

CLP(X)

Na Programação em Lógica, a operação fundamental que se executa na “passagem de parâmetros” é a unificação de termos de Herbrand.

Com efeito, dado um golo g , uma cláusula $h:- b_1, \dots, b_k$ pode ser usada para o resolver se existir uma substituição σ (na prática a substituição mais geral, *mgu*) tal que

$$h\sigma = g\sigma$$

sendo naturalmente inseridos no resolvente os termos $b_1\sigma, \dots, b_k\sigma$.

Desta forma, a resolução das perguntas (queries) implica a imposição de um conjunto de restrições de igualdade de termos de Herbrand. A questão tem resposta sse for possível satisfazer estas restrições.

A ideia base da Programação em Lógica com Restrições é permitir a utilização de restrições de outros domínios. O esquema CLP(X) é a adaptação de programação em lógica para o domínio X.

CLP(X)

Naturalmente, para que essa extensão se possa fazer é necessário dispor de procedimentos de unificação, i.e. que determinam inequivocamente se dois termos podem ser unificados.

Isso distingue os esquemas com resolvedores completos (como PL) e incompletos.

Nos resolvedores completos, a cada restrição uma memória de restrições (***constraint store***) vai sendo adaptada sse for possível satisfazer as restrições. Esse é o caso da unificação de termos de Herbrand.

Nos resolvedores incompletos, a memória de restrições vai guardando as restrições de uma qualquer forma adequada, mesmo que se venha a provar impossível satisfazer as restrições.

CLP(X)

Como exemplo de resolvedores incompletos temos CLP(FD). Com efeito, dadas as restrições

$$[A, B, C] \text{ in } 0..1, A \neq B, A \neq C, B \neq C$$

O sistema não detecta a impossibilidade de satisfazer estas restrições em domínios finitos (2 valores para 3 variáveis diferentes).

Como resolvedores completos temos CLP(Q) e CLP(B), sendo

- B o domínio de restrições Booleanas
- Q o domínio de expressões lineares com coeficientes racionais.

Vamos centrar a nossa atenção em CLP(Q) que permite verificar que, por exemplo

- $X+Y = 5$ e $X-Y = 1$ tem solução única $X = 3$ e $Y = 2$.
- $X \geq Y$, $Y > Z$ e $X < Z$ não tem solução!

CLP(Q)

No primeiro caso, um resolvidor de restrições vai mantendo as restrições na forma

$$X + Y = 5 \qquad \{X / 5 - Y\}$$

$$X - Y = 1 \qquad (5 - Y) - Y = 1 \qquad \{Y / 2, X / 3\}$$

No segundo caso temos

$$\bullet \ X \geq Y, \qquad \{X / Y + S1, S1 \geq 0\}$$

$$\bullet \ Y > Z \qquad \{Z / Y - S2, S2 > 0\}$$

$$\bullet \ X < Z \qquad Y + S1 < Y - S2$$

$$S1 + S2 < 0, S1 \geq 0, S2 > 0, \text{impossível !}$$

Naturalmente, existem casos mais complicados, que requerem um tratamento geral de restrições lineares.

Formalização

Os problemas de **satisfação** de restrições lineares sobre variáveis (de decisão) X_i reais (ou racionais, se todos os parâmetros a_{ij} são números racionais) têm a forma

$$a_{11} X_1 + a_{12} X_2 + \dots + a_{1n} X_n \rho_1 b_1$$

$$a_{21} X_1 + a_{22} X_2 + \dots + a_{2n} X_n \rho_2 b_2$$

...

$$a_{m1} X_1 + a_{m2} X_2 + \dots + a_{mn} X_n \rho_m b_m$$

em que $\rho_1 \dots \rho_m$ são relações do conjunto $\{\leq, =, \geq, <, >, \neq\}$

Se se pretender a **otimização** das variáveis de decisão, inclui-se a otimização de uma função (linear) objectivo F

$$\text{Opt } F = c_1 X_1 + c_2 X_2 + \dots + c_n X_n$$

em que Opt pode ser um de {Max, Min, Sup, Inf}.

Interpretação Geométrica

Dado um espaço com n dimensões, a restrição

$$a_{i1} X_1 + a_{i2} X_2 + \dots + a_{in} X_n = b_i$$

define uma região admissível, correspondente aos pontos de um **hiper-plano** desse espaço.

Como casos particulares temos

- um espaço tridimensional ($n=3$), em que o hiper-plano corresponde ao **plano** usual, e
- um espaço bidimensional (ou vulgar plano, $n=2$) em que o hiper-plano se reduz a uma **reta**.

A restrição $\neq$ define uma região admissível que corresponde a todos os pontos do espaço n -dimensional, excepto os pontos do hiper-plano

Interpretação Geométrica

As restrições $\leq$ e $\geq$ definem regiões admissíveis do tipo **semi-hiper-espaço** limitado pelo correspondente hiper-plano, incluído na região admissível.

Com $n=3$ temos um **semi-espaço** e com $n=2$ um **semi-plano**.

As restrições $<$ e $>$ definem regiões admissíveis semelhantes, mas excluindo a fronteira.

O conjunto de restrições define, num espaço a n dimensões, um **hiper-poliedro** (**poliedro** com $n=3$ e **polígono** com $n=2$) correspondente à intersecção das m regiões admissíveis ($m+n$ com não negatividade).

Algumas faces do hiper-poliedro pertencem ($\leq$, $\geq$) à região admissível e outras não ($<$, $>$). Devem ainda ser considerados hiper-planos de exclusão ($\neq$).

Interpretação Geométrica

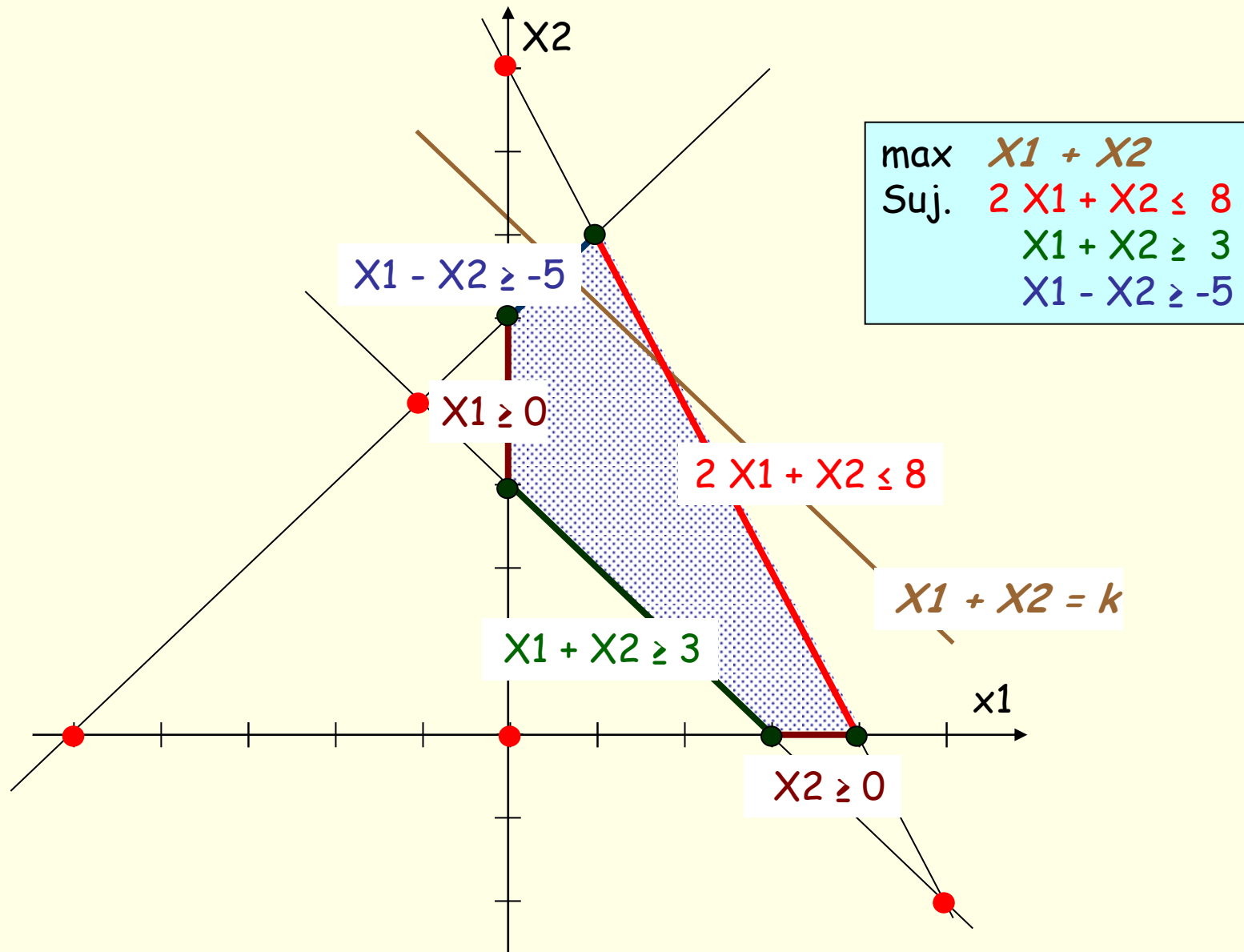
- Com exclusão das restrições $\neq$, a região admissível é **convexa**.
- Em problemas de otimização a função

