf{h}$ :

$$v(\mathbf{z}) = \min_{\mathbf{y} \in \mathbf{Y}} \{\mathbf{q}\mathbf{y} \mid \mathbf{W}\mathbf{y} \sim \mathbf{z}\}. \quad (4.6)$$



De fato, todas as funções de penalidade (4.3), (4.4) e (4.5) apresentadas anteriormente podem ser representadas deste modo, isto é, como uma função de penalidade via ações de recursos para estruturas de recurso  $(\mathbf{q}, \mathbf{W})$  e conjuntos de ações de recurso  $\mathbf{Y}$  adequados:

- Se  $p = 2m$ ,

$$\mathbf{q} = (\bar{\mathbf{q}}, \underline{\mathbf{q}}) \in \mathbb{R}^m \times \mathbb{R}^m = \mathbb{R}^p, \quad \mathbf{W} = \begin{bmatrix} +\mathbf{I}_{m \times m} & -\mathbf{I}_{m \times m} \end{bmatrix}_{m \times 2m} \text{ e}$$

$$\mathbf{Y} = \mathbb{R}_+^m \times \mathbb{R}_+^m = \mathbb{R}_+^p,$$

então a função de recuso  $v$  em (4.6) obtém a função de penalidade em (4.3). Aqui  $\mathbf{I}_{m \times m}$  denota a matriz identidade de tamanho  $m \times m$ .

- Se  $p = 4m$ ,

$$\mathbf{q} = (\bar{\mathbf{q}}^{(2)}, \bar{\mathbf{q}}^{(1)}, \underline{\mathbf{q}}^{(1)}, \underline{\mathbf{q}}^{(2)}) \in \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^m = \mathbb{R}^p,$$

$$\mathbf{W} = \begin{bmatrix} +\mathbf{I}_{m \times m} & +\mathbf{I}_{m \times m} & -\mathbf{I}_{m \times m} & -\mathbf{I}_{m \times m} \end{bmatrix}_{m \times 4m} \text{ e}$$

$$\mathbf{Y} = \{(\bar{\mathbf{y}}^{(2)}, \bar{\mathbf{y}}^{(1)}, \underline{\mathbf{y}}^{(1)}, \underline{\mathbf{y}}^{(2)}) \in \mathbb{R}_+^m \times \mathbb{R}_+^m \times \mathbb{R}_+^m \times \mathbb{R}_+^m \mid \bar{\mathbf{y}}^{(1)} \leq \mathbf{u} \text{ e } \underline{\mathbf{y}}^{(1)} \leq \mathbf{l}\}$$

então a função de recuso  $v$  em (4.6) obtém a função de penalidade em (4.4).

- Se  $p = 1$ ,  $\mathbf{q} = (q_0)$ ,  $\mathbf{W} = [-1 \ -1 \ \dots \ -1]^T$  (de tamanho  $m \times 1$ ) e  $\mathbf{Y} = \mathbb{R}_+$  então a função de recuso  $v$  em (4.6) obtém a função de penalidade em (4.5).

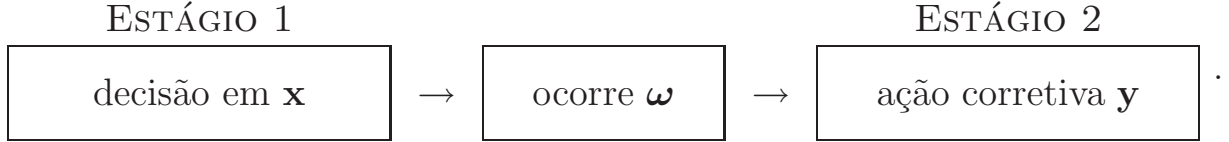
## 4.2 Modelos de recurso em otimização estocástica

Nesta seção veremos como usar a idéia de função de penalidade via ações de recurso para criar um modelo para a seguinte versão estocástica do programa linear determinístico (4.1):

$$\min_{\mathbf{x} \in \mathbf{X}} \{ \mathbf{c}\mathbf{x} \mid \mathbf{A}\mathbf{x} = \mathbf{b} \text{ e } \mathbf{T}(\boldsymbol{\omega})\mathbf{x} \sim \mathbf{h}(\boldsymbol{\omega}) \}, \quad (4.7)$$

De fato, o uso de ações de recurso é muito adequado para uma modelagem do tipo “aqui e agora” para o problema (4.7). Como o agente de decisão deve fazer a escolha da variável  $\mathbf{x}$  sem conhecer os valores de  $\boldsymbol{\omega}$ , podemos pensar nas restrições estocásticas  $\mathbf{T}(\boldsymbol{\omega})\mathbf{x} \sim \mathbf{h}(\boldsymbol{\omega})$  como restrições flexíveis,

que serão ou não satisfeitas dependendo das realizações de  $\omega$ . Os desvios correspondentes são, então, penalizados via uma função de penalidade com ações de recurso:



Aplicando então a estrutura de recurso, obtemos o assim denominado *modelo de recurso em dois estágios* para o problema (4.7):

$$\min_{\mathbf{x} \in \mathbf{X}} \left\{ \mathbf{c}\mathbf{x} + \mathbb{E} \left[ \min_{\mathbf{y} \in \mathbf{Y}} \{ \mathbf{q}(\omega)\mathbf{y} \mid \mathbf{W}(\omega)\mathbf{y} \sim \mathbf{h}(\omega) - \mathbf{T}(\omega)\mathbf{x} \} \right] \mid \mathbf{A}\mathbf{x} = \mathbf{b} \right\}. \quad (4.8)$$

Podemos obter uma formulação mais compacta de (4.8) através da *função de penalidade via ações de recurso* (também denominada *função de valor do segundo estágio*)

$$v(\mathbf{z}, \omega) = \min_{\mathbf{y} \in \mathbf{Y}} \{ \mathbf{q}(\omega)\mathbf{y} \mid \mathbf{W}(\omega)\mathbf{y} \sim \mathbf{z} \} \quad (4.9)$$

e da *função de custo de recurso mínimo esperado* (também denominada *função de valor esperado*)

$$\mathcal{Q}(\mathbf{x}) = \mathbb{E} [v(\mathbf{h}(\omega) - \mathbf{T}(\omega)\mathbf{x}, \omega)]. \quad (4.10)$$

De fato, com estas funções, não é difícil de se ver que o problema (4.8) é equivalente a

$$\min_{\mathbf{x} \in \mathbf{X}} \{ \mathbf{c}\mathbf{x} + \mathcal{Q}(\mathbf{x}) \mid \mathbf{A}\mathbf{x} = \mathbf{b} \}. \quad (4.11)$$

**OBSERVAÇÃO 1.** É muito importante notar que tanto no programa linear do segundo estágio

$$\min_{\mathbf{y} \in \mathbf{Y}} \{ \mathbf{q}(\omega)\mathbf{y} \mid \mathbf{W}(\omega)\mathbf{y} \sim \mathbf{h}(\omega) - \mathbf{T}(\omega)\mathbf{x} \}$$

como no cálculo da função de valor do segundo estágio (4.9), o valor de  $\omega$  está fixo! Neste sentido, os programas lineares envolvidos são determinísticos ( $\omega$  é considerado um parâmetro)!

**OBSERVAÇÃO 2.** Como os exercícios [05] e [07] do capítulo 3 pedem para mostrar, os problemas (3.13) e (3.16) podem ser considerados como modelos de recurso.

OBSERVAÇÃO 3. No problema do fazendeiro, as variáveis de primeiro estágio correspondem à distribuição de terra em milho, trigo e cana-de-açúcar ( $x_1, x_2$  e  $x_3$ ). Elas devem ser escolhidas por João *antes* de ele saber quais serão os rendimentos destes cultivos. As variáveis de segundo estágio correspondem às vendas e às compras destes três cultivos no mercado local ( $w_1, y_1, w_2, y_2, w_3$  e  $w_4$ ). Elas devem ser tomadas como ações corretivas que orientam João a como vender e comprar de maneira ótima no mercado local *depois* da colheita.

OBSERVAÇÃO 4. Sem perda de generalidade, podemos supor que  $\mathbf{T}(\boldsymbol{\omega})$ ,  $\mathbf{h}(\boldsymbol{\omega})$ ,  $\mathbf{W}(\boldsymbol{\omega})$  e  $\mathbf{q}(\boldsymbol{\omega})$  dependem *linearmente* de  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_r) \in \mathbb{R}^r$ :

$$\begin{aligned} \mathbf{T}(\boldsymbol{\omega}) &= \mathbf{T}_0 + \sum_{k=1}^r \omega_k \cdot \mathbf{T}_k, & \mathbf{h}(\boldsymbol{\omega}) &= \mathbf{h}_0 + \sum_{k=1}^r \omega_k \cdot \mathbf{h}_k, \\ \mathbf{W}(\boldsymbol{\omega}) &= \mathbf{W}_0 + \sum_{k=1}^r \omega_k \cdot \mathbf{W}_k & \text{e} & \quad \mathbf{q}(\boldsymbol{\omega}) = \mathbf{q}_0 + \sum_{k=1}^r \omega_k \cdot \mathbf{q}_k. \end{aligned}$$

Aqui  $\mathbf{T}_k$ ,  $\mathbf{h}_k$ ,  $\mathbf{W}_k$  e  $\mathbf{q}_k$  são todos constantes (isto é, não-estocásticos), com dimensões  $m \times n$ ,  $m$ ,  $m \times p$  e  $p$ , respectivamente.

### 4.3 Admissibilidade

Nem todo programa linear tem solução (o conjunto admissível pode ser vazio ou a função objetivo pode ser ilimitada neste conjunto) e nem toda variável aleatória tem média finita. Nesta seção apresentaremos alguns resultados sobre a admissibilidade e finitude do modelo de recurso em dois estágios (4.8). As demonstrações serão omitidas, pois elas usam ferramentas de programação linear e teoria da medida que fogem do escopo deste texto. O leitor interessado pode consultar os livros [03, 13, 14].

Nas considerações que se seguem, vamos supor que a matriz de tecnologia  $\mathbf{W}$  é determinística, isto é, vamos supor que ela não depende de  $\boldsymbol{\omega}$ :

$$\mathbf{W}(\boldsymbol{\omega}) = \mathbf{W} = \text{constante.}$$

Quando isto acontece, dizemos que  $(\mathbf{q}(\boldsymbol{\omega}), \mathbf{W}(\boldsymbol{\omega}))$  possui uma *estrutura de recurso fixa*. Seja então

$$\boldsymbol{\xi}(\boldsymbol{\omega}) = (\mathbf{q}(\boldsymbol{\omega}), \mathbf{h}(\boldsymbol{\omega}), \mathbf{t}_1(\boldsymbol{\omega}), \dots, \mathbf{t}_m(\boldsymbol{\omega})) \in \mathbb{R}^{p+m+(m \times n)}$$

o vetor aleatório formado por todas as componentes aleatórias do proble-

ma (4.8) e seja  $\Xi$  o suporte de  $\xi$ , isto é, o menor subconjunto fechado de  $\mathbb{R}^{p+m+(m \times n)}$  satisfazendo a condição  $\mathbb{P}(\xi(\omega) \in \Xi) = 1$ . Aqui  $\mathbf{t}_i(\omega)$  denota a  $i$ -ésima linha da matriz  $\mathbf{T}(\omega)$ .

Ao estudarmos questões de admissibilidade e finitude em (4.8), é natural considerarmos:

1. O conjunto de decisões  $\mathbf{x}$  que satisfazem as restrições (rígidas) do primeiro estágio:

$$K_1 = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} = \mathbf{b}\}.$$

2. O conjunto de decisões  $\mathbf{x}$  de primeiro estágio para as quais a função de valor esperado é finita:

$$K_2 = \{\mathbf{x} \in \mathbb{R}^n \mid Q(\mathbf{x}) = \mathbb{E}[v(\mathbf{h}(\omega) - \mathbf{T}(\omega)\mathbf{x}, \omega)] < +\infty\}.$$

Com estes conjuntos, o problema (4.8) se reescreve da seguinte maneira:

$\begin{array}{ll} \text{minimizar} & \mathbf{c}\mathbf{x} + Q(\mathbf{x}) \\ \text{sujeito a} & \mathbf{x} \in K_1 \cap K_2. \end{array}$
--

Note, contudo, que o cálculo do conjunto  $K_2$  pode não ser uma tarefa fácil, por conta da esperança envolvida. Mais fáceis de se calcular são:

3. O conjunto de todas as decisões  $\mathbf{x}$  de primeiro estágio para as quais o programa linear do segundo estágio é admissível para um valor de  $\xi \in \Xi$  fixo<sup>1</sup>:

$$K_2(\xi) = \{\mathbf{x} \in \mathbb{R}^n \mid Q(\mathbf{x}, \xi) = v(\mathbf{h}(\omega) - \mathbf{T}(\omega)\mathbf{x}, \omega) < +\infty\}.$$

4. O conjunto de todas as decisões  $\mathbf{x}$  de primeiro estágio para as quais o programa linear do segundo estágio é admissível para “muitos” valores de  $\xi \in \Xi$ :

$$K_2^P = \{\mathbf{x} \in \mathbb{R}^n \mid \text{“}\forall \xi \in \Xi\text{”, } Q(\mathbf{x}, \xi) < +\infty\} = \bigcap_{\text{“}\xi \in \Xi\text{”}} K_2(\xi).$$

Aqui, “ $\xi \in \Xi$ ” significa para todo  $\xi \in \Xi$  exceto possivelmente para valores de  $\xi$  em um conjunto de probabilidade 0.

---

<sup>1</sup>Note que  $v(\mathbf{h}(\omega) - \mathbf{T}(\omega)\mathbf{x}, \omega)$  pode ser escrito em termos de  $\xi \in \Xi$ , já que  $\Xi$  é o suporte do conjunto  $\{(\mathbf{q}(\omega), \mathbf{h}(\omega), \mathbf{t}_1(\omega), \dots, \mathbf{t}_m(\omega)) \in \mathbb{R}^{p+m+(m \times n)} \mid \omega \in \Omega\}$ .

O propósito do conjunto  $K_2^P$  é o de construir um mecanismo que permita identificar se  $Q(\mathbf{x}) < +\infty$  sem que, para isto, seja necessário calcular a esperança  $\mathbb{E}[Q(\mathbf{x}, \boldsymbol{\xi})]$ . Os próximos teoremas descrevem propriedades geométricas de  $K_2^P$  e  $K_2$  e estabelecem condições suficientes para que  $K_2 = K_2^P$ .

**Teorema 4.1** O conjunto  $K_2(\boldsymbol{\xi})$  é um polítopo convexo e fechado para todo  $\boldsymbol{\xi} \in \Xi$  e, em particular,  $K_2^P$  é fechado e convexo. Se  $\Xi$  é finito, então  $K_2^P$  também é um polítopo e  $K_2^P = K_2$ .

**Teorema 4.2** Suponha que  $\mathbf{W}(\boldsymbol{\omega}) = \mathbf{W}$  (recurso fixo) e que  $\boldsymbol{\xi}$  tenha segundos momentos finitos. Então:

- (a) Se  $\mathbb{P}(\boldsymbol{\omega} \in \Omega \mid Q(\mathbf{x}, \boldsymbol{\xi}(\boldsymbol{\omega})) < +\infty) = 1$ , então  $Q(\mathbf{x}) < +\infty$ .
- (b) Vale a igualdade  $K_2^P = K_2$  e, em particular,  $K_2$  é fechado e convexo.
- (c) Se  $\mathbf{T}(\boldsymbol{\omega}) = \mathbf{T} = \text{constante}$ , então o conjunto  $K_2$  é um polítopo.
- (d) Se  $\mathbf{h}(\boldsymbol{\omega})$  e  $\mathbf{T}(\boldsymbol{\omega})$  são variáveis aleatórias independentes e se o suporte da distribuição de  $\mathbf{T}(\boldsymbol{\omega})$  é um polítopo, então  $K_2$  é um polítopo.

#### 4.4 Propriedades das funções de recurso

Nesta seção apresentaremos, sem demonstrações, as propriedades da função de segundo estágio  $Q(\mathbf{x}, \boldsymbol{\xi}) = v(\mathbf{h}(\boldsymbol{\omega}) - \mathbf{T}(\boldsymbol{\omega})\mathbf{x}, \boldsymbol{\omega})$  e da função de valor esperado  $Q(\mathbf{x}) = \mathbb{E}[v(\mathbf{h}(\boldsymbol{\omega}) - \mathbf{T}(\boldsymbol{\omega})\mathbf{x}, \boldsymbol{\omega})] = \mathbb{E}_{\boldsymbol{\xi}}[Q(\mathbf{x}, \boldsymbol{\xi})]$ . O leitor interessado pode consultar os livros [03, 13, 14].

**Teorema 4.3** Se  $\mathbf{W}(\boldsymbol{\omega}) = \mathbf{W}$  (recurso fixo), então  $Q(\mathbf{x}, \boldsymbol{\xi})$  é (a) convexa e linear por partes em  $(\mathbf{h}(\boldsymbol{\omega}), \mathbf{T}(\boldsymbol{\omega}))$  (componentes do vetor aleatório  $\boldsymbol{\xi}$ ), (b) côncava e linear por partes em  $\mathbf{q}(\boldsymbol{\omega})$  e (c) convexa e linear por partes para todo  $\mathbf{x} \in K = K_1 \cap K_2$ .

**Teorema 4.4** Se  $\mathbf{W}(\omega) = \mathbf{W}$  (recurso fixo) e  $\xi$  tem segundos momentos finitos, então (a)  $Q$  é uma função convexa, lipschitziana e finita em  $K_2$ , (b)  $Q$  é linear por partes se  $\Xi$  é finito e (c)  $Q$  é diferenciável em  $K_2$  se a função distribuição de  $\xi$  é absolutamente contínua.

## 4.5 Casos especiais: recurso completo e simples

Dizemos que o modelo de recurso em dois estágios (4.8) tem *recurso relativamente completo* se toda escolha de  $\mathbf{x}$  que satisfaz as restrições (rígidas) do primeiro estágio também satisfaz as condições de admissibilidade e finitude do segundo estágio, isto é, dizemos que (4.8) tem *recurso relativamente completo* se  $K_1 \subseteq K_2$ . Embora a hipótese de recurso relativamente completo seja muito desejável e útil do ponto de vista computacional, pode ser difícil identificar se um determinado problema tem ou não recurso relativamente completo, já que isto exigiria algum conhecimento dos conjuntos  $K_1$  e  $K_2$ .

