Presentation outline

1. Métodos numéricos para optimização não linear sem restrições
   - Métodos de procura directa
   - Métodos de procura unidireccional (geral)
   - Métodos de procura unidireccional (descida máxima, Newton)
   - Métodos de região de confiança
2. Teoria da optimização não linear com restrições
   - Introdução e qualificações de restrições
   - Condições de primeira ordem
   - Condições de segunda ordem
   - Teoria da dualidade
3. Métodos numéricos para optimização não linear com restrições
   - Método da penalização quadrática
   - Método do Lagrangeano aumentado
   - Programação Sequencial Quadrática (SQP)
   - Métodos de pontos interiores

Os métodos de penalização para **optimização com restrições** resolvem o problema original por uma sucessão de subproblemas.

Cada subproblema envolve a minimização de uma função de penalização.

A mais simples **função de penalização** é a **função de penalização quadrática**.

Consideremos um PNL com $\mathcal{I} = \emptyset$

$$\boxed{\begin{array}{ll} \displaystyle\min_{x \in \mathbb{R}^n} & f(x) \\ \text{sujeito a} & c_i(x) \; = \; 0, \quad i \in \mathcal{E} \end{array}} \qquad \mathsf{P}_{\mathcal{E}}$$

e a seguinte função de penalização quadrática

$$Q(x;\mu) = f(x) + \frac{\mu}{2} \sum_{i \in \mathcal{E}} (c_i(x))^2$$

em que $\mu > 0$ é o parâmetro da penalização.

Ao aumentar-se μ, penaliza-se a violação das restrições.

Exemplo:

Consideremos o problema

$$\begin{array}{ll} \displaystyle\min_{(x_1,x_2)} & x_1 + x_2 \\ \text{sujeito a} & x_1^2 + x_2^2 - 2 = 0 \end{array}$$

em que $x_* = (-1,-1)^\top$ é a solução.