$$C = c_1 x_1 + c_2 x_2 + \dots + c_n x_n$$

define uma família de hiper-planos paralelos (1 para cada valor de C).

- O supremo e o ínfimo dessa função na região admissível corresponde a um **(hiper-)vértice** do hiper-poliedro, o “último” ponto em que um hiper-plano da função objectivo “toca” a região admissível quando se desloca para valores crescentes (Sup) ou decrescentes (Inf) de C.

Exemplo (2 dimensões)



Complexidade Potencial

- Apesar de o espaço de pesquisa ser infinito, o problema de satisfação reduz-se à verificação de que as restrições definem um hiper-poliedro não vazio.
- Assim, pelo menos um dos vértices definidos pelos hiper-planos de fronteira das restrições (excepto restrições $\neq$) deve pertencer à região admissível.
- Como n restrições de igualdade (independentes) num espaço n -dimensional definem univocamente um ponto (solução de um sistema de n equações a n incógnitas), no pior caso, ter-se-á de verificar a pertença à região admissível de C_n^m pontos.
- Na prática, o número de vértices a testar é muito inferior.
- No pior caso, a complexidade do problema de satisfação é idêntica a um problema de otimização.

Formas Resolvidas

- Há pois que determinar uma forma expedita de representar os vértices, e pesquisem eficientemente se pertencem à região admissível.
- Em particular, estaremos interessados em algoritmos que
 - Associem vértices a uma **Forma Resolvida**,
 - Implementem a pesquisa através de um conjunto de manipulações algébricas.
- A associação deverá ser possível se e apenas se o conjunto de restrições inicial for satisfazível.
- Podemos abordar estas formas resolvidas para sistemas de restrições com riqueza de expressão crescente.

Restrições Não-Estritas / Variáveis Não-negativas

- As restrições de desigualdade podem ser substituídas por restrições de igualdade através da adição de variáveis de desvio (*slacks*), igualmente não negativas

$$a_{i1} x_1 + \dots + a_{in} x_n \leq b_i \Rightarrow a_{i1} x_1 + \dots + a_{in} x_n + s_i = b_i$$

$$a_{i1} x_1 + \dots + a_{in} x_n \geq b_i \Rightarrow a_{i1} x_1 + \dots + a_{in} x_n - s_i = b_i$$

- A existência de uma forma resolvida para estes sistemas de restrições é justificada pelo seguinte

Lema 1: Todo o sistema de m equações a $m+n$ incógnitas não negativas, é satisfazível sse admitir uma solução em que n variáveis são nulas.

Forma Resolvida SF0

Definição: Dado um sistema de equações de m equações a $m+n$ incógnitas (não negativas)

$$(S1) \quad a_{i,1} x_1 + \dots + a_{i,m+n} x_{m+n} = b_i \quad (i: 1..m)$$

a sua forma resolvida SF0 tem a forma

$$(SF0) \quad x_i = d_i + c_{i,m+1} x_{m+1} + \dots + c_{i,m+n} x_{m+n}$$

$$\text{sendo } d_i \geq 0 \quad i: 1..m$$

As variáveis no lado esquerdo são as variáveis básicas, sendo as outras as não básicas (as variáveis básicas podem ser quaisquer e não as primeiras m variáveis como são apresentadas para simplificar a notação).

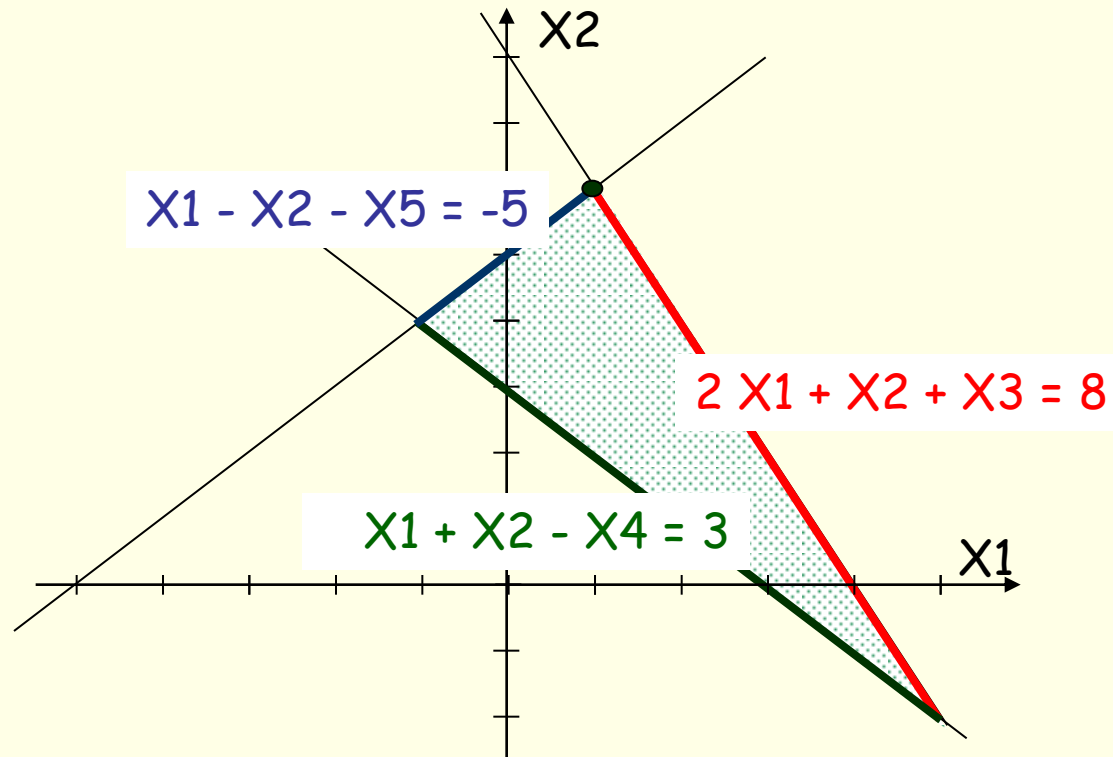
O lema anterior permite justificar o seguinte teorema.

Teorema: Um sistema de equações de m equações a $m+n$ variáveis não-negativas é satisfazível sse se puder reescrever na forma SF0.