Existe, contudo, um tipo particular de recurso relativamente completo que é fácil de se identificar a partir da matriz de tecnologia  $\mathbf{W}$  (determinística). Esta forma, denominada *recurso completo*, ocorre quando a matriz de tecnologia  $\mathbf{W}$  de (4.8) satisfaz a seguinte condição:

$$\text{para todo } \mathbf{z} \in \mathbb{R}^m, \text{ existe } \mathbf{y} \geq \mathbf{0} \text{ tal que } \mathbf{W}\mathbf{y} = \mathbf{z}. \quad (4.12)$$

Observe que se  $\mathbf{W}$  satisfaz esta condição, então  $Q(\mathbf{x}, \xi) < +\infty$  para qualquer realização de  $\xi$  em  $\Xi$ . Wets e Witzgall propuseram, em [18], um algoritmo para verificar se uma matriz  $\mathbf{W}$  satisfaz ou não (4.12).

Uma outra situação especial é a de *recurso simples*. Ela ocorre quando a matriz de tecnologia  $\mathbf{W}$  é da forma  $\begin{bmatrix} +I & -I \end{bmatrix}$ ,  $\mathbf{y}$  é dividido em  $(\bar{\mathbf{y}}, \underline{\mathbf{y}})$  e  $\mathbf{q}(\omega) = (\bar{\mathbf{q}}(\omega), \underline{\mathbf{q}}(\omega))$ . De fato, se (4.8) tem recurso simples, então vale o seguinte resultado (veja [03], página 92):

**Teorema 4.5** Se o modelo de recurso em dois estágios (4.8) tem recurso simples e se  $\xi$  tem segundos momentos finitos, então  $Q$  é finita se, e somente se,  $\bar{\mathbf{q}}(\omega) + \underline{\mathbf{q}}(\omega) \geq \mathbf{0}$  com probabilidade 1.

## 4.6 Mínimos e esperanças

Nesta seção veremos uma reformulação de (4.8) que será útil tanto do ponto de vista teórico como do ponto de vista computacional. Considere os dois problemas de otimização:

$$\begin{aligned} & \text{PL}_1 \\ & \parallel \\ & \min_{\mathbf{x} \in \mathbf{X}} \left\{ \mathbf{c}\mathbf{x} + \mathbb{E} \left[ \min_{\mathbf{y} \in \mathbf{Y}} \{ \mathbf{q}(\boldsymbol{\omega})\mathbf{y} \mid \mathbf{W}(\boldsymbol{\omega})\mathbf{y} \sim \mathbf{h}(\boldsymbol{\omega}) - \mathbf{T}(\boldsymbol{\omega})\mathbf{x} \} \right] \mid \mathbf{A}\mathbf{x} = \mathbf{b} \right\} \\ & \parallel \\ & \min_{\mathbf{x} \in \mathbf{X}} \left\{ \mathbb{E} \left[ \min_{\mathbf{y} \in \mathbf{Y}} \{ \mathbf{c}\mathbf{x} + \mathbf{q}(\boldsymbol{\omega})\mathbf{y} \mid \mathbf{W}(\boldsymbol{\omega})\mathbf{y} \sim \mathbf{h}(\boldsymbol{\omega}) - \mathbf{T}(\boldsymbol{\omega})\mathbf{x} \} \right] \mid \mathbf{A}\mathbf{x} = \mathbf{b} \right\} \end{aligned}$$

e

$$\begin{aligned} & \text{PL}_0 \\ & \parallel \\ & \min_{\mathbf{x} \in \mathbf{X}, \mathbf{y}(\cdot) \in \mathcal{Y}} \left\{ \mathbb{E} [\mathbf{c}\mathbf{x} + \mathbf{q}(\boldsymbol{\omega})\mathbf{y}(\boldsymbol{\omega})] \mid \begin{array}{l} \mathbf{A}\mathbf{x} = \mathbf{b} \\ \mathbf{W}(\boldsymbol{\omega})\mathbf{y}(\boldsymbol{\omega}) \sim \mathbf{h}(\boldsymbol{\omega}) - \mathbf{T}(\boldsymbol{\omega})\mathbf{x}, \forall \boldsymbol{\omega} \in \boldsymbol{\Omega} \end{array} \right\}, \end{aligned}$$

onde  $\mathcal{Y}$  é o conjunto das funções (mensuráveis)  $\mathbf{y}: \boldsymbol{\Omega} \rightarrow \mathbb{R}^p$ . Note que existirão infinitas restrições em  $\text{PL}_0$  se  $\boldsymbol{\Omega}$  for um conjunto infinito.

Apesar de “ $\min\{\cdot\}$ ” e “ $\mathbb{E}[\cdot]$ ” não comutarem (veja o exercício [05] deste capítulo), sob certas condições (por exemplo,  $\boldsymbol{\Omega}$  finito), vale que

$$\text{PL}_1 = \text{PL}_0.$$

A demonstração deste fato usa teoria da medida, o que foge do escopo deste texto. Ao leitor interessado, indicamos o livro [17], páginas 16 a 21.

## 4.7 Cotas para o valor ótimo

Se o agente de decisão pode esperar pelas realizações de  $\boldsymbol{\xi}$  para escolher o valor de  $\mathbf{x}$ , ele então irá resolver cada um dos problemas de otimização

$$\min_{\mathbf{x} \in \mathbf{X}} \left\{ f(\mathbf{x}, \boldsymbol{\xi}) = \mathbf{c}\mathbf{x} + \min_{\mathbf{y} \in \mathbf{Y}} \{ \mathbf{q}(\boldsymbol{\omega}) \mid \mathbf{W}(\boldsymbol{\omega}) = \mathbf{h}(\boldsymbol{\omega}) - \mathbf{T}(\boldsymbol{\omega})\mathbf{x} \} \mid \mathbf{A}\mathbf{x} = \mathbf{b} \right\} \quad (4.13)$$

um para cada realização possível de  $\xi$ . Se  $\tilde{\mathbf{x}}(\xi)$  representa a solução deste problema otimização (que, evidentemente, depende de  $\xi$ ) então, em média, ele ganhará

$$\text{WS} = \mathbb{E}_\xi [f(\tilde{\mathbf{x}}(\xi), \xi)] = \mathbb{E}_\xi \left[ \min_{\mathbf{x} \in \mathbf{X}} \{f(\mathbf{x}, \xi) \mid \mathbf{Ax} = \mathbf{b}\} \right] \quad (4.14)$$

Este é o valor ótimo para a abordagem “espere e veja” (WS = *wait and see*). Por outro lado, se o agente de decisão deve fazer a escolha de  $\mathbf{x}$  antes das realizações de  $\xi$ , então ele ganhará

$$\text{RP} = \min_{\mathbf{x} \in \mathbf{X}} \{\mathbb{E}_\xi [f(\mathbf{x}, \xi)] \mid \mathbf{Ax} = \mathbf{b}\}. \quad (4.15)$$

resolvendo um modelo de recurso em dois estágios (RP = *recourse problem*). Seja  $\mathbf{x}^*$  representa a solução ótima deste problema, então  $\text{RP} = \mathbb{E}_\xi [f(\mathbf{x}^*, \xi)]$ . O *valor esperado de informação perfeita* (EVPI = *expected value of perfect information*) é, por definição, a diferença entre os valores ótimos para as abordagens “espere e veja” e “modelo de recurso”:

$$\text{EVPI} = \text{RP} - \text{WS}. \quad (4.16)$$

No problema do fazendeiro, o valor ótimo para a abordagem “espere e veja” foi igual a  $\text{WS} = -\text{R\$ } 115\,406,00$  (quando convertido para um problema de minimização), enquanto que para a abordagem de “modelo de recurso”, o valor ótimo foi igual a  $\text{RP} = -\text{R\$ } 108\,390,00$ . O valor esperado de informação perfeita do fazendeiro foi, então, igual a  $\text{RP} - \text{WS} = \text{R\$ } 7\,016,00$ . *Esta é a quantidade de dinheiro que o fazendeiro deveria pagar em cada ano para obter uma informação perfeita sobre o clima no próximo ano. Um bom meteorologista poderia cobrar este valor do fazendeiro para assessorá-lo em questões climáticas.*

Pela definição de  $\tilde{\mathbf{x}}(\xi)$ , segue-se que  $f(\tilde{\mathbf{x}}(\xi), \xi) \leq f(\mathbf{x}, \xi)$  para todo  $\xi \in \Xi$  e para todo  $\mathbf{x} \in \mathbf{X}$  satisfazendo  $\mathbf{Ax} = \mathbf{b}$ . Em particular,

$$f(\tilde{\mathbf{x}}(\xi), \xi) \leq f(\mathbf{x}^*, \xi)$$

para todo  $\xi \in \Xi$ . Tomando-se a esperança dos dois lados, concluímos que

$$\text{WS} = \mathbb{E}_\xi [f(\tilde{\mathbf{x}}(\xi), \xi)] \leq \mathbb{E}_\xi [f(\mathbf{x}^*, \xi)] = \text{RP}.$$

Isto mostra que WS dá uma cota inferior para o valor ótimo de RP (que é o problema (4.8)). Como corolário, obtemos também que  $\text{EVPI} \geq 0$ .



Para usar a abordagem “espere e veja”, o agente de decisão deve ser capaz de (1) resolver cada um dos problemas (4.13) e (2) calcular a média dos valores ótimos correspondentes. Como o problema da mistura mostrou, estas tarefas podem ser muito trabalhosas! Uma alternativa tentadora é a de substituir todas as variáveis aleatórias pelas suas médias e, então, resolver o problema correspondente. Este problema é denominado *problema do valor esperado* ou *problema do valor médio*:

$$EV = \min_{\mathbf{x} \in \mathbf{X}} \left\{ f(\mathbf{x}, \hat{\boldsymbol{\xi}}) \mid \mathbf{A}\mathbf{x} = \mathbf{b} \right\} \quad (4.17)$$

onde  $\hat{\boldsymbol{\xi}} = \mathbb{E}_{\boldsymbol{\xi}}[\boldsymbol{\xi}]$  representa a média de  $\boldsymbol{\xi}$ . Vamos denotar por  $\hat{\mathbf{x}}(\hat{\boldsymbol{\xi}})$  a solução ótima de (4.17), denominada *solução do valor esperado*. Em princípio, não existe nenhum motivo para esperar que  $\hat{\mathbf{x}}(\hat{\boldsymbol{\xi}})$  esteja, de alguma maneira, próxima da solução  $\mathbf{x}^*$  do modelo de recurso (4.15). O *valor da solução estocástica* (VSS = *value of stochastic solution*) (apresentado no problema do fazendeiro) é o conceito que justamente mede o quão bom (ou o quão ruim) é a decisão  $\mathbf{x} = \hat{\mathbf{x}}(\hat{\boldsymbol{\xi}})$  em termos de (4.15). Vamos primeiro definir o *resultado esperado no uso da solução do valor esperado* (EEV = *Expected result of using the EV solution*):

$$EEV = \mathbb{E}_{\boldsymbol{\xi}} \left[ f \left( \hat{\mathbf{x}}(\hat{\boldsymbol{\xi}}), \boldsymbol{\xi} \right) \right]. \quad (4.18)$$

EEV mede a eficácia da decisão  $\hat{\mathbf{x}}(\hat{\boldsymbol{\xi}})$ , permitindo-se que decisões do segundo estágio sejam escolhidas de maneira ótima como funções de  $\hat{\mathbf{x}}(\hat{\boldsymbol{\xi}})$  e  $\boldsymbol{\xi}$ . O valor da solução estocástica é, então, definida como

$$VSS = EEV - RP. \quad (4.19)$$

O valor de VSS dá o custo que se paga quando incertezas são ignoradas no processo de decisão. No problema do fazendeiro, vimos que  $EEV = -R\$ 107 240,00$  e  $RP = -R\$ 108 390,00$ , de modo que  $VSS = -R\$ 1 150,00$ .

Pela definição de RP, segue-se que  $\mathbb{E}_{\boldsymbol{\xi}}[f(\mathbf{x}^*, \boldsymbol{\xi})] \leq \mathbb{E}_{\boldsymbol{\xi}}[f(\mathbf{x}, \boldsymbol{\xi})]$  para qualquer  $\mathbf{x} \in \mathbf{X}$  satisfazendo  $\mathbf{A}\mathbf{x} = \mathbf{b}$ . Em particular,

$$RP = \mathbb{E}_{\boldsymbol{\xi}}[f(\mathbf{x}^*, \boldsymbol{\xi})] \leq \mathbb{E}_{\boldsymbol{\xi}} \left[ f(\hat{\mathbf{x}}(\hat{\boldsymbol{\xi}}), \boldsymbol{\xi}) \right] = EEV.$$

Isto mostra que EEV dá uma cota superior para o valor ótimo de RP (que é o problema (4.8)). Como corolário, obtemos também que  $VSS \geq 0$ .

### 4.8 O caso $\Omega$ finito

Se  $\Omega$  é finito, digamos  $\Omega = \{\omega^1, \dots, \omega^S\} \subseteq \mathbb{R}^r$ , o modelo de recurso (4.8), via a equivalência apresentada na seção 4.6, produz o seguinte programa linear (basta abrir a esperança em  $PL_0!$ ) denominada *forma estendida* de

minimizar $\mathbf{c}\mathbf{x} + p_1\mathbf{q}^1\mathbf{y}^1 + p_2\mathbf{q}^2\mathbf{y}^2 + \dots + p_S\mathbf{q}^S\mathbf{y}^S$ $\mathbf{x} \in \mathbf{X}$ $\mathbf{y}^1, \mathbf{y}^2, \dots, \mathbf{y}^S \in \mathbf{Y}$	$\mathbf{A}\mathbf{x}$ $\mathbf{T}^1\mathbf{x} + \mathbf{W}\mathbf{y}^1$ $\mathbf{T}^1\mathbf{x} + \mathbf{W}\mathbf{y}^2$ $\vdots$ $\mathbf{T}^S\mathbf{x} + \mathbf{W}\mathbf{y}^S$	$= \mathbf{b},$ $\sim \mathbf{h}^1,$ $\sim \mathbf{h}^2,$ $\vdots$ $\sim \mathbf{h}^S,$	$(4.20)$
---	---	---	----------

onde  $p_s = \mathbb{P}(\omega = \omega^s)$ ,  $\mathbf{q}^s = \mathbf{q}(\omega^s)$ ,  $\mathbf{y}^s = \mathbf{y}(\omega^s)$ ,  $\mathbf{T}^s = \mathbf{T}(\omega^s)$  e  $\mathbf{h}^s = \mathbf{h}(\omega^s)$ . Aqui estamos supondo que o recurso é fixo ( $\mathbf{W}(\omega) = \mathbf{W}$ ).

A vantagem deste modelo é que ele é um programa linear. A única desvantagem é o seu tamanho: ele possui  $n + pS$  variáveis e  $m_1 + mS$  restrições explícitas (isto é, não contando as possíveis restrições na definição de  $\mathbf{X}$  e  $\mathbf{Y}$ ).

### Exercícios

- [01] Mostre que para os valores de  $p$ ,  $\mathbf{q}$ ,  $\mathbf{W}$  e  $\mathbf{Y}$  indicados em cada um dos três casos na página 36, a função de recuso  $v$  em (4.6) obtém as funções de penalidade (4.3), (4.4) e (4.5) como casos especiais.
- [02] Considere um exemplo de (4.8) onde  $Q(x, \xi) = \min_{y \geq 0} \{y \mid \xi y = 1 - x\}$ . Mostre que se  $\xi$  tem distribuição triangular em  $[0, 1]$  ( $\mathbb{P}(\xi \leq u) = u^2$ ), então  $K_2^P \neq K_2$ .
- [03] Mostre que a matriz  $\mathbf{W} = [ +I \quad -I ]$  satisfaz a condição (4.12).
- [04] Mostre que a matriz  $\mathbf{W} = [ \mathbb{1}^T \quad -I ]$  satisfaz a condição (4.12). Aqui  $\mathbb{1}$  representa o vetor com todas as componentes iguais a 1.
- [05] Considere  $f(x, \omega) = (x - \omega)^2$ , onde  $x \in \mathbb{R}$  e  $\omega$  é uma variável aleatória com distribuição uniforme no intervalo  $[0, 1]$ . Mostre que, para esta

situação, “ $\min\{\cdot\}$ ” e “ $\mathbb{E}[\cdot]$ ” não comutam:

$$\mathbb{E} \left[ \min_{x \in \mathbb{R}} \{f(x, \omega)\} \right] \neq \min_{x \in \mathbb{R}} \left\{ \mathbb{E} [f(x, \omega)] \right\}.$$

# Capítulo 5

## O método L-shaped

O L-shaped é possivelmente o método de resolução e aproximação de problemas de otimização estocástica mais conhecido e tradicional. Ele se originou do método de *decomposição de Benders*, desenvolvido na década de sessenta por J. F. Benders para resolver de maneira mais eficiente problemas de otimização com uma determinada estrutura. Na realidade, o L-shaped pode ser visto como uma aplicação do método de Benders em otimização estocástica. Esse é o ponto de vista que será adotado nesse texto.