A função de penalização quadrática é dada por

$$Q(x;\mu) = x_1 + x_2 + \frac{\mu}{2}(x_1^2 + x_2^2 - 2)^2$$

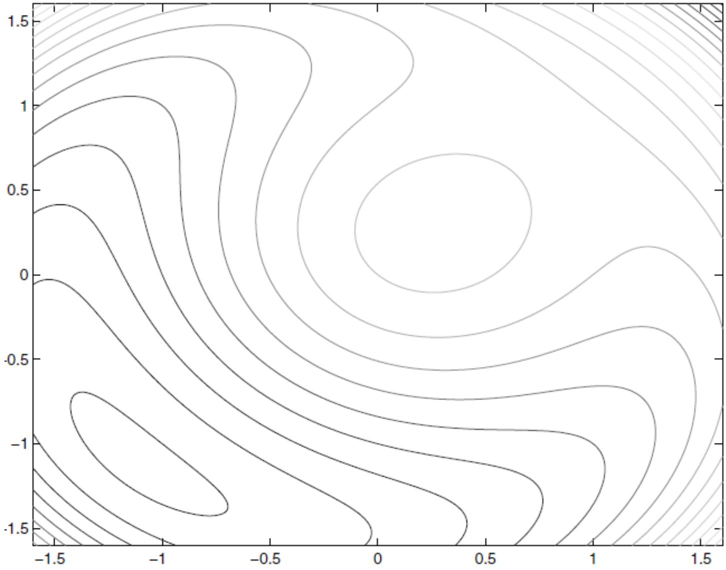

Figure: **curvas de nível de** $Q(x; 1)$

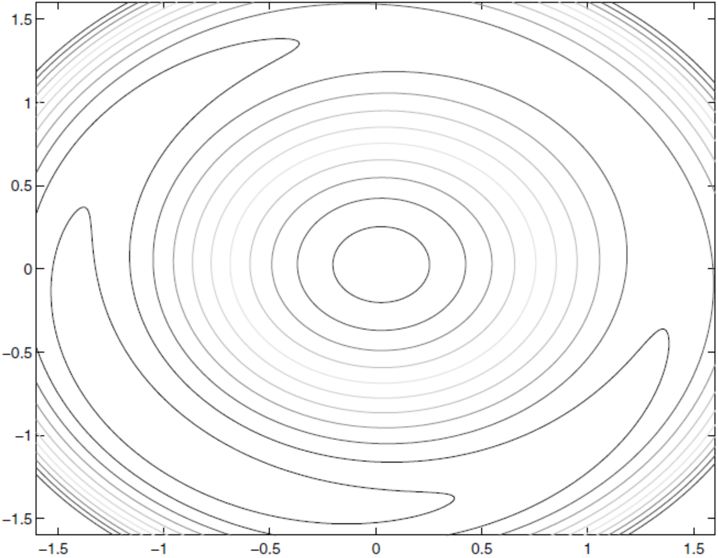

Figure: **curvas de nível de** $Q(x; 10)$

Algoritmo (**Método da penalização quadrática**)

Inicialização

Dado $\mu_0 > 0$, $x_0 = x_0^s$ um ponto inicial e $\{\tau_k\}$ em que $\tau_k \geq 0$ e $\tau_k \downarrow 0$.

Para $k = 0, 1, 2, \ldots$

1. Minimizar $Q(\,\cdot\,; \mu_k)$ partindo de x_k^s e terminando quando $\|\nabla_x Q(\,\cdot\,; \mu_k)\| \;\leq\; \tau_k$ (resultado em x_k).
2. Escolher $\mu_{k+1} > \mu_k$.
3. Escolher x_{k+1}^s (e.g. $x_{k+1}^s = x_k$).

Teorema

Seja x_k *um minimizante global de* $Q(\,\cdot\,;\mu_k)$ *e* $\{\mu_k\}$ *em que* $\mu_k \geq 0$ *e* $\mu_k \uparrow +\infty$*.*

Então qualquer ponto de acumulação x_* *de* $\{x_k\}$ *é um minimizante global de* $\mathrm{P}_{\mathcal{E}}$*.*

Demonstração: Seja $\bar{x}$ um minimizante global de $\mathrm{P}_{\mathcal{E}}$, ou seja,

$$f(\bar{x}) \;\leq\; f(x) \quad \forall x :\; c_i(x) = 0,\; i \in \mathcal{E}$$

Uma vez que x_k minimiza $Q(\,\cdot\,;\mu_k)\;\forall k$, temos

$$Q(x_k;\mu_k) \;\leq\; Q(\bar{x};\mu_k)$$

Logo,

$$f(x_k) + \frac{\mu_k}{2} \sum_{i \in \mathcal{E}} c_i^2(x_k) \; \leq \; f(\bar{x}) + \frac{\mu_k}{2} \underbrace{\sum_{i \in \mathcal{E}} c_i^2(\bar{x})}_{=0} = f(\bar{x})$$

$$\Downarrow$$

$$\sum_{i \in \mathcal{E}} c_i^2(x_k) \; \leq \; \frac{2}{\mu_k}[f(\bar{x}) - f(x_k)]$$

Suponhamos, agora, que x_* é um ponto de acumulação de $\{x_k\}$.

Assim, existe uma subsucessão K tal que

$$\lim_{k \in K} x_k = x_*$$

Por sua vez (atenção...),

$$\sum_{i\in\mathcal{E}} c_i^2(x_*) = \lim_{k\in K} \sum_{i\in\mathcal{E}} c_i^2(x_k) \ \leq \ \lim_{k\in K} \frac{2}{\mu_k}[f(\bar{x}) - f(x_k))] \underbrace{=}_{\mu_k\uparrow+\infty} 0$$

Temos, então, que

$$c_i(x_*) = 0, \quad \forall i \in \mathcal{E}$$

e, desta forma, x_* é admissível.