Interpretação Geométrica

Exemplo:

Dadas **R3**: $2X_1 + X_2 \leq 8$; **R4**: $X_1 + X_2 \geq 3$; **R5**: $X_1 - X_2 \geq -5$, a região admissível é constituída pela interseção das sub-regiões definidas por cada uma das restrições.

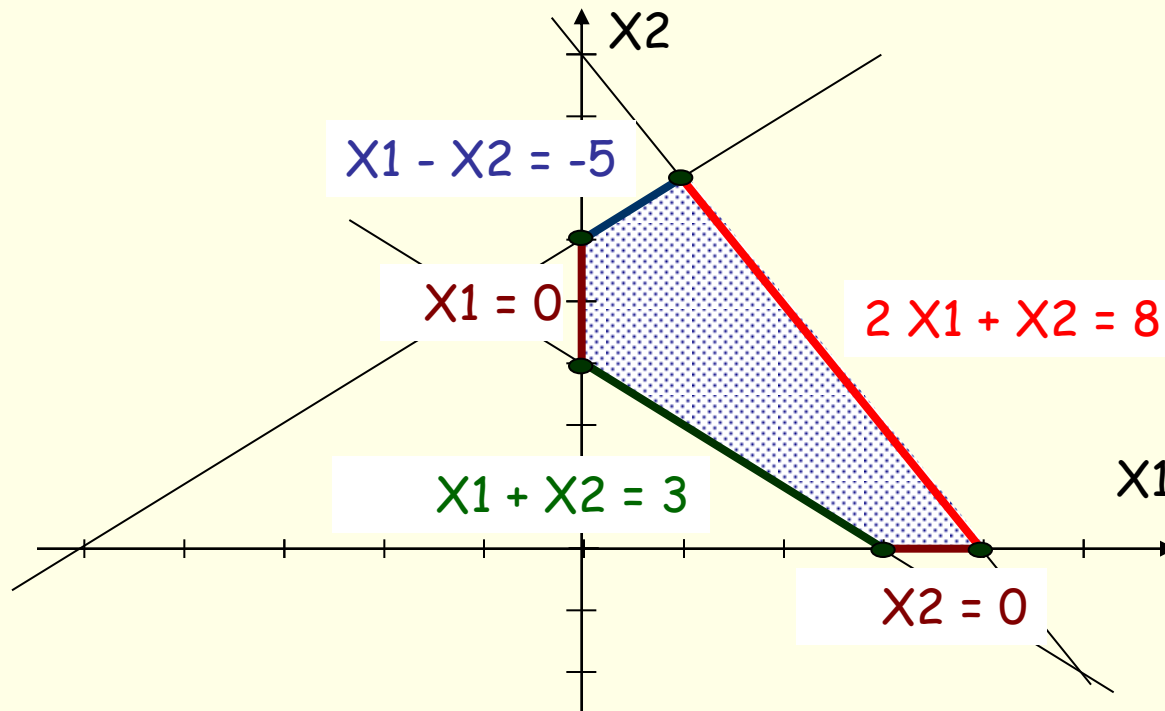


Interpretação Geométrica

Na realidade, pretendendo-se a não-negatividade de X_1 e X_2 , deverão ser consideradas mais 2 restrições

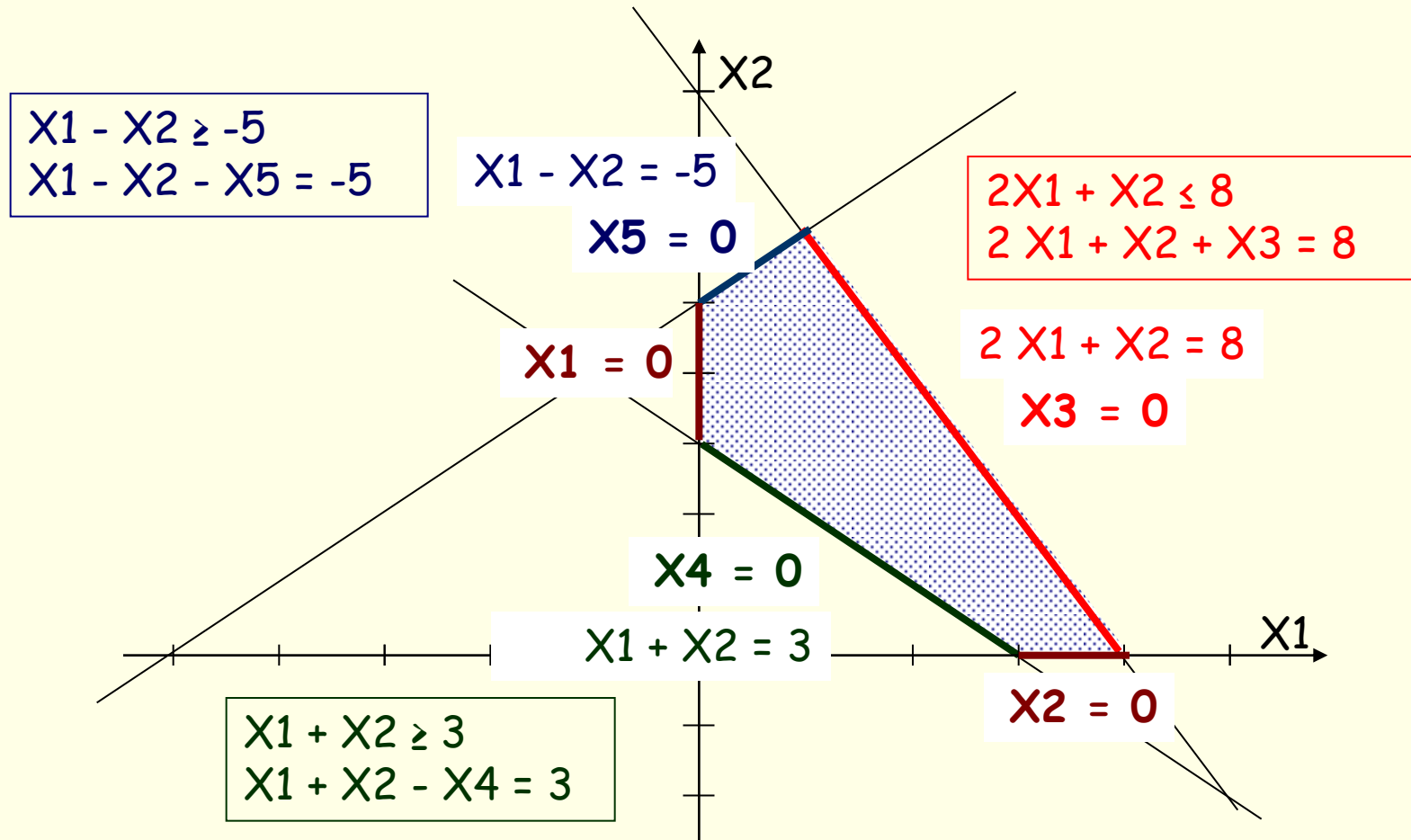
$$R1: X_1 \geq 0 \quad \text{e} \quad R2: X_2 \geq 0$$

A região admissível é constituída pela interseção das sub-regiões definidas pelas 5 restrições $\{R1, R2, R3, R4, R5\}$.