### 5.1 A decomposição de Benders

Considere o problema de otimização linear abaixo

$$\begin{array}{l} \text{VAL} = \min_{\mathbf{x}, \mathbf{y}} \quad \mathbf{c}^T \mathbf{x} + \mathbf{q}^T \mathbf{y} \\ \text{sujeito a} \quad \mathbf{Ax} = \mathbf{b}, \\ \quad \quad \quad \mathbf{Tx} + \mathbf{Wy} = \mathbf{h} \\ \quad \quad \quad \mathbf{x}, \mathbf{y} \geq \mathbf{0}, \end{array} \quad (5.1)$$

onde  $\mathbf{c}$ ,  $\mathbf{q}$ ,  $\mathbf{x}$  e  $\mathbf{y}$  são vetores em  $\mathbb{R}^n$ ,  $\mathbf{h}$  e  $\mathbf{b}$  são vetores de  $\mathbb{R}^m$  e  $\mathbf{A}$ ,  $\mathbf{T}$ ,  $\mathbf{W}$  são matrizes  $m \times n$ . O primeiro passo é eliminar a variável  $\mathbf{y}$  da formulação do problema, criando um subproblema parametrizado por  $\mathbf{x}$ . Para isso, a decomposição de Benders reescreve o problema da seguinte forma:

$$\begin{array}{l} \text{VAL} = \min_{\mathbf{x}} \quad \mathbf{c}^T \mathbf{x} + Q(\mathbf{x}) \\ \text{sujeito a} \quad \mathbf{Ax} = \mathbf{b}, \\ \quad \quad \quad \mathbf{x} \geq \mathbf{0}, \end{array} \quad (5.2)$$

onde  $Q(\mathbf{x})$  é o valor ótimo do subproblema

$$\boxed{\begin{array}{l} Q(\mathbf{x}) = \min_{\mathbf{y}} \quad \mathbf{q}^T \mathbf{y} \\ \text{sujeito a } \quad \mathbf{W}\mathbf{y} = \mathbf{h} - \mathbf{T}\mathbf{x}, \\ \quad \quad \quad \mathbf{y} \geq \mathbf{0}. \end{array}} \quad (\text{P})$$

Dualizando (P), temos o problema (D) definido a seguir:

$$\boxed{\begin{array}{l} Q(x) = \max_{\mathbf{p}} \quad \mathbf{p}^T (\mathbf{h} - \mathbf{T}\mathbf{x}) \\ \text{sujeito a } \quad \mathbf{W}^T \mathbf{p} \leq \mathbf{q}. \end{array}} \quad (\text{D})$$

Repare que a dualidade tirou a dependência do conjunto admissível  $\mathcal{D}$  de (D) em relação a  $\mathbf{x}$ , ou seja, para qualquer escolha de  $\mathbf{x}$  o conjunto admissível de (D) é o mesmo:

$$\mathcal{D} = \{\mathbf{p} \in \mathbb{R}^m \mid \mathbf{W}^T \mathbf{p} \leq \mathbf{q}\}.$$

Vamos assumir que o conjunto  $\mathcal{D}$  é não vazio e denotar seus pontos extremos por  $\mathbf{p}^1, \dots, \mathbf{p}^I$  e seus raios extremos por  $\mathbf{r}^1, \dots, \mathbf{r}^J$ . O problema (D) dá origem a duas situações distintas:

- Se  $Q(\mathbf{x}) = +\infty$ , então o simplex devolve o raio extremo  $\bar{\mathbf{r}}$  de  $\mathcal{D}$ . Em particular  $(\bar{\mathbf{r}})^T (\mathbf{h} - \mathbf{T}\mathbf{x}) > 0$ .
- Se  $Q(\mathbf{x}) < +\infty$ , então o simplex devolve um ponto extremo  $\bar{\mathbf{p}}$  de  $\mathcal{D}$ .

Usando o teorema fundamental da programação linear (D.1), podemos reescrever o problema (D) como

$$\boxed{\begin{array}{l} Q(\mathbf{x}) = \min_z \\ \text{sujeito a } \quad (\bar{\mathbf{p}}^i)^T (\mathbf{h} - \mathbf{T}\mathbf{x}) \leq z, \quad i = 1, \dots, I, \\ \quad \quad \quad (\bar{\mathbf{r}}^j)^T (\mathbf{h} - \mathbf{T}\mathbf{x}) \leq 0, \quad j = 1, \dots, J. \end{array}} \quad (\text{D}')$$

Não é difícil ver que esse problema possui o mesmo valor ótimo de (D). As primeiras  $I$  restrições representam os valores da função objetivo de (D) avaliada nos pontos extremos de  $\mathcal{D}$ . Como queremos minimizar  $z$ , o valor ótimo de (D') corresponde ao maior valor de  $(\bar{\mathbf{p}}^i)^T (\mathbf{h} - \mathbf{T}\mathbf{x})$ , que é exatamente o ótimo de (D). As  $J$  restrições restantes servem para garantir que estamos na situação onde (D) é limitado.

Finalmente, usando (D'), chegamos em uma reformulação do problema original (5.1), conhecida por *problema mestre completo* (PMC), que é a base para a decomposição de Benders:

$$\begin{array}{r}
 \text{VAL} = \min_{\mathbf{x}, z} \quad \mathbf{c}^T \mathbf{x} + z \\
 \text{sujeito a} \quad \mathbf{Ax} = \mathbf{b}, \quad \mathbf{x} \geq \mathbf{0}, \\
 \quad (\mathbf{p}^i)^T (\mathbf{h} - \mathbf{T}\mathbf{x}) \leq z, \quad i = 1, \dots, I \quad (*), \\
 \quad (\mathbf{r}^j)^T (\mathbf{h} - \mathbf{T}\mathbf{x}) \leq 0, \quad j = 1, \dots, J \quad (*).
 \end{array} \tag{5.3}$$

Essa formulação, apesar de equivalente a (5.1), possui várias diferenças que serão exploradas pela decomposição de Benders. Primeiramente, a variável  $\mathbf{y}$  não aparece no PMC. Em seu lugar surge a variável unidimensional  $z$ . Além disso, nessa formulação é preciso conhecer os pontos e raios extremos do conjunto  $\mathcal{D}$ . Possivelmente o número de restrições de (5.3) é gigantesco se comparado a (5.1): se o conjunto admissível  $\mathcal{D}$  for muito facetado, então teremos uma quantidade muito grande de pontos extremos. Aliás, em geral, não temos nem os pontos nem os raios extremos do conjunto  $\mathcal{D}$  de imediato. A idéia do algoritmo da decomposição de Benders é considerar um problema semelhante a (5.3) nas etapas intermediárias do algoritmo e ir acrescentando restrições do tipo (\*) ao problema em cada passo. Mais precisamente, no passo  $k$  do algoritmo, o problema mestre restrito de ordem  $k$  ( $\text{PMR}^k$ ) com apenas  $k$  restrições do tipo (\*) é

$$\begin{array}{r}
 \text{VAL}^k = \min_{\mathbf{x}, z} \quad \mathbf{c}^T \mathbf{x} + z \\
 \text{sujeito a} \quad \mathbf{Ax} = \mathbf{b}, \quad \mathbf{x} \geq \mathbf{0}, \\
 \quad (\mathbf{p}^i)^T (\mathbf{h} - \mathbf{T}\mathbf{x}) \leq z, \quad i = 1, \dots, k - l, \\
 \quad (\mathbf{r}^j)^T (\mathbf{h} - \mathbf{T}\mathbf{x}) \leq 0, \quad j = 1, \dots, l.
 \end{array} \tag{5.4}$$

Para obter pontos e raios extremos de  $\mathcal{D}$  temos que resolver o problema (D) em cada passo do algoritmo. Se ele for finito, então ganhamos um ponto extremo e, se for ilimitado, ganhamos um raio extremo. Eles serão adicionado ao problema  $\text{PMR}^k$ , originando o problema  $\text{PMR}^{k+1}$ .

## 5.2 O algoritmo de Benders

Vamos apresentar o algoritmo de Benders passo a passo. Como a finalidade de um algoritmo é geralmente sua implementação em computador, incluiremos na sua descrição tanto a condição de parada teórica quanto a computacional. A teórica simplesmente diz que se uma certa condição for atingida, então a solução encontrada é ótima. A computacional pode parar antes da solução ótima ser encontrada: basta que a diferença entre uma cota superior e uma cota inferior sejam menores que uma dada tolerância  $\epsilon$ .

PASSO 1: Defina a cota inferior  $CI = +\infty$  e a cota superior  $CS = -\infty$ .

PASSO 2: Resolva o problema  $PMR^k$  (5.4). Seja  $VAL^k$  o valor ótimo e  $(\bar{\mathbf{x}}, \bar{z})$  a solução encontrada. Atualize  $CI = VAL^k$ .

PASSO 3: (a) Se  $Q(\bar{\mathbf{x}}) < +\infty$ , então resolva também (P) para  $\mathbf{x} = \bar{\mathbf{x}}$  e guarde as soluções  $(\bar{\mathbf{y}}, \bar{\mathbf{p}})$  de (P) e (D), respectivamente.

(b) Se  $Q(\bar{\mathbf{x}}) > \bar{z}$ , então atualize a cota superior

$$CS = \min\{CS, \mathbf{c}^T \bar{\mathbf{x}} + \mathbf{q}^T \bar{\mathbf{y}}\}.$$

Além disso, faça

$$\text{MELHOR SOLUÇÃO} = \{(\bar{\mathbf{x}}, \bar{\mathbf{y}})\}.$$

Por fim, se  $CS - CI < \epsilon$ , pare. Caso contrário acrescente a restrição  $\bar{\mathbf{p}}^T(\mathbf{h} - \mathbf{T}\mathbf{x})$  ao problema  $PMR^k$  para obter  $PMR^{k+1}$  e volte ao PASSO 1.

(c) Se  $Q(\bar{\mathbf{x}}) = \bar{z}$ , então  $(\bar{\mathbf{x}}, \bar{z})$  é a solução ótima do problema original (5.1).

PASSO 4: Se  $Q(\mathbf{x}) = +\infty$ , seja  $\bar{\mathbf{r}}$  o raio extremo gerado pelo resolvidor. Acrescente a restrição  $(\bar{\mathbf{r}})^T(\mathbf{h} - \mathbf{T}\mathbf{x}) \leq 0$  a  $PMR^k$  para obter  $PMR^{k+1}$ . Volte ao PASSO 1.

Vamos analisar em detalhe alguns pontos do algoritmo. No PASSO 1, afirmamos que  $VAL^k$  é uma cota inferior para o problema original (5.1). De

fato, se pensarmos na formulação (5.3), vemos que o conjunto admissível está contido no conjunto admissível de  $\text{PMR}^k$  e, portanto,  $\text{VAL}^k$  é menor ou igual ao valor ótimo de (5.1). Dessa forma, a cota inferior vai sempre aumentando. A cota superior não é tão bem comportada. O valor  $\mathbf{c}^T \bar{\mathbf{x}} + \mathbf{q}^T \bar{\mathbf{y}}$  é uma cota superior para (5.1) simplesmente por ser admissível. No entanto, novos pontos admissíveis podem aumentar a cota superior CS, um efeito indesejado. É por isso que a atualização de CS no PASSO 3 (b) é um mínimo entre a cota antiga e o novo candidato gerado.

Para ver que a condição  $\mathcal{Q}(\bar{\mathbf{x}}) = \bar{z}$  implica em otimalidade para PMC, note que  $\mathcal{Q}(\bar{\mathbf{x}}) = \bar{z}$  implica que  $\bar{z}$  é maior ou igual a  $(\bar{\mathbf{p}}^i)^T (\mathbf{h} - T\bar{\mathbf{x}})$ , para  $i$  de 1 até  $I$ . Então a solução  $(\bar{\mathbf{x}}, \bar{z})$  também é admissível para PMC e, portanto, ela é ótima. Mas o problema original é nas variáveis  $\mathbf{x}$  e  $\mathbf{y}$ . Será que nesse caso  $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$  é solução de (5.1)? De fato,

$$\text{VAL}^k = \mathbf{c}^T \bar{\mathbf{x}} + \bar{z} = \mathbf{c}^T \bar{\mathbf{x}} + \mathcal{Q}(\bar{\mathbf{x}}) = \mathbf{c}^T \bar{\mathbf{x}} + \mathbf{q}^T \bar{\mathbf{y}} \geq \text{VAL},$$

onde a última desigualdade segue pois  $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$  é admissível. Por outro lado, o conjunto admissível de  $\text{PMR}^k$  contém o de PMC, que é uma reescrita do problema original (5.1). Assim, é trivial que  $\text{VAL}^k \leq \text{VAL}$ , completando a prova.

### 5.3 Um exemplo completo

Considere o problema

$$\boxed{\begin{array}{ll} \text{VAL} = & \min_{\mathbf{x}, \mathbf{y}} \quad 42x_1 + 18x_2 + 33x_3 - 8y_1 - 6y_2 + 2y_3 \\ & \text{sujeito a} \quad 10x_1 + 8x_2 \quad - 2y_1 - y_2 + y_3 \geq 4, \\ & \quad \quad \quad 5x_1 + \quad \quad 8x_3 - y_1 - y_2 - y_3 \geq 3, \\ & \quad \quad \quad x_i \in \{0, 1\}, y_i \geq 0, i = 1, 2, 3. \end{array}} \quad (5.5)$$

Para ver que esse problema é da forma (5.1) basta considerar as matrizes

$$\begin{aligned} \mathbf{x} &= [x_1 \ x_2 \ x_3]^T, \quad \mathbf{y} = [y_1 \ y_2 \ y_3]^T, \\ \mathbf{c} &= [42 \ 18 \ 33]^T, \quad \mathbf{q} = [-8 \ -6 \ 2]^T, \quad \mathbf{h} = [4 \ 3]^T, \\ \mathbf{T} &= \begin{bmatrix} 10 & 8 & 0 \\ 5 & 0 & 8 \end{bmatrix} \quad \text{e} \quad \mathbf{W} = \begin{bmatrix} -2 & -1 & 1 \\ -1 & -1 & -1 \end{bmatrix}, \end{aligned}$$



com  $\mathbf{A} = \mathbf{0}$ . Esse problema é linear, mas as variáveis  $x_1, x_2$  e  $x_3$  só podem assumir os valores 0 e 1. Como a decomposição de Benders fixa os valores de  $\mathbf{x}$  e resolve o problema em  $\mathbf{y}$ , ela também se aplica nesse caso. Vamos chamar o conjunto de pontos extremos obtidos até a iteração  $k$  de  $\mathcal{P}_k$  e os raios de  $\mathcal{R}_k$ . Os arquivos em AMPL necessários para obter as soluções dos problemas intermediários estão em <http://www.mat.puc-rio.br/~bernardo/>.

ITERAÇÃO 0  $CS = +\infty$ ,  $CI = -\infty$ ,  $\mathcal{P}_0 = \emptyset$ ,  $\mathcal{R}_0 = \emptyset$ .

Resolvendo o problema  $\text{PMR}^0$ , vemos que ele claramente é não limitado (arquivo `benders1.mod`), pois não existem restrições sobre  $z$ . Arbitramos  $\bar{\mathbf{x}}_0 = (0, 0, 0)$  e atualizamos a cota inferior:  $CI = 0$ . Resolvendo o problema (D) (arquivo `bendersdu0.mod`) obtemos  $\mathcal{Q}(\bar{\mathbf{x}}_0) = \infty$  e caímos no PASSO 4 do algoritmo. O resolvidor CPLEX nos devolve o raio extremo  $r_0 = (0, 1)$ , que será acrescentado a  $\text{PMR}^1$ .

ITERAÇÃO 1  $CS = +\infty$ ,  $CI = -\infty$ ,  $\mathcal{P}_1 = \emptyset$ ,  $\mathcal{R}_1 = \{(0, 1)\}$ .

Novamente não temos restrições para a variável  $z$  e, portanto, o problema é ilimitado. Escolhemos  $\bar{\mathbf{x}}_1 = (0, 0, 1)$ , pois ele é admissível para  $\text{PMR}^1$ . Resolvendo o problema (D) (arquivo `bendersdu1.mod`) obtemos  $\mathcal{Q}(\bar{\mathbf{x}}_1) = 6$  e  $\bar{\mathbf{p}}_1 = (4, 2)$ . Como  $\mathcal{Q}(\bar{\mathbf{x}}_1) = 6 > -\infty = \bar{z}_1$ , resolvemos (arquivo `benderspr1.mod`) o problema (P) e encontramos  $\bar{\mathbf{y}}_1 = (0, 0.5, 4.5)$ . Assim,  $CS = \min\{+\infty, \mathbf{c}^T \bar{\mathbf{x}}_1 + \mathbf{q}^T \bar{\mathbf{y}}_1\} = 39$ . Por fim acrescentamos o ponto extremo  $(4, 2)$  ao problema  $\text{PMR}^1$  para obter  $\text{PMR}^2$ .

ITERAÇÃO 2  $CS = 39$ ,  $CI = -\infty$ ,  $\mathcal{P}_2 = \{(4, 2)\}$ ,  $\mathcal{R}_2 = \{(0, 1)\}$ .

Resolvendo (arquivo `benders2.mod`)  $\text{PMR}^2$  encontramos que  $\text{VAL}^2 = 0$ ,  $\bar{\mathbf{x}}_2 = (1, 1, 0)$  e  $\bar{z}_2 = -60$ . Após atualizar  $CI = -60$ , resolve (arquivo `bendersdu2.mod`) (D) para  $\mathbf{x} = \bar{\mathbf{x}}_2$ . A solução encontrada foi  $\mathcal{Q}(\bar{\mathbf{x}}_2) = -16$  e  $\bar{\mathbf{p}}_2 = (0, 8)$ . Como  $-16 > -60$ , temos que resolver (arquivo `benderspr2.mod`) o problema (P). A solução é  $\bar{\mathbf{y}}_2 = (2, 0, 0)$ . A cota superior fica inalterada nesse caso pois o valor de  $\mathbf{c}^T \bar{\mathbf{y}}_2 + \mathbf{q}^T \bar{\mathbf{y}}_2$  é maior que 39. Acrescentando-se o ponto extremo  $\bar{\mathbf{p}}_2$  a  $\text{PMR}^2$  obtemos  $\text{PMR}^3$ .

ITERAÇÃO 3  $CS = 39$ ,  $CI = -60$ ,  $\mathcal{P}_3 = \{(4, 2), (0, 8)\}$ ,  $\mathcal{R}_3 = \{(0, 1)\}$ .

Resolvendo (arquivo benders3.mod)  $PMR^3$  encontramos que  $VAL^3 = 17$ ,  $\bar{\mathbf{x}}_3 = (1, 1, 1)$  e  $\bar{z}_3 = -76$ . Atualize  $CI = 17$  e resolva (bendersdu3.mod) (D) para  $\mathbf{x} = \bar{\mathbf{x}}_3$ . O resultado é  $\mathcal{Q}(\bar{\mathbf{x}})_3 = -68$ , com  $\bar{\mathbf{p}}_3 = (2, 4)$ . Como  $-68 > -76$ , temos que resolver (arquivo benderspr3.mod) o problema P. A solução é  $\bar{\mathbf{y}}_3 = (4, 6, 0)$ . A nova cota superior é  $CS = 25$ . Acrescentando-se o ponto extremo  $\bar{\mathbf{p}}_3$  a  $PMR^3$  obtemos  $PMR^4$ .