Vimos já que

$$\Big(f(x_k) \;\le\; \Big)\; f(x_k) + \frac{\mu_k}{2} \sum_{i \in \mathcal{E}} c_i^2(x_k) \;\le\; f(\bar{x}) + \frac{\mu_k}{2} \sum_{i \in \mathcal{E}} c_i^2(\bar{x}) = f(\bar{x})$$

Tomando limites, temos

$$f(x_*) \;\le\; f(\bar{x})$$

Assim, x_* é um minimizante global.

$\blacksquare$

Este resultado é pouco prático.

Supõe a minimização global de $Q(\,\cdot\,;\mu_k)$ em x.

O resultado seguinte apenas requer de x_k uma estacionaridade aproximada.

Teorema

Sejam f *e* $c_i,\ i \in \mathcal{E}$, *funções continuamente diferenciáveis.*

Seja $\{\tau_k\}$ *em que* $\tau_k \geq 0$ *e* $\tau_k \downarrow 0$.

Seja $\{\mu_k\}$ *em que* $\mu_k \geq 0$ *e* $\mu_k \uparrow +\infty$.

Seja x_* *um ponto de acumulação de* $\{x_k\}$ $\left(\exists K : \lim_{k \in K} x_k = x_*\right)$.

(i) *Se* x_* *for não admissível ele é um ponto estacionário do problema*

$$\min \ \|c(x)\|^2$$

(ii) *Se* x_* *verificar LICQ, então satisfaz as CN1O-KKT e*

$$\lim_{k \in K} -\mu_k c_i(x_k) = \lambda_i^*, \quad \forall i \in \mathcal{E}$$

Demonstração: (i) Comecemos por observar que

$$\nabla_x Q(x_k;\mu_k) = \nabla f(x_k) + \sum_{i\in\mathcal{E}} \mu_k c_i(x_k)\nabla c_i(x_k)$$

Assim, vem

$$\begin{array}{rcll} \|\nabla_x Q(x_k;\mu_k)\| & \leq & \tau_k & \iff \\ \left\|\nabla f(x_k) + \sum_{i\in\mathcal{E}} \mu_k c_i(x_k)\nabla c_i(x_k)\right\| & \leq & \tau_k & \Longrightarrow \\ \left\|\sum_{i\in\mathcal{E}} c_i(x_k)\nabla c_i(x_k)\right\| & \leq & \frac{1}{\mu_k}\left(\tau_k + \|\nabla f(x_k)\|\right) & \end{array}$$

Logo, tomando limites, vem que

$$\sum_{i \in \mathcal{E}} c_i(x_*) \nabla c_i(x_*) = 0$$

e temos que x_* é um ponto estacionário da função $\|c(x)\|^2$.

(ii) Seja $A(x) = \left[\nabla c_i(x)^\top \right]_{i \in \mathcal{E}}$.

Seja $\lambda_k = -\mu_k c(x_k)$, em que $c(x) = [c_i(x)]_{i \in \mathcal{E}}$.

Assim, temos

$$A(x_k)^\top \lambda_k = \nabla f(x_k) - \nabla_x Q(x_k; \mu_k)$$

com $\|\nabla_x Q(x_k; \mu_k)\| \leq \tau_k$.

Uma vez que, para k suf. grande, temos car$(A(x_k)) = |\mathcal{E}|$, então $A(x_k)A(x_k)^\top$ é não-singular.

Multiplicando

$$A(x_k)^\top \lambda_k = \nabla f(x_k) - \nabla_x Q(x_k; \mu_k)$$

por $A(x_k)$, vem

$$A(x_k)A(x_k)^\top \lambda_k = A(x_k)\left[\nabla f(x_k) - \nabla_x Q(x_k; \mu_k)\right]$$

Logo,

$$\lambda_k = \left[A(x_k) A(x_k)^\top \right]^{-1} A(x_k) \left[\nabla f(x_k) - \nabla_x Q(x_k; \mu_k) \right]$$

Tomando limites, vem

$$\lim_{k \in K} \lambda_k = \left[A(x_*) A(x_*)^\top \right]^{-1} A(x_*) \nabla f(x_*) \overbrace{=}^{\text{def}} \lambda_*$$

Por outro lado, tomando limites em

$$A(x_k)^\top \lambda_k = \nabla f(x_k) - \nabla_x Q(x_k; \mu_k)$$

vem que

$$A(x_*)^\top \lambda_* = \nabla f(x_*) \quad \Longleftrightarrow \quad \nabla f(x_*) - A(x_*)^\top \lambda_* = 0$$

Logo, x_* satisfaz as CN1O-KKT.