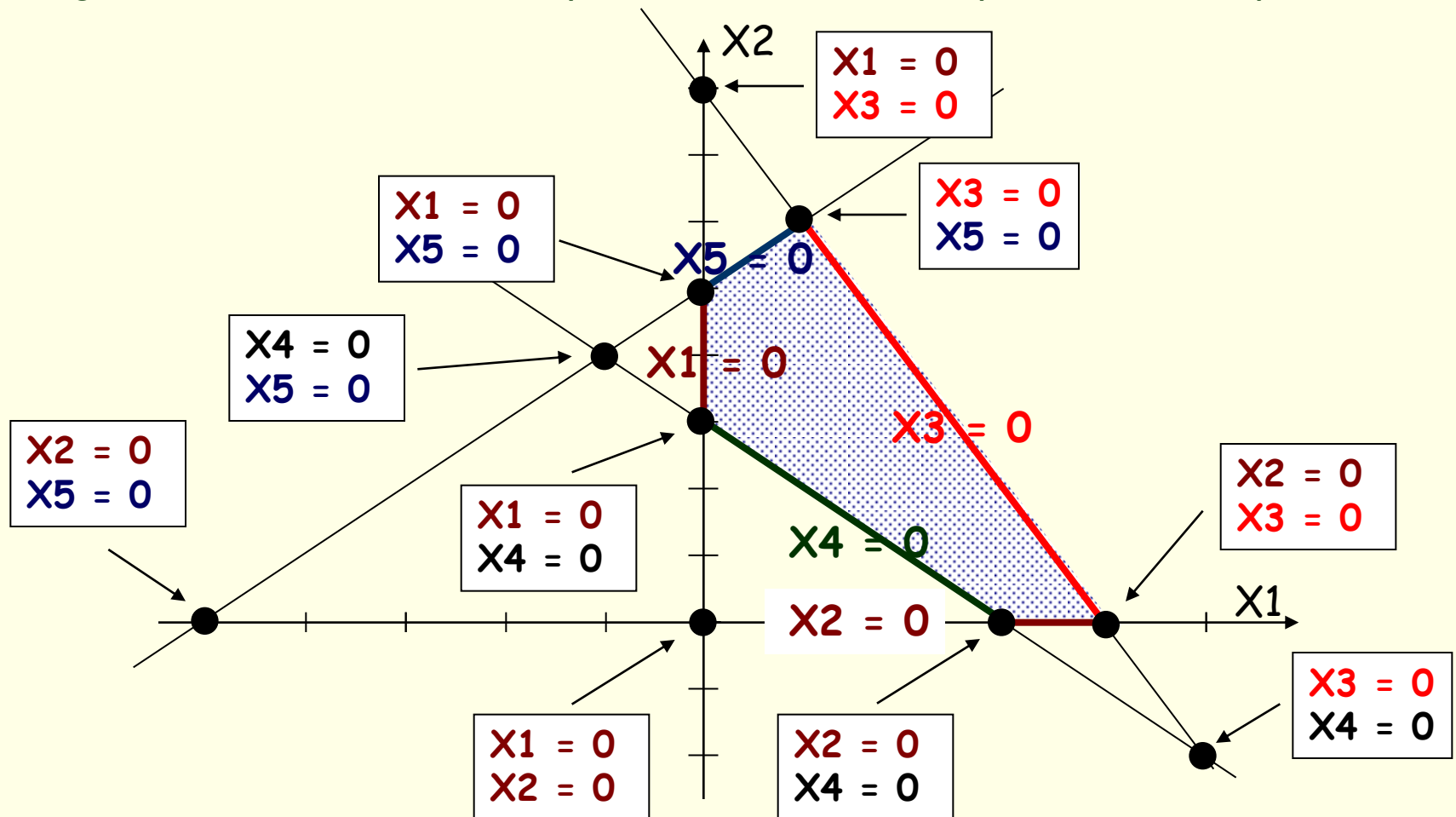
Interpretação Geométrica

Cada uma das retas correspondentes a uma restrição pode ser identificada pelo anulamento da correspondente variável de desvio.



Interpretação Geométrica

A interseção de quaisquer duas retas, define um **potencial** vértice da região admissível, e é obtida pelo anulamento de 2 (das 5 variáveis).



Interpretação Geométrica

Neste caso temos um sistema de 3 igualdades a 5 variáveis (2 variáveis de decisão, X_1 e X_2 , e 3 variáveis de desvio (X_3 , X_4 e X_5)).

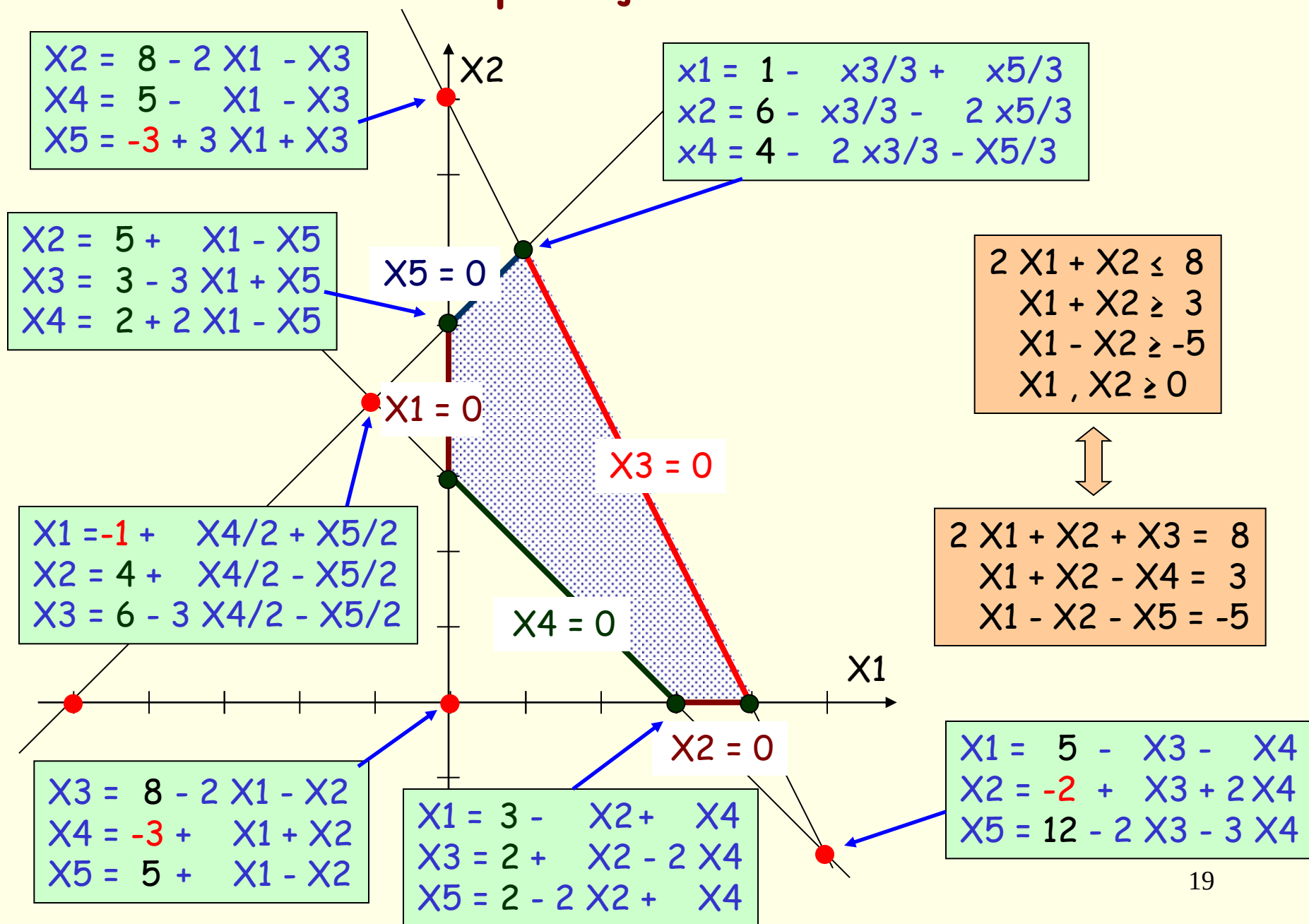
$2 X_1 + X_2 \leq 8$	$\iff$	$2 X_1 + X_2 + X_3 = 8$
$X_1 + X_2 \geq 3$		$X_1 + X_2 - X_4 = 3$
$X_1 - X_2 \geq -5$		$X_1 - X_2 - X_5 = -5$
$X_1, X_2 \geq 0$		

Existem naturalmente 10 combinações de variáveis 2 a 2. O anulamento de quaisquer 2 variáveis, corresponde a um vértice **potencial** da região admissível.