ITERAÇÃO 4  $CS = 25$ ,  $CI = 17$ ,  $\mathcal{P}_4 = \{(4, 2), (0, 8), (2, 4)\}$ ,  $\mathcal{R}_4 = \{(0, 1)\}$ .

Resolvendo (arquivo benders4.mod)  $PMR^4$  encontramos que  $VAL^4 = 17$ ,  $\bar{\mathbf{x}}_4 = (1, 1, 1)$  e  $\bar{z}_4 = -68$ . Atualize  $CI = 25$  e note que como  $\bar{\mathbf{x}}_3 = \bar{\mathbf{x}}_4$ , então  $\mathcal{Q}(\bar{\mathbf{x}}_4) = -68$ . Finalmente, temos  $\mathcal{Q}(\bar{\mathbf{x}}_4) = \bar{z}_4 = -68$  e, portanto, a condição de parada foi satisfeita! A solução ótima é  $\bar{\mathbf{x}}^* = (1, 1, 1)$  e  $\bar{\mathbf{y}}^* = (4, 6, 0)$ , com  $VAL^* = 25$ .

## 5.4 Decomposição de Benders em otimização estocástica: o método L-shaped

Considere um problema de otimização estocástica na forma estendida, como em (4.20). Repare que o problema (5.1) tem o mesmo formato de (4.20), porém com menos restrições. Vamos apresentar duas versões do método L-shaped: o clássico e o multicortes. O método L-shaped multicortes consiste em usar decomposição de Benders na forma estendida de um problema estocástico.

O primeiro passo é reescrever o problema (4.8), de maneira análoga ao que foi feito para (5.1). Denotando por  $\Omega = \{\omega^1, \omega^2, \dots, \omega^S\}$  o espaço amostral da variável aleatória subjacente a (4.20) e definindo  $p_i = \mathbb{P}(\omega^i)$ , temos

$$\boxed{\begin{array}{ll} VAL = & \min_{\mathbf{x}} \quad \mathbf{c}^T \mathbf{x} + \sum_{s=1}^S p_s Q^s(\mathbf{x}, \omega^s) \\ & \text{sujeito a} \quad \mathbf{Ax} = \mathbf{b}, \\ & \quad \quad \quad \mathbf{x} \geq \mathbf{0}, \end{array}} \quad (5.6)$$

onde

$$Q^s(\mathbf{x}, \omega^s) = \min_{\mathbf{y}} \mathbf{q}^T \mathbf{y}^s$$

$$\text{sujeito a } \mathbf{W}\mathbf{y}^s = \mathbf{h}^s - \mathbf{T}^s \mathbf{x},$$

$$\mathbf{y}^s \geq \mathbf{0}.$$
(5.7)

para cada  $i$  de 1 a  $S$ .

Repare que a estrutura da formulação estendida (4.20) naturalmente desacopla os  $s$  problemas (5.7). Dualizando cada um deles, tem-se

$$Q^s(\mathbf{x}, \omega^s) = \max_{\mathbf{p}} \mathbf{p}^T (\mathbf{h}^s - \mathbf{T}^s \mathbf{x})$$

$$\text{sujeito a } \mathbf{W}\mathbf{p} \leq \mathbf{q}^s.$$
(5.8)

Agora vamos reescrever o problema (5.8) da mesma forma que foi feito para (D')

$$Q^s(\mathbf{x}, \omega^s) = \min_{z_s} z_s$$

$$\text{sujeito a } (\mathbf{p}_s^{i(s)})^T (\mathbf{h}^s - \mathbf{T}\mathbf{x}) \leq z_s, \quad i(s) = 1, \dots, I_s,$$

$$(\mathbf{r}_s^{j(s)})^T (\mathbf{h}^s - \mathbf{T}\mathbf{x}) \leq 0, \quad j(s) = 1, \dots, J_s,$$
(5.9)

onde  $I^s$  é o número de pontos extremos do conjunto  $\mathcal{D}^s = \{\mathbf{p} \mid \mathbf{W}\mathbf{p} \leq \mathbf{q}^s\}$ ,  $J^s$  é o número de raios extremos de  $\mathcal{D}^s$ ,  $\mathbf{p}_s^{i(s)}$  são os pontos extremos de  $\mathcal{D}$  e  $\mathbf{r}_s^{j(s)}$  são os raios extremos, para cada  $s = 1, \dots, S$ . Portanto, podemos reescrever o problema na forma estendida como

$$\text{VAL} = \min_{\mathbf{x}, z} \mathbf{c}\mathbf{x} + \sum_{s=1}^S p_s z_s$$

$$\text{sujeito a } (\mathbf{p}_s^{i(s)})^T (\mathbf{h}^s - \mathbf{T}\mathbf{x}) \leq z_s, \quad i(s) = 1, \dots, I_s, \quad s = 1, \dots, S,$$

$$(\mathbf{r}_s^{j(s)})^T (\mathbf{h}^s - \mathbf{T}\mathbf{x}) \leq 0, \quad j(s) = 1, \dots, J_s, \quad s = 1, \dots, S,$$
(5.10)

que é também chamado de PMC. A partir daí o algoritmo L-shaped multicortes é análogo a Benders. Considera-se problema mestre restrito PMR<sup>k</sup> e, em cada passo do algoritmo, vão se adicionando cortes da mesma forma que foi feita na seção 5.2. A única diferença é que no caso do L-shaped serão acrescentados até  $S$  cortes em cada passo, oriundos de cada um dos  $S$  problemas (5.8). Além disso, a condição de otimalidade tem que se verificar para todo  $s$ , ou seja,  $\mathcal{Q}(\bar{\mathbf{x}})^s = \bar{z}^s$ . para  $s$  de 1 a  $S$ .

O método L-shaped clássico pode ser pensado como uma variante do L-shaped multicorte. Por simplicidade, vamos assumir que os  $s$  problemas (5.8) possuem solução ótima. Ao invés de se construir  $S$  cortes a partir dessas soluções, constrói-se apenas um corte, que é uma espécie de média dos cortes obtidos:

$$\sum_{s=1}^S p_s (\mathbf{p}_s^{i(s)})^T (\mathbf{h}_s - T_s \mathbf{x}) \leq z, \quad i(s) \in \{1, \dots, I_s\}, \quad s = 1, \dots, S. \quad (5.11)$$

Naturalmente, temos apenas uma variável  $z$  nesse caso, pois temos apenas um corte.

## Exercícios

- [01] Considere o problema (4.8) com  $c = 0$ ,  $X = [0, 10]$ ,  $(q^+, q^-) = (1, 1)$  e  $\xi$  discreta assumindo os valores 1, 2 e 4 com probabilidade  $1/3$ . Resolva o problema usando o método L-shaped, clássico ou multicorte.
- [02] Considere o problema (4.8) com  $c = 1/2$ ,  $X = [0, 5]$ ,  $(q^+, q^-) = (1, 3)$  e  $\xi$  discreta uniforme assumindo os valores 1, 2, 3 e 4. Resolva o problema usando o método L-shaped (clássico ou multicorte).

# Capítulo 6

## Métodos Amostrais

Nessa seção vamos estudar métodos amostrais. Esses métodos diferem dos anteriores por uma razão fundamental: eles supõem que podemos sortear em computador números aleatórios com respeito a uma dada distribuição. Por essa razão, conceitos de estatística vão aparecer no desenvolvimento dos algoritmos. O primeiro deles, a *aproximação pela média amostral*, é um método de amostragem exterior: obtém-se uma amostra e, em seguida, se resolve o problema. Já a *decomposição estocástica* é um método interior: a amostragem ocorre durante a execução do algoritmo.

### 6.1 Aproximação pela média amostral

Considere o problema

$$\boxed{\begin{array}{l} \text{VAL} = \min_{\mathbf{x} \in \mathbf{X}} \left\{ f(\mathbf{x}) = \mathbb{E}_{\boldsymbol{\omega}}[F(\mathbf{x}, \boldsymbol{\omega})] = \int_{\Omega} G(\mathbf{x}, \boldsymbol{\omega})g(\mathbf{x})d\mathbf{x} \right\} \\ \text{sujeito a} \quad \mathbf{Ax} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}, \end{array}} \quad (6.1)$$

onde  $g$  é a densidade de  $\boldsymbol{\omega}$ . O problema de recurso com dois estágios (4.8) é um caso particular de (6.1). Para ver isso defina

$$\mathbf{X} = \{\mathbf{x} \mid \mathbf{Ax} = \mathbf{b}\}, \quad f(\mathbf{x}) = \mathbf{c}^T \mathbf{x} + Q(\mathbf{x}), \quad Q(\mathbf{x}) = \mathbb{E}[Q(\mathbf{x}, \boldsymbol{\omega})], \\ Q(\mathbf{x}, \boldsymbol{\omega}) = \min_{\mathbf{y} \in \mathbf{Y}} \{\mathbf{q}^T \mathbf{y} \mid \mathbf{Wy} = \mathbf{h}(\boldsymbol{\omega}) - \mathbf{T}(\boldsymbol{\omega})\mathbf{x}\}.$$

A principal dificuldade em resolver o problema (6.1) é a presença da esperança, que é uma integral. Em geral, não é possível resolver explicitamente

essa integral, que pode ser  $n$ -dimensional se o vetor de variáveis aleatórias tiver dimensão  $n$ . Como essa integral é a esperança de  $F(\mathbf{x}, \boldsymbol{\omega})$ , podemos aproximá-la pela soma de  $n$  sorteios da variável aleatória  $\boldsymbol{\omega}$  e dividir o resultado por  $n$ .

Essa idéia é natural: pense num lançamento de uma moeda honesta modelado por uma variável aleatória  $Y$ , que assume o valor 1 quando o resultado é cara e 0 quando é coroa. Se quisermos aproximar a média dessa variável aleatória, que é  $1/2$ , podemos lançar uma moeda digamos 100 vezes, somar os valores obtidos e dividir o total por 100. Espera-se que o valor seja perto de  $1/2$ .

Vamos definir a *aproximação da média amostral* (AMA), que é uma aproximação para o problema (6.1):

$$\boxed{\hat{v}_n = \min_{\mathbf{x} \in \mathbf{X}} \left\{ \tilde{f}_N(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^N F(\mathbf{x}, \boldsymbol{\omega}^i) \right\}.} \quad (6.2)$$

O estimador  $\tilde{f}_N(\mathbf{x})$  é não-viesado para o valor ótimo  $f(\mathbf{x})$  do problema (6.1):

$$\mathbb{E}_{\boldsymbol{\omega}}[\tilde{f}(\mathbf{x})] = \frac{1}{N} \mathbb{E}_{\boldsymbol{\omega}} \left[ \sum_{i=1}^N F(\mathbf{x}, \boldsymbol{\omega}) \right] = \frac{1}{N} N f(\mathbf{x}) = f(\mathbf{x}).$$

## Construindo uma cota inferior

**Lema 6.1** O estimador  $\hat{v}_N$  subestima o valor ótimo  $v^*$  do problema (6.1), isto é,

$$\mathbb{E}[\hat{v}_N] \leq v^*.$$

**Prova.** Note que

$$v^* = \min_{\mathbf{x} \in \mathbf{X}} \mathbb{E}_{\boldsymbol{\omega}} [F(\mathbf{x}, \boldsymbol{\omega})] = \min_{\mathbf{x} \in \mathbf{X}} \left\{ \mathbb{E}_{\boldsymbol{\omega}} \left[ N^{-1} \sum_{j=1}^N F(\mathbf{x}, \boldsymbol{\omega}^j) \right] \right\}.$$

Então

$$\begin{aligned}
 \min_{\mathbf{x} \in \mathbf{X}} \left\{ \frac{1}{N} \sum_{j=1}^N F(\mathbf{x}, \boldsymbol{\omega}^j) \right\} &\leq \frac{1}{N} \sum_{j=1}^N F(\mathbf{x}, \boldsymbol{\omega}^j) \\
 &\Downarrow \\
 \mathbb{E}_{\boldsymbol{\omega}} \left[ \min_{\mathbf{x} \in \mathbf{X}} \left\{ \frac{1}{N} \sum_{j=1}^N F(\mathbf{x}, \boldsymbol{\omega}^j) \right\} \right] &\leq \mathbb{E}_{\boldsymbol{\omega}} \left[ \frac{1}{N} \sum_{j=1}^N F(\mathbf{x}, \boldsymbol{\omega}^j) \right] \\
 &\Downarrow \\
 \mathbb{E}_{\boldsymbol{\omega}}[\hat{v}_N] &\leq \min_{\mathbf{x} \in \mathbf{X}} \mathbb{E}_{\boldsymbol{\omega}} \left[ \frac{1}{N} \sum_{j=1}^N F(\mathbf{x}, \boldsymbol{\omega}^j) \right] = v^*.
 \end{aligned}$$

■

Temos assim uma cota inferior para o problema original (6.1). Infelizmente, não é fácil calcular  $\mathbb{E}_{\boldsymbol{\omega}}[\hat{v}_N]$ . O que faremos é aproximar esse valor através de amostragem. Gere  $M$  amostras independentes  $\boldsymbol{\omega}^{1,j}, \dots, \boldsymbol{\omega}^{N,j}, j = 1, \dots, M$ , de tamanho  $N$ . Para cada pacote  $j$  de  $N$  amostras resolva o problema AMA correspondente:

$$\hat{v}_N^j = \min_{\mathbf{x} \in \mathbf{X}} \left\{ \frac{1}{N} \sum_{i=1}^N F(\mathbf{x}, \boldsymbol{\omega}^{i,j}) \right\}. \quad (6.3)$$

Cada um dos  $j$  problemas (6.3) fornece uma realização da variável aleatória  $\hat{v}_N^j$ . Logo, conseguimos uma aproximação para  $\mathbb{E}[\hat{v}_N^j]$  tirando a média dos  $M$  problemas (6.3):

$$L_{N,M} = \frac{1}{M} \sum_{j=1}^M \hat{v}_N^j. \quad (6.4)$$

É fácil ver que o estimador  $L_{N,M}$  é não viesado para  $\mathbb{E}_{\boldsymbol{\omega}}[\hat{v}_N^j]$  e, portanto, é um bom candidato para aproximar a cota inferior do problema original (6.1).

Podemos construir um intervalo de confiança para  $\mathbb{E}_{\boldsymbol{\omega}}[\hat{v}_N^j]$ , da mesma forma que foi feito no apêndice B, equação (B.7). Pelo teorema central do limite (A.2), em particular os resultados equivalentes (A.3), temos

$$\sqrt{n} (L_{N,M} - \mathbb{E}[\tilde{v}_N]) \xrightarrow{d} \eta(0, \sigma_L^2),$$

onde  $\eta(\mu, \sigma^2)$  é a variável aleatória normal com média  $\mu$  e variância  $\sigma^2$ . Para aproximar a variância  $\sigma_L^2$ , vamos usar o estimador  $S^2$ , definido em (B.2).

$$\hat{\sigma}_L^2 = \frac{1}{M-1} \sum_{j=1}^M (\hat{v}_N^j - L_{N,M})^2. \quad (6.5)$$

Por fim basta escolher um tolerância  $\alpha$  e encontrar o valor  $z_{\alpha/2}$  correspondente (ver equação (B.3)). Assim como foi feito para (B.7), temos um intervalo (aproximado)  $100(1-\alpha)\%$  confiável para o parâmetro  $\mathbb{E}[\hat{v}_N]$ :

$$\left[ L_{N,M} - \frac{z_{\alpha} \hat{\sigma}_L}{\sqrt{M}}, L_{N,M} + \frac{z_{\alpha} \hat{\sigma}_L}{\sqrt{M}} \right] \quad (6.6)$$

### Construindo uma cota superior

Vamos agora tentar encontrar uma cota superior para o problema (6.1). Para isso, considere  $\hat{\mathbf{x}}$  um ponto admissível para o problema (6.1). Por se tratar de um problema de minimização, é imediato que  $f(\hat{\mathbf{x}})$  é uma cota superior para o problema (6.1). Nossa tarefa será encontrar um bom estimador (não-viesado) para  $f(\hat{\mathbf{x}})$  e assim obter uma cota superior para o problema (6.1).

Comece gerando  $T$  amostras independentes  $\omega^{1,j}, \dots, \omega^{\bar{N},j}$  de tamanho  $\bar{N}$ . Para cada um dos  $T$  pacotes de amostras defina

$$\hat{f}_{\bar{N}}^j(\mathbf{x}) = \frac{1}{\bar{N}} \sum_{i=1}^{\bar{N}} F(\mathbf{x}, \omega^{i,j}), \quad \forall \mathbf{x} \in \mathbf{X} \text{ e } j = 1, \dots, T. \quad (6.7)$$

Em particular,  $\hat{f}_{\bar{N}}^j(\hat{\mathbf{x}})$  é uma aproximação para  $f(\hat{\mathbf{x}})$ , para todo  $j$ . Tirando a média entre as  $T$  aproximações temos um estimador não-viesado para  $f(\hat{\mathbf{x}})$ :

$$U_{\bar{N},T}(\hat{\mathbf{x}}) = \frac{1}{T} \sum_{j=1}^T \hat{f}_{\bar{N}}^j(\hat{\mathbf{x}}). \quad (6.8)$$

Aplicando novamente o teorema central do limite A.2, temos

$$\sqrt{T} (U_{\bar{N},T} - f(\hat{\mathbf{x}})) \xrightarrow{d} \eta(0, \sigma_U^2), \quad (6.9)$$



onde  $\sigma_U^2$  é a variância de  $f(\hat{\mathbf{x}})$ . Novamente vamos aproximar a variância pelo estimador  $S^2$ :

$$\hat{\sigma}_U^2 = \frac{1}{T-1} \sum_{j=1}^T (\hat{f}_{\bar{N}}^j(\hat{\mathbf{x}}) - U_{\bar{N},T})^2.$$

Para uma tolerância  $\alpha$ , temos imediatamente um intervalo  $100(1-\alpha)\%$  confiável para  $f(\hat{\mathbf{x}})$ :

$$\left[ U_{\bar{N},T}(\hat{\mathbf{x}}) - \frac{z_\alpha \hat{\sigma}_U(\hat{\mathbf{x}})}{\sqrt{T}}, U_{\bar{N},T} + \frac{z_\alpha \hat{\sigma}_U(\hat{\mathbf{x}})}{\sqrt{T}} \right] \quad (6.10)$$

### Um algoritmo usando as cotas

Agora que sabemos calcular cotas inferiores e superiores para um problema do tipo (6.1), podemos obter uma aproximação do valor ótimo deste problema. Vamos descrever um algoritmo que nada mais é do que uma estruturação das idéias para calcular cotas descrita nas seções anteriores.

O primeiro passo é fixar valores para  $N$  e  $M$ , por exemplo  $N = 100$  e  $M = 12$ . Em seguida, sorteiam-se  $M$  amostras de tamanho  $N$ . Em cada um dos  $M$  sorteios obtemos uma aproximação (6.3) para o valor ótimo de (6.1). O estimador  $L_{N,M}$  para a cota inferior é obtido tirando-se a média destes valores e o intervalo de confiança é construído conforme (6.6). Além do valor ótimo, em cada um dos  $M$  problemas foi obtido um ponto admissível  $\hat{\mathbf{x}}_j$ ,  $j = 1, \dots, M$ . Fazendo  $T = 50$  por exemplo e escolhendo  $\bar{N} = 20\,000$ , construímos aproximações (6.8) para  $f(\hat{\mathbf{x}}_j)$ . Escolhemos a menor entre as  $M$  cotas superiores e construímos o intervalo de confiança correspondente (6.10).

Esse método é efetivamente usado para aproximar as soluções de problemas de recurso com dois estágios de grande porte, inviáveis de serem resolvidos diretamente. Para alguns desses problemas as cotas inferior e superior ficaram extremamente próximas, indicando que a solução ótima provavelmente foi estimada com precisão. Para a descrição precisa dos problemas e maiores detalhes desse método sugerimos o artigo [15].

## 6.2 A decomposição estocástica

Esse método foi desenvolvido por J. Hige e S. Sen em 1991, e posteriormente publicado em livro [10]. Diferentemente do método da aproximação pela média amostral, a decomposição estocástica (DE) é um método de *amostragem interior*, ou seja, as amostras são feitas durante a execução do algoritmo. A DE também supõe que é possível obter amostras de uma variável ou vetor aleatório qualquer. É um método exclusivamente para problemas dois estágios e se baseia em aproximações lineares por partes da função objetivo do problema de recurso (4.8). Para chegar no algoritmo da DE vamos passar por vários algoritmos intermediários.

### 6.2.1 O algoritmo de planos de corte de Kelley

O algoritmo da DE é uma adaptação de algoritmos existentes de otimização, visando eficiência computacional. O primeiro deles é o algoritmo de Kelley, que se aplica a problemas da forma

$$\min_{\mathbf{x} \in \mathbf{X}} \{ \mathbf{c}\mathbf{x} + \mathcal{Q}(\mathbf{x}) \}, \quad (6.11)$$

onde  $\mathcal{Q}$  é uma função convexa e  $\mathbf{X}$  é um conjunto compacto e convexo. A escolha na notação não é por acaso: a função  $\mathcal{Q}$  do problema (4.8) possui estas propriedades.

PASSO 0: Suponha  $\mathbf{x}^1 \in \mathbf{X}$  dado.

Faça  $k = 0$ ,  $\varphi_0(\mathbf{x}) = -\infty$ ,  $u_0 = \mathbf{c}\mathbf{x}^1 + \mathcal{Q}(\mathbf{x}^1)$  e  $l_0 = -\infty$ .

PASSO 1: Faça  $k = k + 1$ . Encontre  $(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k)$  tais que  $\mathcal{Q}(\mathbf{x}^k) = \boldsymbol{\alpha}^k + \boldsymbol{\beta}^k \mathbf{x}^k$  e  $\mathcal{Q}(\mathbf{x}) \leq \boldsymbol{\alpha}^k + \boldsymbol{\beta}^k \mathbf{x}$ ,  $\forall \mathbf{x} \in \mathbf{X}$ .

PASSO 2: Faça  $u_k = \min\{u_{k-1}, \mathbf{c}\mathbf{x}^k + \mathcal{Q}(\mathbf{x}^k)\}$  e defina  $\varphi_k(\mathbf{x}) = \max\{\varphi_{k-1}(\mathbf{x}), \boldsymbol{\alpha}^k + \boldsymbol{\beta}^k \mathbf{x}\}$ .

PASSO 3: Faça  $l_k = \min_{\mathbf{x} \in \mathbf{X}} \{ \mathbf{c}\mathbf{x} + \varphi_k(\mathbf{x}) \}$  e denote por  $\mathbf{x}^{k+1}$  a solução.

PASSO 4: Se  $u_k - l_k = 0$ , então pare. Caso contrário, volte ao PASSO 1.

A idéia central do algoritmo é aproximar a função  $Q$  pelos planos de corte  $\varphi_k(\mathbf{x})$ . Mas como encontrar  $(\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k)$  no PASSO 1? Considere o dual do problema  $Q(\mathbf{x}^k, \boldsymbol{\omega})$ , para cada elemento  $\boldsymbol{\omega}$  no espaço amostral:

$$\boxed{Q(\mathbf{x}, \boldsymbol{\omega}) = \max_{\mathbf{p}} \mathbf{p}^T(\mathbf{h}(\boldsymbol{\omega}) - \mathbf{T}(\boldsymbol{\omega})\mathbf{x}^k) \text{ sujeito a } \mathbf{W}^T \mathbf{p} \leq \mathbf{q}.} \quad (6.12)$$

Denotando por  $\mathbf{p}^k(\boldsymbol{\omega})$  a solução ótima do dual para cada  $\boldsymbol{\omega}$ , temos que

$$\begin{aligned} Q(\mathbf{x}^k) &= \mathbb{E}_{\boldsymbol{\omega}} [\mathbf{p}^k(\boldsymbol{\omega})(\mathbf{h}(\boldsymbol{\omega}) - \mathbf{T}(\boldsymbol{\omega})\mathbf{x}^k)] \\ &= \mathbb{E}_{\boldsymbol{\omega}} [\mathbf{p}^k(\boldsymbol{\omega})\mathbf{h}(\boldsymbol{\omega})] - \mathbb{E}_{\boldsymbol{\omega}} [\mathbf{p}^k(\boldsymbol{\omega})\mathbf{T}(\boldsymbol{\omega})] \mathbf{x}^k. \end{aligned} \quad (6.13)$$

Por outro lado, se deixarmos  $\mathbf{x}$  variar temos que

$$\begin{aligned} Q(\mathbf{x}) &\geq \mathbb{E}_{\boldsymbol{\omega}} [\mathbf{p}^k(\boldsymbol{\omega})(\mathbf{h}(\boldsymbol{\omega}) - \mathbf{T}(\boldsymbol{\omega})\mathbf{x})] \\ &= \mathbb{E}_{\boldsymbol{\omega}} [\mathbf{p}^k(\boldsymbol{\omega})\mathbf{h}(\boldsymbol{\omega})] - \mathbb{E}_{\boldsymbol{\omega}} [\mathbf{p}^k(\boldsymbol{\omega})\mathbf{T}(\boldsymbol{\omega})] \mathbf{x}. \end{aligned} \quad (6.14)$$

Por (6.13) e (6.14), podemos usar o algoritmo de Kelley em otimização estocástica tomando

$$\boldsymbol{\alpha}^k = \mathbb{E}_{\boldsymbol{\omega}} [\mathbf{p}^k(\boldsymbol{\omega})\mathbf{h}(\boldsymbol{\omega})] \quad \text{e} \quad \boldsymbol{\beta}_k = -\mathbb{E}[\mathbf{p}^k(\boldsymbol{\omega})\mathbf{T}(\boldsymbol{\omega})]. \quad (6.15)$$

Tudo parece resolvido, mas temos um problema computacional: esse procedimento pressupõe que em cada passo se resolva um problema de otimização para cada  $\boldsymbol{\omega}$  pertencente ao espaço amostral  $\Omega$ . Se tivermos um vetor aleatório independente com 10 coordenadas onde cada componente aleatória assume 4 valores, temos que resolver um total de  $4^{10} = 1\,048\,576$  problemas de otimização em cada passo!

### 6.2.2 AMA + Kelley

O próximo passo na caminhada rumo a DE é combinar o método AMA descrito em (6.1) com a algoritmo de Kelley. Na realidade, queremos uma versão um pouco diferente do AMA: seja

$$Q_k(\mathbf{x}) = \frac{1}{k} \sum_{t=1}^k Q(\mathbf{x}, \boldsymbol{\omega}^t), \quad (6.16)$$

onde  $\omega^t$  são observações de uma variável aleatória  $\omega$ . Assim com no algoritmo AMA, a função  $Q_k$  é uma aproximação de  $Q(\mathbf{x}) = \mathbb{E}_\omega[Q(\mathbf{x}, \omega)]$ . Essa função é convexa (por que?) e, portanto, podemos usar o algoritmo de Kelley para resolver o problema

$$\min_{\mathbf{x} \in \mathbf{X}} \{f_k(\mathbf{x}) = \mathbf{c}\mathbf{x} + Q_k(\mathbf{x})\}. \quad (6.17)$$

Intuitivamente, esperamos que as funções  $Q_k$  sejam aproximações cada vez melhores de  $Q(\mathbf{x})$ , quando  $k$  cresce. O algoritmo abaixo, chamado *aproximação da média amostral sucessiva* (AMAS), parte dessa premissa.

PASSO 0: Faça  $k = 0$  e  $Q_0(\mathbf{x}) \equiv 0$ .

PASSO 1: Faça  $k = k + 1$  e gere uma observação  $\omega^k$  de  $\omega$  independente das anteriores.

PASSO 2: Faça  $Q_k(\mathbf{x}) = \frac{k-1}{k} Q_{k-1}(\mathbf{x}) + \frac{1}{k} Q(\mathbf{x}, \omega^k)$ .

PASSO 3: Resolva  $\min_{\mathbf{x} \in \mathbf{X}} \{\mathbf{c}\mathbf{x} + Q_k(\mathbf{x})\}$  e denote por  $\mathbf{x}^k$  a solução ótima. Volte ao PASSO 1.

É possível mostrar, usando a lei forte dos grandes números (A.1), que existe uma subsequência de iterações  $Q_k(\mathbf{x})$  que converge para  $Q(\mathbf{x})$  sempre que  $\mathbf{x}_k$  convergem para  $\mathbf{x}^*$ , a solução ótima de (4.8). No entanto, esse algoritmo em geral é ineficiente. A razão principal é que no PASSO 2 temos que resolver um problema de otimização “cru” em cada passo, ou seja, um problema onde não existem restrições sobre onde procurar uma solução. Além disso, em cada passo minimizamos  $Q_k(\mathbf{x})$  baseado numa nova saída de  $\omega$ , mesmo que essa saída não melhore muito a aproximação  $Q_n$ .

Uma maneira de melhorar o algoritmo AMAS é reduzir o custo computacional no PASSO 2, incluindo cortes que auxiliam na resolução do problema. A idéia é aproveitar o aninhamento das amostras, ou seja, o fato que  $\{\omega^t\}_{t=1}^{k-1} \subset \{\omega^t\}_{t=1}^k$ .

Suponha que usamos o método de Kelley para otimizar  $f_{k-1}$  e sejam  $\{\mathbf{x}^j\}_{j=1}^q$  os pontos onde os cortes foram derivados. Os coeficientes desses cortes são  $(\alpha_j^{k-1}, \beta_j^{k-1})$ , onde o subscrito denota que o corte foi derivado

em  $\mathbf{x}^j$  e o sobrescrito lembra que esses cortes foram obtidos na aplicação de Kelley a  $f_{k-1}$ . Queremos achar expressões para cortes baseados em amostras de tamanho  $k$   $(\boldsymbol{\alpha}_j^k, \boldsymbol{\beta}_j^k)$  a partir dos cortes  $(\boldsymbol{\alpha}_j^{k-1}, \boldsymbol{\beta}_j^{k-1})$  para amostras de tamanho  $k-1$ .

Denotando por  $p^j(\boldsymbol{\omega}^t)$  a solução ótima do dual do problema de segundo estágio para cada  $\boldsymbol{\omega}^t$ , temos por definição (6.15) que

$$\boldsymbol{\alpha}_j^{k-1} + \boldsymbol{\beta}_j^{k-1} \mathbf{x} = \frac{1}{k-1} \sum_{t=1}^{k-1} p^j(\boldsymbol{\omega}^t) (\mathbf{h}^t - \mathbf{T}^t \mathbf{x}) \quad \text{e} \quad (6.18)$$

$$\boldsymbol{\alpha}_j^k + \boldsymbol{\beta}_j^k \mathbf{x} = \frac{1}{k} \sum_{t=1}^k p^j(\boldsymbol{\omega}^t) (\mathbf{h}^t - \mathbf{T}^t \mathbf{x}), \quad (6.19)$$

onde  $\mathbf{h}^t = \mathbf{h}(\boldsymbol{\omega}^t)$  e  $\mathbf{T}^t = \mathbf{T}(\boldsymbol{\omega}^t)$ . Assim, usando (6.18) e (6.19), definimos os cortes na iteração  $k$  recursivamente por

$$\boldsymbol{\alpha}_j^k + \boldsymbol{\beta}_j^k \mathbf{x} = \frac{k-1}{k} (\boldsymbol{\alpha}_j^{k-1} + \boldsymbol{\beta}_j^{k-1} \mathbf{x}) + \frac{1}{k} p^j(\boldsymbol{\omega}^k) (\mathbf{h}^k - \mathbf{T}^k \mathbf{x}), \quad (6.20)$$

para cada  $j = 1, \dots, q$ . Essa definição recursiva permite que no PASSO 3 do AMAS + Kelley otimizemos a função  $\mathcal{Q}_k$  com a inclusão destes cortes. No entanto, para calcular  $\mathbf{p}^j(\boldsymbol{\omega}^k)$  temos que guardar todos os vetores  $\{\mathbf{x}^j\}_{j=1}^q$  e resolver os sub-problemas duais (6.12) associados. Infelizmente, com o crescimento do número de iterações, o custo computacional destas duas operações cresce rapidamente e não compensa mais a inclusão dos cortes para evitar a resolução de um problema “cru” no PASSO 2 do AMAS.

A DE que veremos em seguida é exatamente uma estratégia para aliviar o custo computacional do PASSO 2 do AMAS. Ela evita o armazenamento dos vetores  $\{\mathbf{x}^j\}_{j=1}^q$  e a resolução dos sub-problemas duais (6.12). Em contrapartida, o processo de formação de cortes será menos eficiente.

## Método estocástico dos planos de cortes e decomposição estocástica

Vamos ver um último algoritmo antes da DE: o método estocástico dos planos de cortes (MEPC). Assuma que exista um número  $L$  tal que  $L \leq Q(\mathbf{x}, \boldsymbol{\omega})$ . Vamos a descrição do algoritmo:

PASSO 0: Faça  $k = 0$ ,  $\nu_0(\mathbf{x}) = -\infty$  e tome  $\mathbf{x}^1 \in \mathbf{X}$ . Seja  $L$  um número tal que  $L \leq Q(\mathbf{x}, \boldsymbol{\omega})$

PASSO 1: Faça  $k = k + 1$  e gere uma observação  $\boldsymbol{\omega}^t$  de  $\boldsymbol{\omega}$  independente das anteriores.

PASSO 2: Cálculo de  $\nu_k(\mathbf{x})$ , uma aproximação linear por partes de  $Q_k$ :

(a) Cálculo do  $k$ -ésimo plano de corte:

$$\boldsymbol{\alpha}_k^k + \boldsymbol{\beta}_k^k \mathbf{x} = \frac{1}{k} \sum_{t=1}^k \mathbf{p}^k(\boldsymbol{\omega}^t)(\mathbf{h}^t - \mathbf{T}^t \mathbf{x}),$$

onde  $\mathbf{p}^k(\boldsymbol{\omega}^k) \in \operatorname{argmax}_{\mathbf{p}} \{\mathbf{p}^T(\mathbf{h}^k - \mathbf{T}^k \mathbf{x}^k) \mid \mathbf{W}^T \mathbf{p} \leq \mathbf{q}\}$ .

(b) Atualização dos coeficientes de todos os cortes anteriores:

$$\boldsymbol{\alpha}_t^k = \frac{k-1}{k} \boldsymbol{\alpha}_t^{k-1} + \frac{1}{k} L, \quad \boldsymbol{\beta}_t^k = \frac{k-1}{k} \boldsymbol{\beta}_t^{k-1}.$$

(c) Defina  $\nu_k(\mathbf{x}) = \max_{1 \leq t \leq k} \{\boldsymbol{\alpha}_t^k + \boldsymbol{\beta}_t^k \mathbf{x}\}$ .

PASSO 3: Resolva  $\min_{\mathbf{x} \in \mathbf{X}} \{\mathbf{c}\mathbf{x} + \nu_k(\mathbf{x})\}$  e denote por  $\mathbf{x}^{k+1}$  a solução. Volte ao PASSO 1.

Começemos traçando algumas semelhanças e diferenças entre o AMAS e o MEPC. Ambos geram uma observação da variável aleatória no curso do algoritmo e a função a ser otimizada em cada passo é uma aproximação da função de recurso  $Q(\mathbf{x}) = \mathbb{E}_{\boldsymbol{\omega}}[Q(\mathbf{x}, \boldsymbol{\omega})]$ . A natureza dessa aproximação, no entanto, é muito diferente. O AMAS gera uma observação e otimiza a função  $Q_k$ , enquanto que o MEPC usa planos de corte para aproximar a função  $Q(\mathbf{x})$ .

Aqui aparece uma diferença em relação ao algoritmo de Kelley: os planos de corte em MEPC só são tangentes a função  $Q(\mathbf{x})$  no momento em que são criados: em Kelley, os planos permanecem tangentes durante todas as iterações do algoritmo. Mais ainda: cada corte em MEPC é gerado com um número diferente de observações de  $\boldsymbol{\omega}$ .