Esse anulamento corresponde igualmente a uma escolha de variáveis não básicas, e a uma **potencial** forma SF0 do sistema inicial.

Apenas as formas SF0 potenciais em que os coeficientes livres são não negativos correspondem a **formas SF0 reais** (e reais vértices da região admissível).

Interpretação Geométrica



Interpretação Geométrica

A existência de soluções para o conjunto de restrições equivale a que a região admissível seja não vazia. Um problema em que nenhum vértice seja admissível é um problema sem solução.

Por exemplo, se nas restrições anteriores se trocarem o sinal das restrições R3 e R4, o novo sistema

$$R3' : 2x_1 + x_2 \geq 8$$

$$R3' : 2x_1 + x_2 - x_3 = 8$$

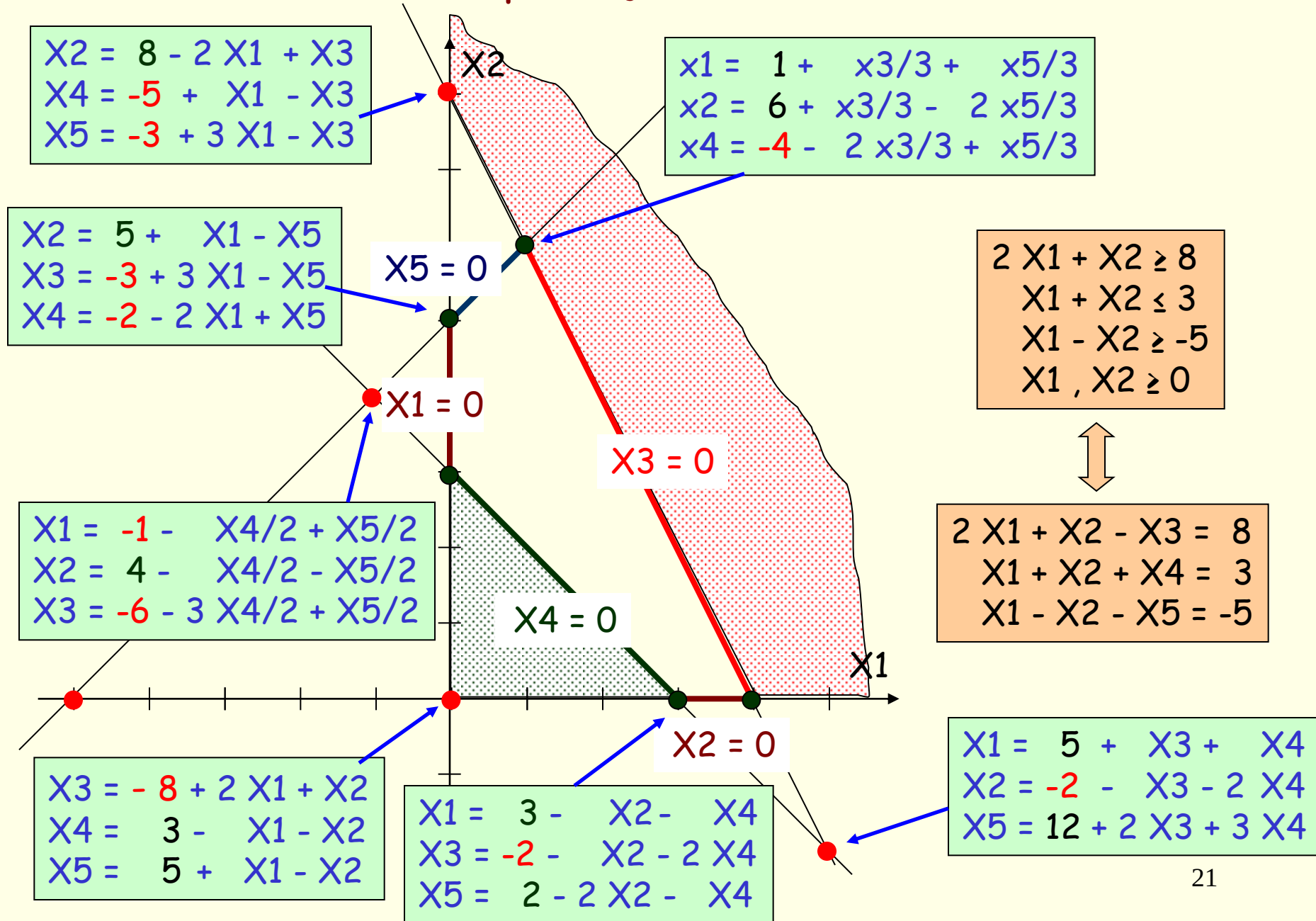
$$R4' : x_1 + x_2 \leq 3 \quad \text{ou} \quad R4' : x_1 + x_2 + x_4 = 3$$

$$R5 : x_1 - x_2 \geq -5$$

$$R5 : x_1 - x_2 - x_5 = -5$$

corresponde a um problema que é impossível e não tem qualquer solução.

Interpretação Geométrica



Passagem para a Forma Resolvida SF0

Para avaliar a satisfazibilidade de um problema basta pois verificar que pelo menos um dos potenciais vértices corresponde a uma solução.

Mas isso corresponde a colocar o sistema de restrições na forma SF0. Por exemplo, o vértice admissível

$$x_1 = 1 - x_3/3 + x_5/3$$

$$x_2 = 6 - x_3/3 - 2 x_5/3$$

$$x_4 = 4 - 2 x_3/3 - x_5/3$$

Corresponde à reescrita na forma SF0 do sistema inicial

$$2 x_1 + x_2 \geq 8$$

$$x_1 + x_2 \leq 3$$

$$x_1 - x_2 \geq -5$$

ou

$$2 x_1 + x_2 - x_3 = 8$$

$$x_1 + x_2 + x_4 = 3$$

$$x_1 - x_2 - x_5 = -5$$

Passagem para a Forma Resolvida SF0

Para colocar um sistema de m restrições de igualdade a $m+n$ variáveis (possivelmente proveniente de um sistema de m restrições de desigualdade $\leq$ ou $\geq$) na forma SF0, basta pois

1. Escolher m variáveis como variáveis básicas
2. Reescrever o sistema nessa base
3. Verificar se os coeficientes livres (valor das variáveis básicas quando as não básicas se anulam) são não negativos

O problema desta abordagem é a existência de um número muito elevado (combinações de $m+n$, n a n , $\binom{m+n}{n}$) de possibilidades de escolha da base.

A escolha é trivial no caso em que todas as restrições são do tipo

$$a_{i1} x_1 + \dots + a_{in} x_n \leq b_i \quad \text{com } b_i \geq 0$$

em que as variáveis não básicas são as de decisão (x_1 a x_n) já que a solução $x_i = 0$ ($1 \leq i \leq n$) é admissível (pois $0 \leq b_i$).

Conversão em SF0 - Simplex

No caso mais geral em que as restrições são do tipo $\geq$ ou $\leq$ coloca-se a questão de garantir a passagem eficiente de um conjunto de restrições lineares sobre variáveis não negativas para uma forma resolvida SF0, se existir, sem escolha arbitrária dos vértices (variáveis básicas).

Por outro lado, se se pretender a integração dum resolvidor de restrições simbólicos num sistema de programação em lógica (CLP) esse resolvidor deve ser **incremental**.

Há pois que definir um procedimento eficiente e incremental para efetuar essa transformação.

Na prática, há que definir uma estratégia para escolher eficientemente as variáveis básicas e as não básicas.

Uma solução possível, geralmente utilizada, é essencialmente baseada no algoritmo **SIMPLEX**.

Otimização em SF0

Há pois que definir uma estratégia para determinar quais as variáveis que devem entrar na base.

Essa estratégia é fácil de entender se se pretender não apenas verificar a satisfação de um conjunto de restrições, mas ainda a otimização de uma função objectivo que, tal como as restrições, seja linear.

Assim, assumamos que, dado um conjunto de m restrições de igualdade nas variáveis X_1 a X_{m+n} , **já colocado na forma SF0**, se pretende adicionalmente maximizar (o caso da minimização é semelhante) uma função

$$\mathbf{max \ F = c_1 \ X_1 + c_2 \ X_2 + \dots + C_{m+n} \ X_{m+n}}$$

Otimização em SF0

$$\max F = c_1 X_1 + c_2 X_2 + \dots + c_{m+n} X_{m+n}$$

Sem perda de generalidade, consideremos que as primeiras m variáveis X_1 a X_m são as variáveis básicas, isto é, que as restrições foram reescritas na forma SF0

$$X_1 = d_1 + c_{11} X_{m+1} + \dots + c_{1n} X_{m+n}$$

...

$$X_m = d_m + c_{m1} X_{m+1} + \dots + c_{mn} X_{m+n}$$

Substituindo na função F as variáveis básicas pelas suas expressões nas não básicas obtemos

$$\max F = k + k_1 X_{m+1} + k_2 X_{m+2} + \dots + k_n X_{m+n}$$

Otimização em SFO

$$\max F = k + k_1 X_{m+1} + k_2 X_{m+2} + \dots + k_n X_{m+n}$$

Esta expressão mostra que no vértice definido pelas variáveis básicas (com as não-básicas nulas) a função a maximizar tem o valor k .

Mostra ainda que, se todos os coeficientes k_i forem negativos, não se pode obter maiores valores de F .

No caso de haver coeficientes k_i positivos, um aumento da variável correspondente aumenta o valor da função F .

Quanto maior o coeficiente, maior o aumento da função objectivo por aumento unitário da variável.

Otimização em SF0

$$\max F = k + k_1 X_{m+1} + k_2 X_{m+2} + \dots + k_n X_{m+n}$$

Numa perspetiva de otimização, dada uma certa base admissível, para a qual existam valores k_i positivos, o valor da função objectivo pode ser melhorado (ser maior que k) tornando positivas as correspondentes variáveis não básicas.

Mas para mantermos o sistema na forma SF0, se se “desanula” uma variável não-básica, essa situação deverá ser compensada anulando uma das variáveis básicas.

Assim, a estratégia de otimização a seguir é a de, partindo de uma base admissível, proceder a um conjunto de mudanças de base até se poder reescrever a função objetivo com coeficientes exclusivamente não positivos.

Otimização em SFO

Obviamente, uma mudança de base, envolve a escolha de

- a) uma variável não básica para entrar na base, e
- b) uma variável básica para sair da base,

A escolha da variável de entrada pode seguir a heurística delineada atrás:

A variável de entrada é aquela a que corresponde na função objetivo o coeficiente mais positivo (maximização) ou mais negativo (minimização).

Uma vez escolhida a variável de entrada, ela deverá aumentar tanto quanto possível. No entanto, na forma SFO corrente, um aumento dessa variável poderá conduzir à diminuição das variáveis básicas.

Otimização em SF0

Com efeito, tome-se a forma SF0 abaixo e, sem perda de generalidade, considere-se X_{m+1} como variável de entrada.

$$\begin{array}{ll}\max & F = k + k_1 X_{m+1} + k_2 X_{m+2} + \dots + k_n X_{m+n} \\ \text{su}j. & X_i = d_i + c_{i1} X_{m+1} + \dots + c_{in} X_{m+n} \quad i:1..m\end{array}$$

Mantendo as outras variáveis não básicas nulas, as equações reescrevem-se como

$$X_i = d_i + c_{i1} X_{m+1} \quad i:1..m$$

Assim, sendo $d_i \geq 0$, um aumento de X_{m+1} anula as variáveis básicas (em que $c_{i1} < 0$), quando tomar o valor $d_i/|c_{i1}|$.

Como se pretende anular uma variável básica, mantendo as outras não negativas, a escolha da variável de saída recai na variável básica X_i para a qual seja **menor** o valor $d_i/|c_{i1}|$.

Otimização em SF0

Resumindo, a maximização da função objetivo de um sistema de m igualdades a $m+n$ variáveis envolve a escrita do sistema e da função objetivo na forma SF0

$$\max \quad F = k + k_1 X_{m+1} + k_2 X_{m+2} + \dots + k_n X_{m+n}$$

$$\text{su}j. \quad X_i = d_i + c_{i1} X_{m+1} + \dots + c_{in} X_{m+n} \quad i:1..m$$

e a execução do seguinte algoritmo:

- ❖ Enquanto a função objetivo tiver coeficientes positivos, mudar a base escolhendo
 - como **variável de entrada** a variável não básica X_j com coeficiente k_j mais positivo na função objetivo
 - como **variável de saída** a variável básica i com $c_{ij} < 0$ e menor valor $d_i/|c_{ij}|$ na respectiva restrição

Interpretação Geométrica

Do ponto de vista geométrico, e como visto atrás, cada forma SF0 corresponde a um vértice da região admissível.

Assim, ao trocar uma variável básica por outra não básica este algoritmo vai percorrendo vários vértices da região admissível, com valores crescentes da função objetivo.

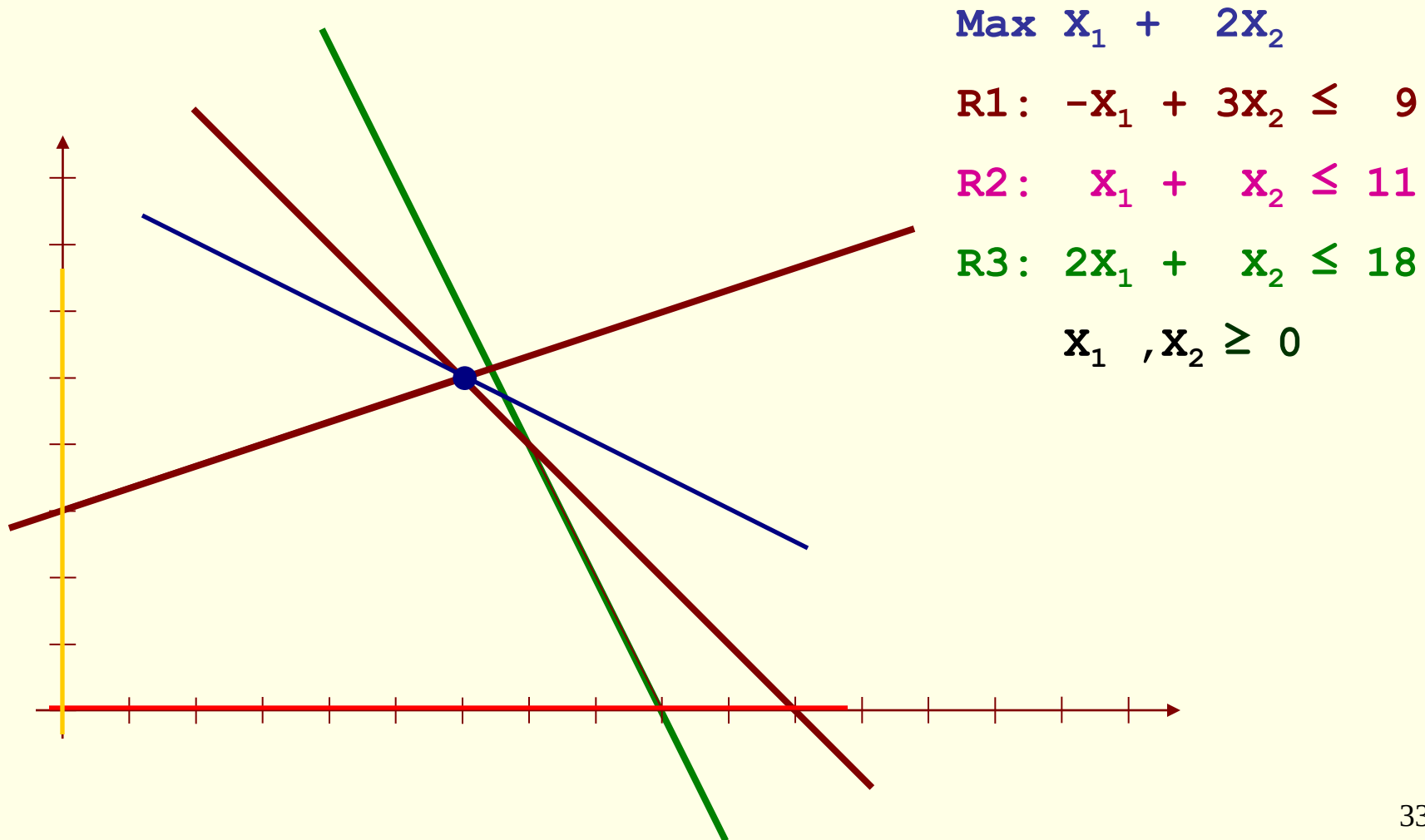
Por apenas trocarem uma variável, dois vértices consecutivos estão unidos por uma aresta do hiper-poliedro que representa a região admissível.