Quando se usa o algoritmo de Kelley combinado com o AMAS, temos que em cada iteração os cortes são gerados de um mesmo número de observações. A aproximação é mais uniforme, enquanto que em MEPC os cortes gerados nas últimas iterações, quando a amostra é maior, fornecem aproximações mais precisas da função  $Q$  do que os cortes gerados nas primeiras iterações.

Em relação a maneira como os cortes são atualizados, podemos interpretar o PASSO 2 (b) de MEPC como uma relaxação do procedimento de atualização de cortes de AMAS usando Kelley (equações (6.20)). É fácil ver que

$$\nu_k(\mathbf{x}) = \max \left\{ \frac{k-1}{k} \nu_{k-1}(\mathbf{x}) + \frac{1}{k} L, \boldsymbol{\alpha}_k^k + \boldsymbol{\beta}_k^k \mathbf{x} \right\} \quad (6.21)$$

Comparando-se com (6.20), podemos ver que o termo  $p^t(\boldsymbol{\omega}^k)(\mathbf{h}^k - T^k \mathbf{x})$  é trocado simplesmente por  $L$  em MEPC.

Agora estamos prontos para a decomposição estocástica. Em cada algoritmo que vimos havia uma pequena diferença que o tornava mais eficiente que o anterior. O AMAS combinado com Kelley é mais eficiente que AMAS, pois a inclusão de cortes permite que se otimize a função  $Q_k$  de maneira mais eficiente.

O método MEPC, por sua vez, apresenta uma maneira mais eficiente de se atualizar cortes, que dá mais peso a cortes gerados quando o tamanho da amostra é maior. No entanto, o PASSO 2 deste algoritmo ainda depende da resolução de um problema de otimização para cada  $\boldsymbol{\omega}$  no espaço amostral o que pode tornar o algoritmo intratável computacionalmente.

É exatamente esse ponto que a DE apresenta uma modificação. Ao invés de resolver estes problemas, o algoritmo resolve apenas um problema por iteração, baseado na amostra recém gerada. Essa solução é um vértice do conjunto admissível  $\Pi = \{\mathbf{p} \mid \mathbf{W}^T \mathbf{p} \leq \mathbf{q}\}$  de (6.12). Os vértices que vão sendo obtidos vão sendo armazenados no conjunto  $V_k$ . Novamente, assumimos que uma cota inferior  $L$  para  $Q(\mathbf{x}, \boldsymbol{\omega})$  é dada.

PASSO 0: Faça  $k = 0$ ,  $\eta_0(\mathbf{x}) = -\infty$  e tome  $\mathbf{x}^1 \in \mathbf{X}$ . Seja  $L$  um número tal que  $L \leq Q(\mathbf{x}, \boldsymbol{\omega})$

PASSO 1: Faça  $k = k + 1$  e gere uma observação  $\boldsymbol{\omega}^t$  de  $\boldsymbol{\omega}$  independente das anteriores.

PASSO 2: Cálculo de  $\eta_k(\mathbf{x})$ , uma aproximação linear por partes de  $Q_k$ :

(a) Resolva o subproblema dual e atualize  $V_k$ :

$$\mathbf{p}^k(\boldsymbol{\omega}^k) = \operatorname{argmax}_{\mathbf{p}} \{ \mathbf{p}^T (\mathbf{h}^k - \mathbf{T}^k \mathbf{x}^k) \mid \mathbf{W}^T \mathbf{p} \leq \mathbf{q} \} \text{ e}$$

$$V_k = V_{k-1} \cup \{ \mathbf{p}^k(\boldsymbol{\omega}^k) \}.$$

(b) Cálculo do  $k$ -ésimo plano de corte:

$$\boldsymbol{\alpha}_k^k + \boldsymbol{\beta}_k^k \mathbf{x} = \frac{1}{k} \sum_{t=1}^k \mathbf{p}^k(\boldsymbol{\omega}^t) (\mathbf{h}^t - \mathbf{T}^t \mathbf{x}),$$

$$\text{onde } \mathbf{p}^k(\boldsymbol{\omega}^t) = \operatorname{argmax}_{\mathbf{p} \in V_k} \{ \mathbf{p}^T (\mathbf{h}^k - \mathbf{T}^k \mathbf{x}) \}.$$

(c) Atualização dos coeficientes de todos os cortes anteriores:

$$\boldsymbol{\alpha}_t^k = \frac{k-1}{k} \boldsymbol{\alpha}_t^{k-1} + \frac{1}{k} L, \quad \boldsymbol{\beta}_t^k = \frac{k-1}{k} \boldsymbol{\beta}_t^{k-1}.$$

(d) Defina  $\eta_k(\mathbf{x}) = \max_{1 \leq t \leq k} \{ \boldsymbol{\alpha}_t^k + \boldsymbol{\beta}_t^k \mathbf{x} \}$ .

PASSO 3: Resolva  $\min_{\mathbf{x} \in \mathbf{X}} \{ \mathbf{c} \mathbf{x} + \eta_k(\mathbf{x}) \}$  e denote por  $\mathbf{x}^{k+1}$  a solução. Volte ao PASSO 1.

Pode-se mostrar que existe uma subsequência do algoritmo da DE que converge para a solução ótima do problema (4.8). O algoritmo que apresentamos aqui é a versão mais simples da DE. Existem melhorias que estabilizam o método e aumentam sua eficiência. Para maiores detalhes sobre o algoritmo e orientações para a implementação, sugerimos ao leitor o excelente livro [10].

A DE foi aplicada com sucesso em problemas de grande porte e permanece até hoje como um algoritmo eficiente na resolução de problemas de otimização estocástica.



## Exercícios

[01] Considere o problema (4.8), com:

$$\begin{aligned} \mathbf{X} &= [0, 5], & \mathbf{x} &= [x_1], & \mathbf{y} &= [y_1 \ y_2 \ y_3 \ y_4]^T, \\ \mathbf{c} &= [-0.75], & \mathbf{q} &= [-1 \ 3 \ 1 \ 1]^T, & \mathbf{h} &= [\omega_1 \ \omega_2]^T, \\ \mathbf{T} &= \begin{bmatrix} 10 & 8 & 0 \\ 5 & 0 & 8 \end{bmatrix} & \text{e} & \mathbf{W} &= \begin{bmatrix} -1 & 1 & -1 & 1 \\ -1 & 1 & 1 & -1 \end{bmatrix}, \end{aligned}$$

onde  $\omega_1$  é a distribuição uniforme no intervalo  $[-1, 0]$  e  $\omega_2 = 1 + \omega_1$ . Faça três passos do algoritmo de decomposição estocástica e encontre uma aproximação para  $x$ .

# Apêndice A

## Probabilidade

Este apêndice não se propõe a ser um resumo completo de probabilidade: queremos definir apenas os conceitos de probabilidade usados no texto. O leitor interessado em se aprofundar no tema pode consultar [12]. O primeiro conceito que vamos definir é o de *variável aleatória*. Para isso precisamos de duas definições preliminares:

**Definição A.1** Uma coleção não-vazia de conjuntos  $\mathcal{F}$  de um conjunto  $\Omega$  é dita uma  $\sigma$ -álgebra de subconjuntos de  $\Omega$  se as seguintes propriedades se verificam:

- (a) Se  $A \in \mathcal{F}$  então o complemento  $A^c$  de  $A$  também está em  $\mathcal{F}$ .
- (b) Se  $A_n$  está em  $\mathcal{F}$ ,  $n = 1, 2, \dots$ , então  $\bigcup_{n=1}^{\infty} A_n$  e  $\bigcap_{n=1}^{\infty} A_n$  estão em  $\mathcal{F}$ .

Vamos agora definir o que é uma medida de probabilidade. A partir desse conceito vamos ser capazes de definir o ambiente onde toda a teoria da probabilidade repousa: o *espaço de probabilidade*.

**Definição A.2** Uma medida de probabilidade  $\mathbb{P}$  numa  $\sigma$ -álgebra de subconjuntos  $\mathcal{F}$  de um conjunto  $\Omega$  é uma função real com domínio  $\mathcal{F}$  satisfazendo as seguintes propriedades:

(a)  $\mathbb{P}(\Omega) = 1$ .

(b)  $\mathbb{P}(A) \geq 0$  para todo  $A \in \mathcal{F}$ .

(c) Se  $A_n$ ,  $n = 1, 2, \dots$ , são conjuntos mutuamente disjuntos em  $\mathcal{F}$  então

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

Um espaço de probabilidade  $(\Omega, \mathcal{F}, \mathbb{P})$  é uma tripla formada por um conjunto  $\Omega$ , chamado *espaço amostral*, uma  $\sigma$ -álgebra  $\mathcal{F}$  e uma medida de probabilidade  $\mathbb{P}$  definida em  $\mathcal{F}$ . Os elementos de  $\mathcal{F}$  são chamados *eventos*.

Não é difícil construir espaços de probabilidade para situações simples. Considere uma urna com  $s$  bolas numeradas e suponha que a probabilidade de se retirar uma dada bola é  $1/s$ , ou seja, todas possuem a mesma probabilidade de serem sorteadas. Nesse caso o conjunto  $\Omega$  é um conjunto com  $s$  pontos,  $\mathcal{F}$  é a coleção de todos os subconjuntos de  $\Omega$  e  $\mathbb{P}$  é tal que  $\mathbb{P}(A) = j/s$  se  $A$  possui  $j$  pontos.

Um dos conceitos mais importantes de probabilidade é o de independência. Vamos primeiramente definir o conceito de eventos independentes:

**Definição A.3** Seja  $(\Omega, \mathcal{F}, \mathbb{P})$  um espaços de probabilidade e sejam  $A$  e  $B$  eventos. Dizemos que os eventos  $A$  e  $B$  são *independentes* se

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B).$$

Tendo construído o espaço de probabilidade  $(\omega, \mathcal{F}, \mathbb{P})$ , podemos construir o conceito de variável aleatória:

**Definição A.4** Uma variável aleatória discreta  $X$  num espaço de probabilidade  $(\Omega, \mathcal{F}, \mathbb{P})$  é uma função  $X$  com domínio  $\Omega$  assumindo valores em um subconjunto finito ou enumerável  $\{x_1, x_2, \dots\}$  de  $\mathbb{R}$  tal que  $\{\omega \mid X(\omega) = x_i\} \in \mathcal{F}$  para todo  $i$ .

Essa definição se estende imediatamente para mais dimensões:

**Definição A.5** Sejam  $X_1, X_2, \dots, X_n$  variáveis aleatórias discretas. O vetor aleatório discreto  $\mathbf{X}$  é definido por

$$\mathbf{X} = (X_1, X_2, \dots, X_n).$$

A variável aleatória é uma tradução do conjunto  $\Omega$  para os números reais, onde sabemos trabalhar melhor. Seguindo o exemplo anterior, podemos definir uma variável aleatória no espaço  $(\Omega, \mathcal{F}, \mathbb{P})$  dado. Se  $\omega_i$ ,  $i = 1, \dots, s$  é o elemento de  $\Omega$  associado a  $i$ -ésima bola, então  $X(\omega_i) = i$  é uma variável aleatória em  $(\Omega, \mathcal{F}, \mathbb{P})$ .

**Definição A.6** A função real  $f$  definida por  $f(x) = \mathbb{P}(X = x)$  é chamada função densidade da variável aleatória  $X$ . De maneira similar, sejam  $X_1, X_2, \dots, X_n$  variáveis aleatórias discretas. A função  $g$  de  $\mathbb{R}^n$ , definida por  $g(\mathbf{x}) = \mathbb{P}(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$ , é a densidade conjunta do vetor aleatório  $(X_1, X_2, \dots, X_n)$ , onde  $\mathbf{x}$  pertence a  $\mathbb{R}^n$ .

Ainda no exemplo das bolas, temos que  $f(x) = 1/s$  para  $x = 1, 2, \dots, s$  e  $f(x) = 0$  caso contrário. Passemos ao conceito de independência para variáveis aleatórias discretas:

**Definição A.7** Sejam  $X_1, X_2, \dots, X_n$  variáveis aleatórias discretas, com densidades  $f_1, f_2, \dots, f_n$  e com densidade conjunta  $f$ . Estas variáveis são ditas *independentes* se

$$f(x_1, x_2, \dots, x_n) = f_1(x_1) \cdot f_2(x_2) \cdots f_n(x_n).$$

A seguir vamos definir a esperança de uma variável aleatória discreta.

**Definição A.8** Seja  $X$  uma variável aleatória discreta com densidade  $f$  assumindo valores  $x_1, x_2, \dots, x_j, \dots$ . A esperança de  $X$  é definida por

$$\mathbb{E}[X] = \sum_{j=1}^{\infty} x_j f(x_j),$$

desde que esse valor seja finito. Se for infinito então dizemos que a esperança de  $X$  não está definida.

Voltando ao nosso exemplo, temos que a esperança de  $X$  é  $\mathbb{E}[X] = (1 + s)/2$ . A esperança é linear:

**Lema A.1** Sejam  $\alpha_1, \dots, \alpha_n$  constantes reais e  $X_1, \dots, X_n$  variáveis aleatórias. Então

$$\mathbb{E} \left[ \sum_{i=1}^n \alpha_i X_i \right] = \sum_{i=1}^n \alpha_i \mathbb{E}[X_i].$$

Outro conceito importante é a variância de uma variável aleatória:

**Definição A.9** Seja  $X$  uma variável aleatória discreta com densidade  $f$  assumindo valores  $x_1, x_2, \dots, x_j, \dots$ . A variância  $\sigma^2$  de  $X$  é definida por

$$\sigma^2 = \sum_{j=1}^{\infty} (x_j - \mathbb{E}[X])^2 f(x_j),$$

desde que esse valor seja finito. Se for infinito, então dizemos que a variância de  $X$  não está definida.

O número  $\sigma$  é chamado de *desvio-padrão*.

No exemplo da urna, a variância de  $X$  não é tão fácil de calcular. É preciso usar a fórmula da soma de quadrados a seguir:

$$\sum_{j=1}^s j^2 = \frac{(2s + 1)(s + 1)(s)}{6}. \quad (\text{A.1})$$

Usando (A.1), temos que uma expressão para a variância de  $X$ :

$$\sigma^2(X) = \frac{(s+1)(s-1)}{12}.$$

Para variáveis aleatórias independentes, temos o seguinte resultado:

**Lema A.2** Sejam  $X_1, \dots, X_n$  variáveis aleatórias independentes identicamente distribuídas, com variância  $\sigma^2$ . Então

$$\text{variância}(X_1 + \dots + X_n) = n\sigma^2.$$

Precisamos generalizar esses conceitos para o caso contínuo:

**Definição A.10** Uma variável aleatória contínua  $X$  num espaço de probabilidade  $(\Omega, \mathcal{F}, \mathbb{P})$  é uma função real  $X: \Omega \rightarrow \mathbb{R}$ , tal que, para todo  $x \in \mathbb{R}$ ,  $\{\omega \in \Omega \mid X(\omega) \leq x\} \in \mathcal{F}$ .

A seguir temos a definição de função distribuição de uma variável aleatória contínua ou discreta:

**Definição A.11** A função distribuição  $F$  de uma variável aleatória discreta ou contínua  $X$  definida em  $(\omega, \mathcal{F}, \mathbb{P})$  é definida por

$$F(x) = \mathbb{P}(X \leq x), \quad -\infty < x < \infty.$$

**Definição A.12** Se  $X_1, X_2, \dots, X_n: \Omega \rightarrow \mathbb{R}$  são variáveis aleatórias definidas em um espaço de probabilidade  $(\Omega, \mathcal{F}, \mathbb{P})$ , então a função  $F: \mathbb{R}^n \rightarrow \mathbb{R}$  definida por

$$F(x_1, x_2, \dots, x_n) = \mathbb{P}\{X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n\}$$

é denominada *função distribuição conjunta* do vetor aleatório  $(X_1, X_2, \dots, X_n)$ .

**Definição A.13** Uma função densidade  $f$  é qualquer função não negativa tal que

$$\int_{-\infty}^{\infty} f(x) dx = 1.$$

É comum descrever uma variável aleatória por sua densidade, pois a partir dela podemos definir uma medida de probabilidade e, conseqüentemente, sua função distribuição. Mais precisamente se  $X$  é uma variável aleatória com densidade  $f$  então definimos

$$P(X \leq x) = F(x) = \int_{-\infty}^x f(t) dt. \quad (\text{A.2})$$

**Definição A.14** Seja  $X$  uma variável aleatória contínua com densidade  $f$ . Dizemos que  $X$  tem esperança finita se

$$\int_{-\infty}^{\infty} |x|f(x) dx < \infty,$$

e, nesse caso, definimos a esperança de  $X$  como

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} xf(x) dx.$$

Uma generalização importante da definição A.14 vem a seguir:

**Definição A.15** Seja  $g$  uma função contínua e  $X$  uma variável aleatória contínua com função densidade  $f$ . Então

$$\mathbb{E}[g(x)] = \int_{-\infty}^{\infty} g(x)f(x) dx,$$

caso a integral exista.

Vamos considerar a variável aleatória  $\xi$  contínua do problema do jornalista, que era uniforme entre 50 e 150. Dizer que uma variável aleatória

contínua é uniforme é o mesmo que dizer que a densidade de  $\xi$  é  $f(x) = 1/100$  se  $50 \leq x \leq 150$  e 0 caso contrário. Assim, temos que a função distribuição de  $\xi$  é

$$F(x) = \int_{50}^x \frac{1}{100} dt = \frac{x - 50}{100}.$$

Logo, a esperança de  $\xi$  é obtida usando a definição A.14:

$$\mathbb{E}[\xi] = \int_{50}^{150} x \frac{1}{100} dx = 100.$$

Vamos dar um salto conceitual para enunciar dois dos mais importantes teorema de probabilidade. As demonstrações não são simples e necessitam de conceitos mais avançados de teoria da probabilidade.

**Teorema A.1 (Lei Forte dos Grandes Números)** Sejam  $X_1, X_2, \dots, X_n, \dots$  variáveis aleatórias independentes identicamente distribuídas, com esperança  $\mu$ . Então

$$\frac{X_1 + \dots + X_n}{n} \rightarrow \mu \text{ (quase sempre).}$$

Lembrando que a expressão “quase sempre” significa que o resultado vale a menos de um conjunto com probabilidade 0. Uma conseqüência familiar da lei dos grandes números ocorre no experimento de se lançar repetidamente uma moeda honesta. Associando o valor 1 ao evento “cara” e 0 a “coroa” esperamos que, após um número grande de lançamentos, aproximadamente metade deles dê cara e outra metade coroa. Nesse caso, a lei dos grande números nos diz, de fato, que o número de caras dividido pela quantidade de lançamentos se aproxima de 1/2.