Vamos ilustrar este processo com restrições do tipo $\sum a_i x_i \leq b_i$, ($b_i \geq 0$), que garantem que o vértice $x_1 = x_2 = 0$ está na região admissível.

Uma forma SF0 inicial pode pois ter x_1 e x_2 como variáveis não básicas.

Interpretação Geométrica

Exemplo.



Interpretação Geométrica

Exemplo.

$$\text{Max } X_1 + 2X_2$$

$$-X_1 + 3X_2 + X_3 = 9$$

$$X_1 + X_2 + X_4 = 11$$

$$2X_1 + X_2 + X_5 = 18$$

$$X_1, X_2 \geq 0$$

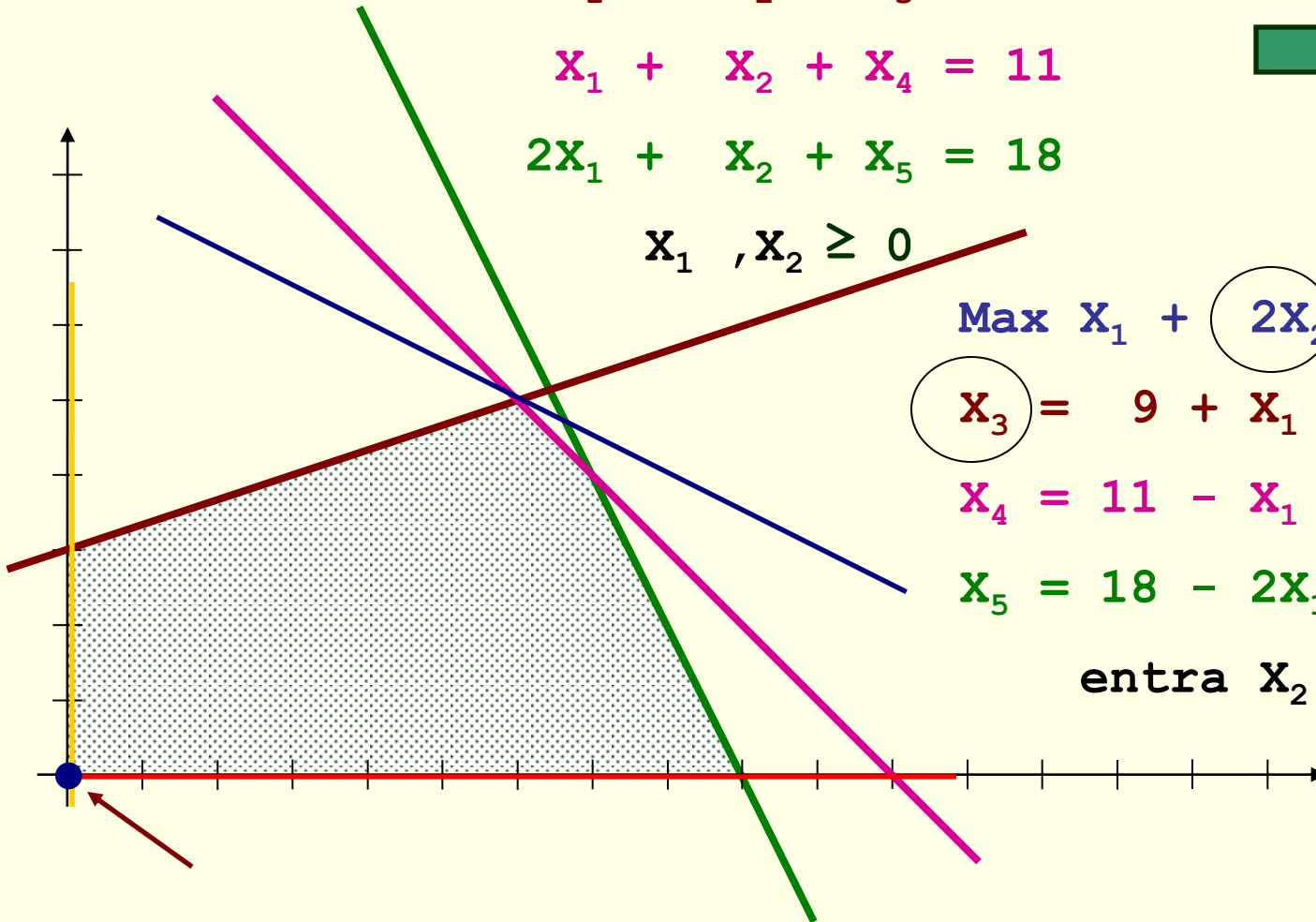
$$\text{Max } X_1 + 2X_2$$

$$X_3 = 9 + X_1 - 3X_2 \rightarrow 9/3$$

$$X_4 = 11 - X_1 - X_2 \rightarrow 11/1$$

$$X_5 = 18 - 2X_1 - X_2 \rightarrow 18/1$$

entra X_2 , sai X_3



Interpretação Geométrica

Exemplo.