**Teorema A.2 (Teorema Central do Limite)** Considere  $X_1, X_2, \dots, X_n, \dots$  variáveis aleatórias independentes identicamente distribuídas, com esperança  $\mu$  e variância  $\sigma^2 > 0$ . Se  $S_n = X_1 + \dots + X_n$ , então

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x \right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{y^2}{2}} dy.$$

Essa convergência ocorre *em distribuição*.



Em palavras, o que o teorema central do limite diz é que as funções distribuição das somas parciais  $S_n$  devidamente normalizadas, convergem para a função distribuição da variável aleatória normal, não importando a distribuição inicial de  $X_i$ ! O teorema central do limite possui enunciados equivalentes que serão úteis no texto. Denotando-se por  $\eta(\mu, \sigma^2)$  a normal com parâmetros  $\mu$  e  $\sigma^2$  e considerando-se as hipóteses do Teorema A.2, temos que

$$\sqrt{n} \left( \frac{\bar{X} - \mu}{\sigma} \right) \xrightarrow{d} \eta(0, 1) \quad \text{e} \quad \sqrt{n}(\bar{X} - \mu) \xrightarrow{d} \eta(0, \sigma^2). \quad (\text{A.3})$$

A convergência também é em *distribuição*, ou seja, as funções distribuição das variáveis aleatórias  $S_n$  normalizadas convergem para a função distribuição da normal.

Por fim, apresentamos na Tabela A.1 as variáveis aleatórias usadas no texto e suas respectivas densidades, esperanças e variâncias.

	Densidade	$\mathbb{E}[X]$	$\sigma^2(X)$
Uniforme discreta	$f(x s) = \frac{1}{s},$ $x = 1, 2, \dots, s, \quad s = 1, 2, \dots$	$\frac{s+1}{2}$	$\frac{(s+1)(s-1)}{12}$
Uniforme contínua	$f(x a, b) = \frac{1}{b-a},$ $a \leq x \leq b.$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Normal (gaussiana)	$f(x \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$ $-\infty \leq x \leq \infty, \quad \mu \in \mathbb{R},$ $\sigma > 0.$	$\mu$	$\sigma^2$

Tabela A.1: Variáveis aleatória usadas no texto.

# Apêndice B

## Estatística

Este apêndice define os conceitos básicos de estatística, necessários à compreensão do capítulo 6. Estatística é uma área importante para otimização estocástica, especialmente na parte de algoritmos. É comum usar técnicas estatísticas para obter aproximações de problemas de otimização estocástica difíceis de se resolver analiticamente.

O conceito mais importante de estatística é o de estimador.

**Definição B.1** Um estimador é simplesmente uma função  $\theta = \theta(X_1, \dots, X_n)$  com contradomínio em  $\mathbb{R}^n$ , onde  $X_1, \dots, X_n$  são variáveis aleatórias identicamente distribuídas.

Nesse texto vamos considerar  $n = 1$ . Geralmente, um estimador aproxima um parâmetro da densidade de uma variável aleatória  $X$ , sua esperança ou sua variância. É o problema inverso da probabilidade: dados  $n$  valores de uma variável aleatória desconhecida, queremos obter informação sobre essa variável. Vamos ver dois exemplos de estimadores:

$$\bar{X} = \frac{X_1 + \dots + X_n}{n} \text{ e} \tag{B.1}$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \tag{B.2}$$

Uma propriedade desejável de um estimador é que ele seja não-viesado, ou seja, que em média o seu valor seja igual ao parâmetro que ele estima. O

estimador (B.1) é não viesado para a esperança  $\mu$  de uma variável aleatória  $X$ , pois

$$\mathbb{E}[\bar{X}] = \frac{1}{n} n \mathbb{E}[X] = \mu.$$

O estimador  $S^2$  em (B.2) é não-viesado para a variância  $\sigma^2$  de uma variável aleatória  $X$ , mas não vamos mostrar esse fato aqui.

$$\mathbb{E}[S^2] = \sigma^2.$$

Vamos ver um exemplo: suponha que seja dada uma amostra  $(x_1, \dots, x_n)$  de uma variável aleatória  $X$  normal, com parâmetro  $\mu$  desconhecido. Deseja-se construir um estimador que aproxime esse valor, baseado na amostra fornecida. Uma possibilidade é o estimador  $\bar{X}$ , que em media vale  $\mu$ . Assim, uma aproximação  $\tilde{\mu}$  para o parâmetro  $\mu$  é

$$\tilde{\mu} = \frac{x_1 + \dots + x_n}{n}.$$

Como o estimador é uma variável aleatória, podemos calcular sua variância. É preferível trabalhar com estimadores de baixa variância, pois um número pequeno de amostras permite aproximar bem o parâmetro a ser estimado. Como exemplo, note que a variância do estimador  $\bar{X}$  é  $\sigma^2/n$ , onde  $\sigma^2$  é a variância de  $X_i$ .

Apesar do estimador  $\bar{X}$  ser não-viesado para a esperança, é importante obter mais informação sobre a qualidade da aproximação que esse estimador fornece. Essa a idéia dos intervalos de confiança: eles fornecem um intervalo na reta com a propriedade que o parâmetro estimado pertence ao intervalo com probabilidade, digamos, 90%.

Antes de prosseguirmos, vamos enunciar um resultado de probabilidade que será extremamente útil para a construção a seguir. A demonstração pode ser encontrada em [12].

**Lema B.1** Sejam  $X_1, \dots, X_n$  variáveis aleatórias normais com parâmetros  $(\mu_i, \sigma_i^2)$ ,  $i = 1, \dots, n$ . Então a variável aleatória  $Y = X_1 + \dots + X_n$  é normal com parâmetros  $(\mu, \sigma^2)$ , onde

$$\mu = \mu_1 + \dots + \mu_n \quad \text{e} \quad \sigma^2 = \sigma_1^2 + \dots + \sigma_n^2.$$

Para construir um intervalo de confiança, vamos supor que a variável aleatória  $Z$  é normal com parâmetros 0 e 1. O estimador  $\bar{X}$  (B.1) se baseia nas variáveis aleatórias independentes  $X_1, \dots, X_n$ , com distribuição normal de parâmetros  $\mu$  desconhecido e  $\sigma^2$  conhecido. Comece escolhendo um nível de confiabilidade  $\alpha$ . O parâmetro a ser estimado, no caso a esperança  $\mu$ , pertencerá ao chamado *intervalo de confiança* com probabilidade  $1 - \alpha$ . Escolhido  $\alpha$ , defina  $z_\alpha$  como sendo número real tal que

$$\mathbb{P}(\{Z > z_\alpha\}) = \alpha. \quad (\text{B.3})$$

Os valores da normal são tabelados: para um dado  $\alpha$ , basta consultar uma tabela da distribuição normal ([11]) para obter o valor  $z_\alpha$  correspondente. A variável normal tem a propriedade

$$\mathbb{P}(Z > z_\alpha) = 1 - \mathbb{P}(Z > -z_\alpha). \quad (\text{B.4})$$

Usando (B.4), podemos escrever

$$\mathbb{P}(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha. \quad (\text{B.5})$$

O próximo passo é calibrar o estimador  $\bar{X}$ . Como a esperança de  $\bar{X}$  é  $\mu$  e a variância é  $\sigma^2/n$ , a variável aleatória  $\sqrt{n}(\bar{X} - \mu)/\sigma$  é uma normal com parâmetros 0 e 1. Assim, podemos reescrever a equação (B.5) como

$$\mathbb{P}\left(-z_{\alpha/2} < \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} < z_{\alpha/2}\right). \quad (\text{B.6})$$

Isolando-se  $\mu$  em (B.6) temos um intervalo  $(1 - \alpha)\%$  confiável para  $\mu$ :

$$\mathbb{P}\left(\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha. \quad (\text{B.7})$$

Esse intervalo contém o parâmetro  $\mu$  a ser estimado com 95% de certeza. Naturalmente, maior precisão implica em um intervalo maior e menor precisão implica em um intervalo menor.

Apesar de termos feito as contas supondo que a amostra  $X_1, \dots, X_n$  tivesse distribuição normal, isso não é necessário. Pelo teorema central do limite (A.2), a distribuição de  $\sqrt{n}(\bar{X} - \mu)/\sigma$  é aproximadamente normal para  $n$  suficientemente grande. No entanto, é importante ressaltar que o intervalo de confiança construído nessa situação é aproximado.

Outra hipótese que pode ser relaxada é a do conhecimento da variância  $\sigma^2$ . Na maioria das situações, não conhecemos com precisão seu valor, mas podemos substituí-lo pelo estimador  $S^2$  em (B.2), ou por  $\sqrt{S}$  no caso do desvio-padrão. Novamente, não teremos um intervalo de confiança exato, mas aproximado.

# Apêndice C

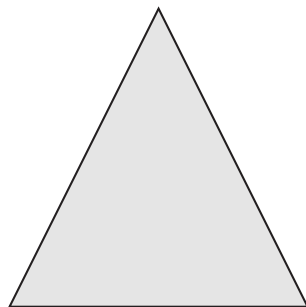
## Convexidade

Neste apêndice apresentaremos as definições e propriedades básicas de funções convexas necessárias no texto. As demonstrações omitidas podem ser encontradas em [04].

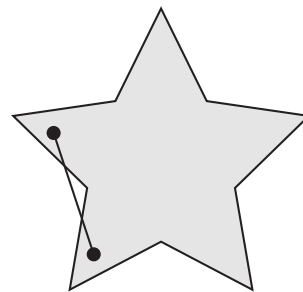
**Definição C.1 (CONJUNTOS CONVEXOS)** Dizemos que  $U \subset \mathbb{R}^n$  é um *conjunto convexo* se, e somente se, para todo  $\mathbf{p}, \mathbf{q} \in U$  tem-se

$$(1 - t) \cdot \mathbf{p} + t \cdot \mathbf{q} \in U,$$

para todo  $t \in [0, 1]$ , isto é, se o segmento de reta que une dois pontos quaisquer de  $U$  está sempre contido em  $U$ .



(a)



(b)

Figura C.1: O conjunto da esquerda é convexo enquanto que o da direita não o é.

**Teorema C.1** Seja  $\{U_\lambda\}_{\lambda \in \Lambda}$  uma família de conjuntos convexos em  $\mathbb{R}^n$ .  
Então

$$\bigcap_{\lambda \in \Lambda} U_\lambda$$

também é um conjunto convexo em  $\mathbb{R}^n$ .

**Definição C.2** (FUNÇÕES CONVEXAS E CÔNCAVAS)

(a) Dizemos que uma função  $f: U \subset \mathbb{R}^n \rightarrow \mathbb{R}$  definida em um subconjunto convexo  $U$  de  $\mathbb{R}^n$  é *convexa* se, e somente se,

$$f((1-t) \cdot \mathbf{p} + t \cdot \mathbf{q}) \leq (1-t) \cdot f(\mathbf{p}) + t \cdot f(\mathbf{q}), \quad (\text{C.1})$$

para todo  $\mathbf{p}, \mathbf{q} \in U$  e todo  $t \in [0, 1]$ .

A interpretação geométrica de (C.1) é a seguinte: o segmento de reta secante que passa pelos pontos  $(\mathbf{p}, f(\mathbf{p}))$  e  $(\mathbf{q}, f(\mathbf{q}))$  sempre está *acima* ou *coincide* com o gráfico de  $f$  para qualquer escolha de pontos  $\mathbf{p}$  e  $\mathbf{q}$  em  $U$  (veja a Figura C.2).

(b) Dizemos que uma função  $f: U \subset \mathbb{R}^n \rightarrow \mathbb{R}$  definida em um subconjunto convexo  $U$  de  $\mathbb{R}^n$  é *côncava* se, e somente se,

$$f((1-t) \cdot \mathbf{p} + t \cdot \mathbf{q}) \geq (1-t) \cdot f(\mathbf{p}) + t \cdot f(\mathbf{q}), \quad (\text{C.2})$$

para todo  $\mathbf{p}, \mathbf{q} \in U$  e todo  $t \in [0, 1]$ .

A interpretação geométrica de (C.2) é a seguinte: o segmento de reta secante que passa pelos pontos  $(\mathbf{p}, f(\mathbf{p}))$  e  $(\mathbf{q}, f(\mathbf{q}))$  sempre está *abaixo* ou *coincide* com o gráfico de  $f$  para qualquer escolha de pontos  $\mathbf{p}$  e  $\mathbf{q}$  em  $U$ .

O próximo teorema estabelece o motivo de convexidade ser uma propriedade tão desejável em otimização.

**Teorema C.2** Se  $f: U \subset \mathbb{R}^n \rightarrow \mathbb{R}$  é convexa, então todo *ponto de mínimo local* de  $f$  em  $U$  também é *ponto de mínimo global* de  $f$  em  $U$ .

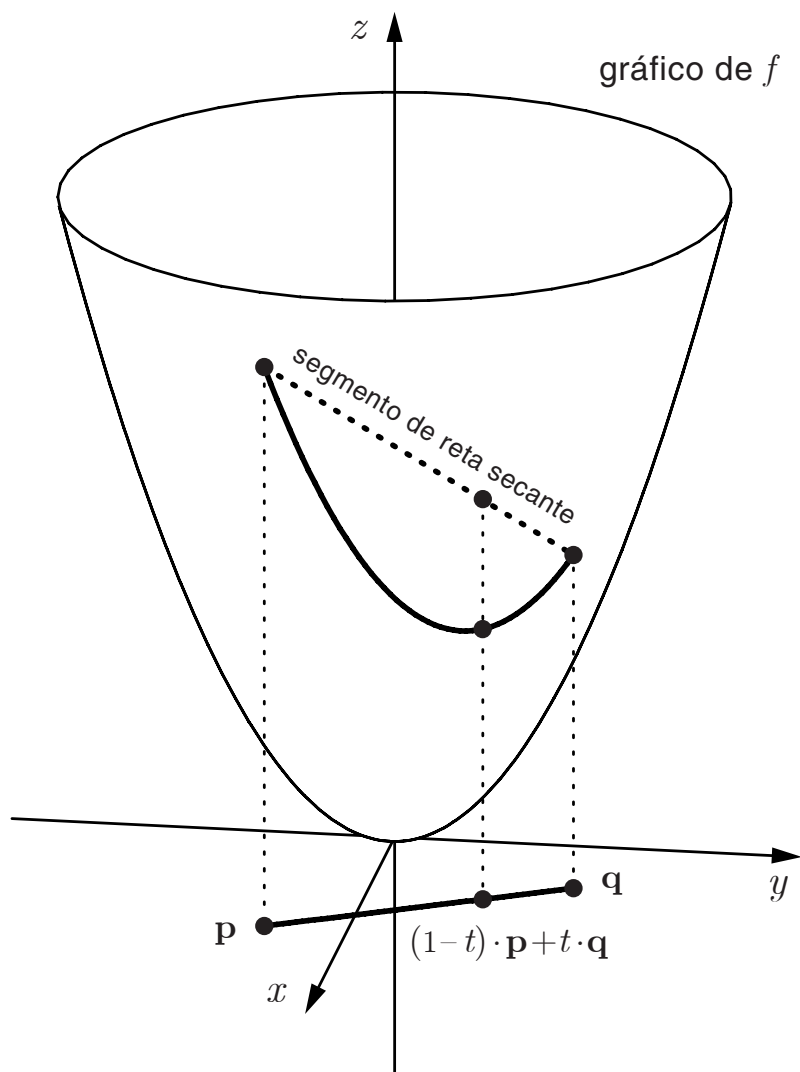


Figura C.2: Para uma função convexa, o segmento de reta secante fica sempre acima ou coincide com o gráfico da função, para quaisquer escolhas de  $\mathbf{p}$  e  $\mathbf{q}$ .

**Teorema C.3** Se  $f_1: U \subset \mathbb{R}^n \rightarrow \mathbb{R}$  e  $f_2: U \subset \mathbb{R}^n \rightarrow \mathbb{R}$  são funções convexas, então:

- (a) A função  $\alpha_1 f_1 + \alpha_2 f_2$  é convexa para todo  $\alpha_1, \alpha_2 \in \mathbb{R}_+$ .
- (b) A função  $\mathbf{x} \mapsto \max\{f_1(\mathbf{x}), f_2(\mathbf{x})\}$  é convexa.



**Teorema C.4** Seja  $f: U \subset \mathbb{R}^n \rightarrow \mathbb{R}$  uma função de classe  $C^1$  definida em um subconjunto convexo  $U$  de  $\mathbb{R}^n$ .

(a)  $f$  é uma função convexa em  $U$  se, e somente se,

$$f(\mathbf{q}) \geq f(\mathbf{p}) + \nabla f(\mathbf{p}) \cdot (\mathbf{q} - \mathbf{p}), \quad (\text{C.3})$$

para todo  $\mathbf{p}, \mathbf{q} \in U$ , isto é, se, e somente se, cada hiperplano tangente ao gráfico de  $f$  está sempre *abaixo* ou *coincide* com o gráfico de  $f$ .

(b)  $f$  é uma função côncava em  $U$  se, e somente se,

$$f(\mathbf{q}) \leq f(\mathbf{p}) + \nabla f(\mathbf{p}) \cdot (\mathbf{q} - \mathbf{p}), \quad (\text{C.4})$$

para todo  $\mathbf{p}, \mathbf{q} \in U$ , isto é, se, e somente se, cada hiperplano tangente ao gráfico de  $f$  está sempre *acima* ou *coincide* com o gráfico de  $f$ .

Aqui  $\nabla f(\mathbf{p})$  denota o vetor gradiente de  $f$  em  $\mathbf{p}$ .