$$\text{Max } X_1 + 2X_2$$

$$X_3 = 9 + X_1 - 3X_2 \rightarrow 9/3$$

$$X_4 = 11 - X_1 - X_2 \rightarrow 11/1$$

$$X_5 = 18 - 2X_1 - X_2 \rightarrow 18/1$$

entra X_2 , sai X_3

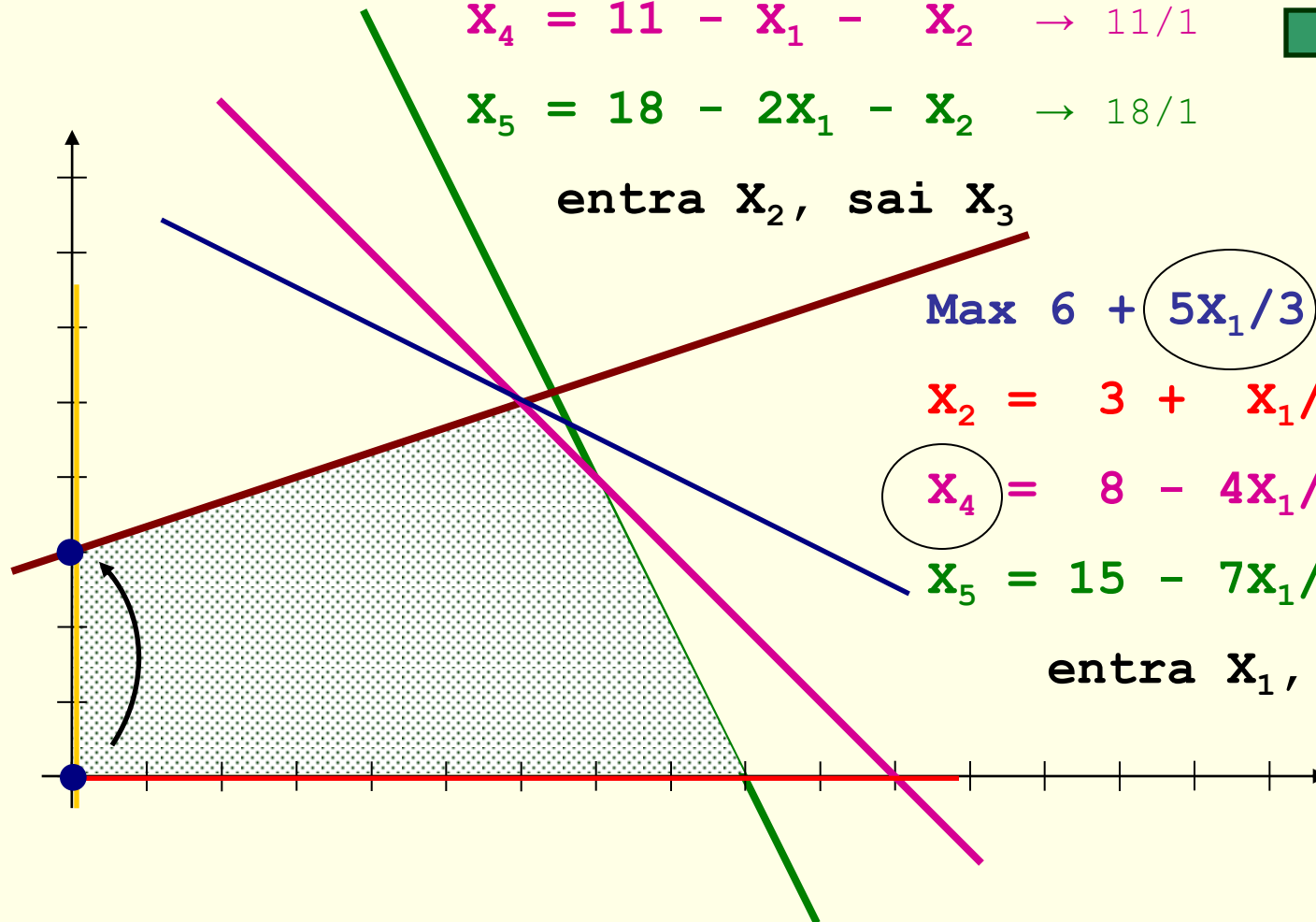
$$\text{Max } 6 + \frac{5X_1}{3} - \frac{2X_3}{3}$$

$$X_2 = 3 + \frac{X_1}{3} - \frac{X_3}{3} \rightarrow 9$$

$$X_4 = 8 - \frac{4X_1}{3} + \frac{X_3}{3} \rightarrow 6$$

$$X_5 = 15 - \frac{7X_1}{3} + \frac{X_3}{3} \rightarrow 45/7$$

entra X_1 , sai X_4



Interpretação Geométrica

Exemplo.

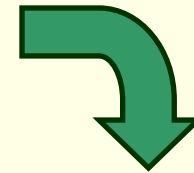
$$\text{Max } 6 + 5x_1/3 - 2x_3/3$$

$$x_2 = 3 + x_1/3 - x_3/3$$

$$x_4 = 8 - 4x_1/3 + x_3/3$$

$$x_5 = 15 - 7x_1/3 + x_3/3$$

entra x_1 , sai x_4



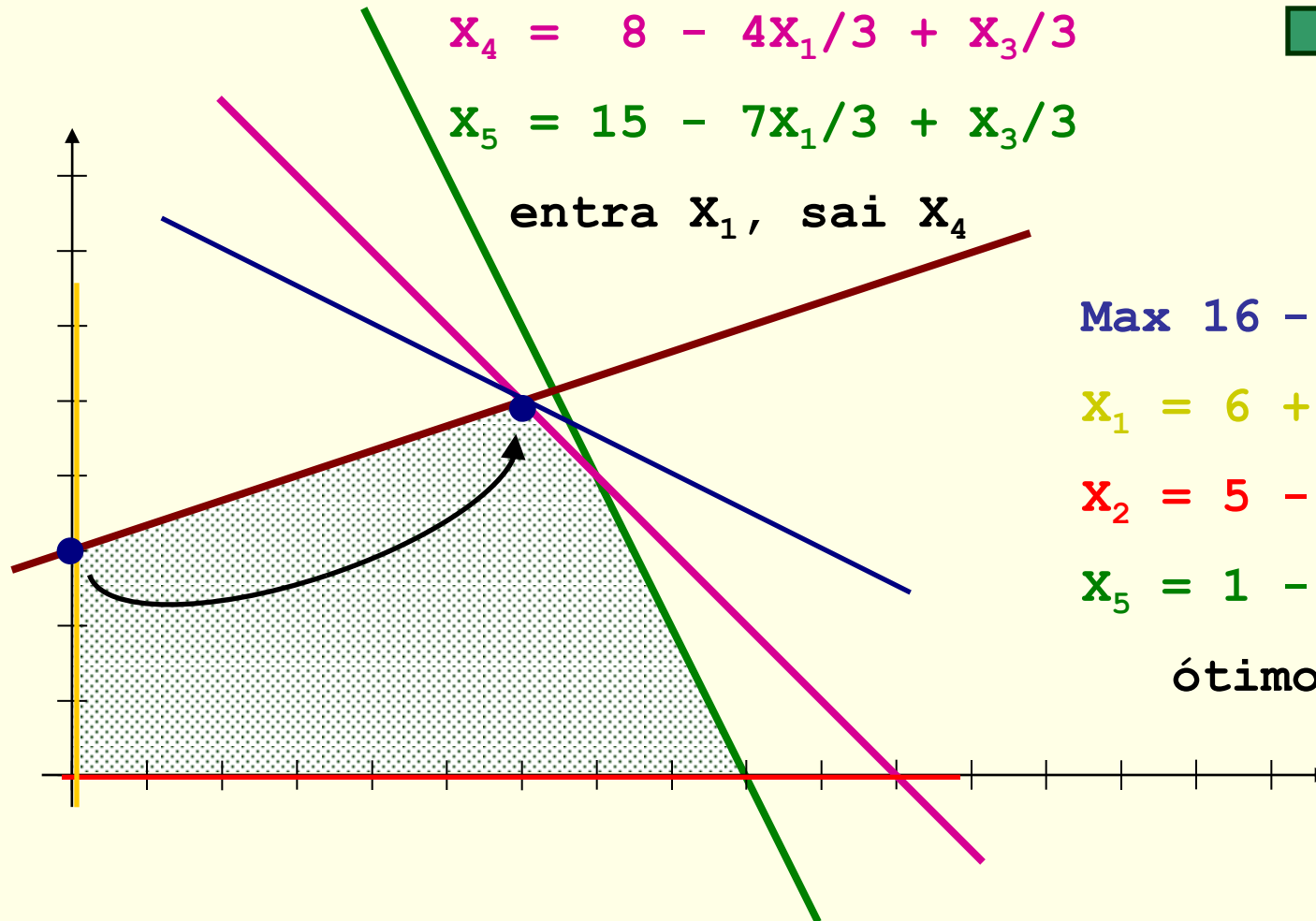
$$\text{Max } 16 - x_3/4 - 5x_4/4$$

$$x_1 = 6 + x_3/4 - 3x_4/4$$

$$x_2 = 5 - x_3/4 - x_4/4$$

$$x_5 = 1 - x_3/4 - 7x_4/4$$

ótimo encontrado



CLP(Q/R) no SICStus

```
:- use_module(library(clpq)) .
```

```
:- use_module(library(clpr)) .
```

`{+Constraint}` (com `=`, `=\=`, `>`, `=<`, ...)

`minimize/maximize(+Expr)`

`inf/sup/2, 4`: ínfimo e supremo duma expressão

- Ver manual...