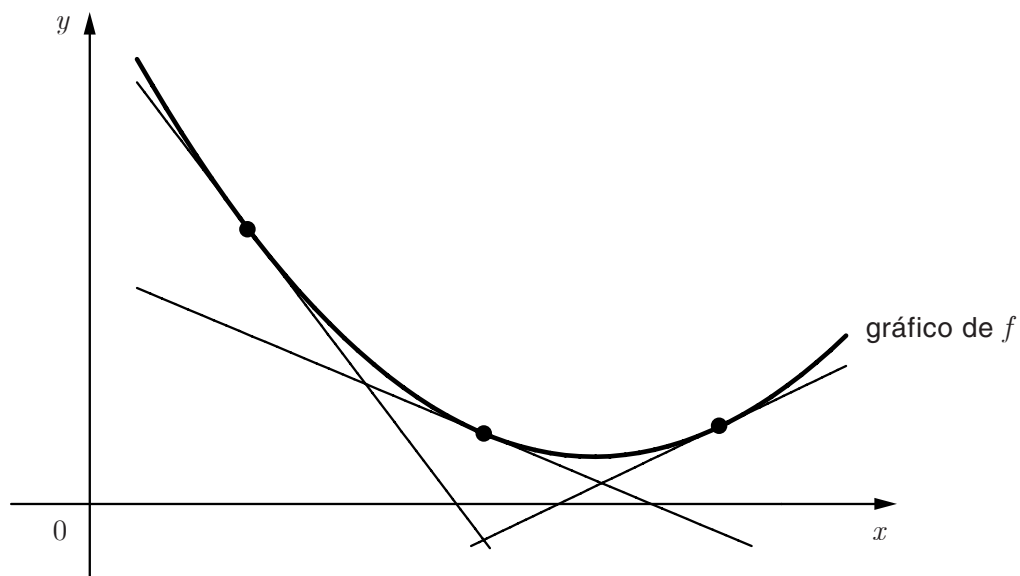


Figura C.3: Para uma função convexa, cada reta tangente ao gráfico de  $f$  está sempre abaixo do gráfico de  $f$ .

**Definição C.3 (Semiplanos e Semi-Espaços)** Seja  $\mathbf{a}$  um vetor não-nulo em  $\mathbb{R}^n$  e seja  $c$  um número real. Os conjuntos

$$\mathbf{H}_+ = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{a}\mathbf{x} \geq c\} \quad \text{e} \quad \mathbf{H}_- = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{a}\mathbf{x} \leq c\}$$

são denominados, respectivamente, *semi-espacos fechados* correspondentes ao *semiplano*  $\mathbf{H} = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{a}\mathbf{x} = c\}$ .

Por linearidade, segue-se que semiplanos e semi-espacos são conjuntos convexos.

**Definição C.4 (Politopos e Poliedros)** Um *politopo* é um conjunto que pode ser expresso como a interseção de um número finito de semi-espacos fechados. Um *poliedro* é um politopo limitado. Note que politopos e poliedros são conjuntos convexos.

# Apêndice D

## Programação Linear

Neste apêndice apresentaremos as definições e propriedades básicas da teoria de programação linear necessárias no texto. Para detalhes, demonstrações e extensões, recomendamos os excelentes livros [05, 06, 07, 16].

Um *programa linear* é um problema de otimização onde a função objetivo e as restrições são todas lineares. Dizemos que um programa linear está na *forma padrão* se todas as variáveis de decisão são não-negativas e se todas as restrições são em igualdade:

minimizar $\in \mathbb{R}$ <small><math>x_1, x_2, \dots, x_n</math></small>	$c_1x_1 + c_2x_2 + \dots + c_nx_n$
sujeito a	$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1,$
	$a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2,$
	$\vdots$
	$a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m,$
	e $x_1 \geq 0, x_2 \geq 0, \dots, x_n \geq 0.$

Todo programa linear pode ser reescrito na forma padrão com o uso de *variáveis de folga*. Por exemplo, uma restrição da forma

$$a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n \leq b_i$$

pode ser substituída, de maneira equivalente, pelas restrições

$$a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n + y_i = b_i \text{ e } y_i \geq 0.$$

Se uma variável de decisão  $x_i$  pode assumir qualquer valor real, isto é, se não existe restrição de não-negatividade em  $x_i$ , então podemos substituir  $x_i$  por  $u_i - v_i$ , a diferença de dois números positivos.

Um programa linear pode ser escrito de forma mais compacta usando-se matrizes e vetores:

$$\boxed{\begin{array}{ll} \text{minimizar} & \mathbf{c}^T \mathbf{x} \\ \text{sujeito a} & \mathbf{Ax} = \mathbf{b} \text{ e } \mathbf{x} \geq \mathbf{0}, \end{array}} \quad (\text{D.1})$$

onde  $\mathbf{x} \in \mathbb{R}^n$ ,  $\mathbf{c} \in \mathbb{R}^n$ ,  $\mathbf{h} \in \mathbb{R}^m$  e  $\mathbf{A}$  é uma matriz  $m \times n$ .

Na teoria de programação linear, assume-se que  $m < n$  (existem mais incógnitas do que restrições em igualdade) e que o posto da matriz  $\mathbf{A}$  é  $m$ , isto é, as  $m$  linhas de  $\mathbf{A}$  são linearmente independentes. Da teoria de Álgebra Linear sabemos, então, que existem  $m$  colunas de  $\mathbf{A}$  que são linearmente independentes. Renomeando-se índices se necessário, podemos assumir que estas colunas sejam as  $m$  primeiras. Isto induz uma decomposição de  $\mathbf{A}$  e de  $\mathbf{x}$ :

$$\mathbf{A} = [ \mathbf{B} \ \mathbf{C} ], \quad \mathbf{x} = \begin{bmatrix} \mathbf{x}_B \\ \mathbf{x}_C \end{bmatrix},$$

onde  $\mathbf{B}$  é uma matriz  $m \times m$  inversível. Como o sistema linear  $\mathbf{Ax} = \mathbf{b}$  é equivalente a  $\mathbf{Bx}_B + \mathbf{Cx}_C = \mathbf{b}$ , segue-se então que existe uma solução  $\mathbf{x}$  de  $\mathbf{Ax} = \mathbf{b}$  na forma

$$\begin{bmatrix} \mathbf{x}_B \\ \mathbf{0} \end{bmatrix}.$$

Esta solução é denominada *solução básica* do sistema linear  $\mathbf{Ax} = \mathbf{b}$  associada à base  $\mathbf{B}$ . As componentes de  $\mathbf{x}_B$  são denominadas *variáveis básicas*.

**Teorema D.1 (Teorema Fundamental da Programação Linear)** Considere um programa linear na forma padrão (D.1), com  $\mathbf{A}$  matriz  $m \times n$  de posto  $m$ .

- (a) Se o programa linear possui um ponto admissível, então ele possui um ponto admissível que é uma solução básica do sistema linear  $\mathbf{Ax} = \mathbf{b}$ .
- (b) Se o programa linear possui um ponto ótimo, então ele possui um ponto ótimo que é uma solução básica do sistema linear  $\mathbf{Ax} = \mathbf{b}$ .

O próximo teorema dá uma interpretação geométrica para pontos admissíveis que são soluções básicas: eles correspondem aos pontos extremos (vértices) do politopo  $\mathbf{K} = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{Ax} = \mathbf{b} \text{ e } \mathbf{x} \geq \mathbf{0}\}$ .

**Definição D.1 (Ponto Extremo)** Dizemos que um ponto  $\mathbf{x}$  em um conjunto convexo  $U$  é *ponto extremo* de  $U$  se não existem dois outros pontos distintos  $\mathbf{x}_1$  e  $\mathbf{x}_2$  em  $U$  tais que  $\mathbf{x} = \alpha \mathbf{x}_1 + (1 - \alpha) \mathbf{x}_2$  para algum  $\alpha$  no intervalo  $(0, 1)$ .

**Teorema D.2 (Equivalência entre Pontos Extremos e Soluções Básicas)** Seja  $\mathbf{A}$  uma matriz  $m \times n$  de posto  $m$ ,  $\mathbf{b}$  um vetor em  $\mathbb{R}^m$  e  $\mathbf{K} = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{Ax} = \mathbf{b} \text{ e } \mathbf{x} \geq \mathbf{0}\}$  o conjunto admissível de (D.1). Então  $\mathbf{x}$  é um ponto extremo de  $\mathbf{K}$  se, e somente se,  $\mathbf{x}$  é um ponto admissível que é solução básica de  $\mathbf{Ax} = \mathbf{b}$ .

Os teoremas (D.1) e (D.2) dizem que, para se resolver o problema (D.1), não é preciso considerar todos os pontos do conjunto admissível  $\mathbf{K}$ : basta procurar pelo ponto ótimo entre os pontos extremos (vértices) de  $\mathbf{K}$ ! O método *simplex* explora esta estrutura para construir um algoritmo muito popular para se resolver (D.1). Outra categoria de métodos que recentemente ganhou bastante popularidade é a classe dos métodos de ponto interior. Não é nosso propósito estudar estes algoritmos aqui. O leitor interessado poderá consultar os livros [05, 06, 07, 16]. O que é preciso se ter em mente é que programas lineares podem ser resolvidos numericamente de maneira muito eficiente nos dias de hoje. É por este motivo que, em otimização estocástica, existe esta quase que obsessiva diretiva de converter tudo o que se pode para programas lineares.

A seguir estabeleceremos resultados sobre dualidade, um conceito fundamental e muito útil em programação linear.

**Definição D.2 (O problema dual)** O problema dual de

$$\begin{array}{ll} \underset{\mathbf{x} \in \mathbb{R}^n}{\text{minimizar}} & \mathbf{c}^T \mathbf{x} \\ \text{sujeito a} & \mathbf{Ax} = \mathbf{b} \text{ e } \mathbf{x} \geq \mathbf{0}, \end{array} \quad (\text{D.1})$$

é o programa linear

$$\begin{array}{ll} \text{maximizar} & \boldsymbol{\lambda}^T \mathbf{b} \\ \lambda \in \mathbb{R}^m & \\ \text{sujeito a} & \mathbf{A}^T \boldsymbol{\lambda} \leq \mathbf{c}, \end{array} \quad (\text{D.2})$$

onde  $\boldsymbol{\lambda}^T \mathbf{b} = \sum_{i=1}^m \lambda_i b_i$ . (D.2) é denominado o *problema dual* de (D.1). Neste contexto, (D.1) é denominado *problema primal*.

**Teorema D.3 (Teorema fraco de dualidade)** Se  $\mathbf{x}$  e  $\boldsymbol{\lambda}$  são admissíveis para os problemas (D.1) e (D.2), respectivamente, então  $\mathbf{c}\mathbf{x} \geq \boldsymbol{\lambda}\mathbf{b}$ .

Este teorema mostra que um ponto admissível para um dos problemas fornece uma cota para o valor da função objetivo do outro problema. Os valores associados com o problema primal são sempre maiores ou iguais aos valores associados com o problema dual. Como corolário, vemos que se um par de pontos admissíveis pode ser encontrado para os problemas primal e dual com valores iguais da função objetivo, então estes pontos são ótimos.

**Teorema D.4** Se um dos problemas (D.1) ou (D.2) tem uma solução ótima finita, então o outro também terá uma solução ótima finita e, neste caso, os valores das respectivas funções objetivo são iguais. Se a função objetivo do problema primal não é limitada inferiormente, então o conjunto admissível do problema dual é vazio e, se a função do objetivo do problema dual não limitada superiormente, então o conjunto admissível do problema primal é vazio.

Por fim, gostaríamos de observar que se a função objetivo do programa linear (D.1) não é limitada inferiormente no conjunto admissível

$$\mathbf{K} = \{ \mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} = \mathbf{b} \text{ e } \mathbf{x} \geq \mathbf{0} \},$$

então existem ponto extremo  $\tilde{\mathbf{x}}$  e raio extremo  $\tilde{\mathbf{r}}$  de  $\mathbf{K}$  tal que o valor da função objetivo de (D.1) em  $\mathbf{x} = \tilde{\mathbf{x}} + t\tilde{\mathbf{r}}$  tende a  $-\infty$  quando  $t$  tende a  $+\infty$ . Em particular,

$$\mathbf{c}^T \tilde{\mathbf{r}} < 0.$$

Dizemos que  $\tilde{\mathbf{r}}$  é um raio de  $\mathbf{K}$  se, e somente se,  $\tilde{\mathbf{r}} \neq \mathbf{0}$  e o conjunto  $\{\mathbf{p} \in \mathbb{R}^n \mid \mathbf{p} = \mathbf{x} + t\tilde{\mathbf{r}} \text{ e } t \geq 0\}$  está contido em  $\mathbf{K}$  para todo  $\mathbf{x} \in \mathbf{K}$ . Um raio  $\tilde{\mathbf{r}}$  de  $\mathbf{K}$  é *extremo*, se não existem outros dois raios  $\mathbf{r}_1$  e  $\mathbf{r}_2$  de  $\mathbf{K}$  (com  $\mathbf{r}_1 \neq t\mathbf{r}_2$  para todo  $t > 0$ ) e um escalar  $s$  no intervalo  $(0, 1)$  tal que  $\tilde{\mathbf{r}} = s\mathbf{r}_1 + (1-s)\mathbf{r}_2$ .

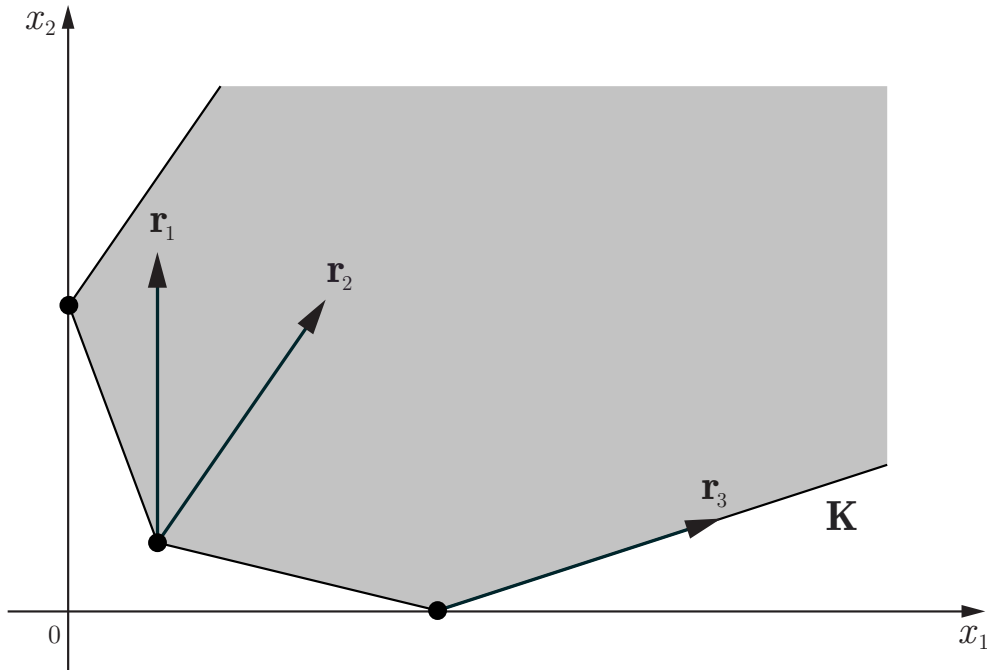


Figura D.1: Os vetores  $\mathbf{r}_1$  e  $\mathbf{r}_2$  não são raios extremos de  $\mathbf{K}$ . O vetor  $\mathbf{r}_3$  é um raio extremo de  $\mathbf{K}$ .

# Bibliografia

- [01] G. Beek, J. Blatt, M. Koschat, N. Kunz, M. LePore e S. Blyakher, *News Vendors Tackle The News Vendor Problem*. Interfaces, v. 33, n. 3, pp. 72–84, 2003.
- [02] D. Bell, *Incorporating The Customer's Perspective into The News vendor Problem*. Preprint, Harvard Business School, 2003.
- [03] J. R. Birge e F. Louveaux, *Introduction to Stochastic Programming*. Springer Series in Operations Research, Springer-Verlag, 1997.
- [04] S. Boyd e L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2005.  
Disponível em: <http://www.stanford.edu/~boyd/cvxbook/>.
- [05] V. Chvátal, *Linear Programming*. W. H. Freeman and Company, 1983.
- [06] G. B. Dantzig e M. N. Thapa, *Linear Programming 1: Introduction*. Springer Series in Operations Research, Springer-Verlag, 1997.
- [07] G. B. Dantzig e M. N. Thapa, *Linear Programming 2: Theory and Extensions*. Springer Series in Operations Research, Springer-Verlag, 2003.
- [08] R. Fourer, D. M. Gay, B. W. Kernighan, *AMPL: A Modeling Language for Mathematical Programming*. Duxbury Press, 2002.
- [09] W. K. K. Haneveld e M. H. van der Vlerk, *Stochastic Programming*. Lecture notes, Departament of Econometrics & OR, University of Groningen, 2004.
- [10] J. Higle e S. Sen, *Stochastic Decomposition: A Statistical Method for Large Scale Stochastic Linear Programming*. Nonconvex Optimization and Its Applications, Kluwer Academic Research, 1996.



- 
- [11] W. Hines e D. Montgomery, *Probability and Statistics in Engineering and Management Science*. John Willey & Sons, 1990.
- [12] P. Hoel, S. Port e C. Stone, *Introduction to Probability Theory*. Houghton Mifflin Company, 1971.
- [13] P. Kall e J. Mayer, *Stochastic Linear Programming: Models, Theory, and Computation*. International Series in Operations Research & Management Science, Springer-Verlag, 2005.
- [14] P. Kall e S. W. Wallace, *Stochastic Programming*. Wiley-Interscience Series in Systems and Optimization, John Willey & Sons, 1995.
- [15] J. Linderoth, A. Shapiro, S. Wright, *The Empirical Behavior of Sampling Methods for Stochastic Programming*. Annals of Operations Research, n. 142, pp. 215–241, 2006.
- [16] D. G. Luenberger, *Linear and Nonlinear Programming*. Addison-Wesley Publishing Company, 1989.
- [17] A. Ruszczyński e A. Shapiro, *Stochastic Programming*. Handbooks in Operations Research and Management Science, vol. 10, Elsevier, 2003.
- [18] R. J.-B. Wets e C. Witzgall, *Algorithms for Frames and Linearity Spaces of Cones*. Journal of Research of the National Bureau of Standards Section, B 71B, pp. 1–7, 1967.