$\blacksquare$

Notas:

1. A Hessiana da função de penalização

$$\nabla^2_{xx} Q(x;\mu_k) = \nabla^2 f(x) + \sum_{i\in\mathcal{E}} \mu_k c_i(x) \nabla^2 c_i(x) + \mu_k A(x)^\top A(x)$$

é mal condicionada para μ_k suf. grande, o que dificultará a aplicação do método de Newton (ou métodos nele baseados).

De facto, quando x está próximo de x_*

$$\nabla^2_{xx} Q(x;\mu_k) \approx \nabla^2_{xx} L(x,\lambda_*) + \mu_k A(x)^\top A(x)$$

e o segundo termo torna-se (com μ_k suf. grande) arbitrariamente mal condicionado.

Porém, o sistema de Newton

$$\nabla^2_{xx} Q(x;\mu_k) p = -\nabla_x Q(x;\mu_k)$$

é equivalente a

$$\begin{bmatrix} \nabla^2 f(x) + \sum_{i\in\mathcal{E}} \mu_k c_i(x) \nabla^2 c_i(x) & A(x) \\ A(x)^\top & -\frac{1}{\mu_k} I \end{bmatrix} \begin{bmatrix} p \\ \zeta \end{bmatrix} = \begin{bmatrix} -\nabla_x Q(x;\mu_k) \\ 0 \end{bmatrix}$$

A matriz, 2×2 por blocos, deste sistema é, em princípio, bem condicionada quando x está próximo de x_* e μ_k é suf. grande (sob LICQ e CS2O) — ver final do capítulo anterior...

No entanto, note-se que o método de Newton pode, mesmo assim, ser lento (pois a sua região de convergência local pode ser muito pequena).

2. Na presença de restrições de desigualdade pode considerar-se a função de penalidade quadrática

$$Q(x;\mu) = f(x) + \frac{\mu}{2} \sum_{i \in \mathcal{E}} (c_i(x))^2 + \frac{\mu}{2} \sum_{i \in \mathcal{I}} \left([c_i(x)]^-\right)^2$$

em que $[y]^- = \max(-y, 0)$.

Observe-se que esta função apresenta uma ordem de suavidade a menos que o PNL original, pois tem uma derivada de segunda ordem descontínua.

Presentation outline

1. Métodos numéricos para optimização não linear sem restrições
   - Métodos de procura directa
   - Métodos de procura unidireccional (geral)
   - Métodos de procura unidireccional (descida máxima, Newton)
   - Métodos de região de confiança
2. Teoria da optimização não linear com restrições
   - Introdução e qualificações de restrições
   - Condições de primeira ordem
   - Condições de segunda ordem
   - Teoria da dualidade
3. Métodos numéricos para optimização não linear com restrições
   - Método da penalização quadrática
   - **Método do Lagrangeano aumentado**
   - Programação Sequencial Quadrática (SQP)
   - Métodos de pontos interiores

Consideremos um PNL com $\mathcal{I} = \emptyset$

$$\boxed{\begin{array}{rl} \min\limits_{x \in \mathbb{R}^n} & f(x) \\ \text{sujeito a} & c_i(x) \;=\; 0, \quad i \in \mathcal{E} \end{array}} \qquad \mathsf{P}_{\mathcal{E}}$$

O **método do Lagrangeano aumentado** (também conhecido por **método dos multiplicadores**) está relacionado com o **método da penalização quadrática**.

Pretende evitar o mau condicionamento quando μ_k se torna grande, forçando as restrições a serem satisfeitas mais rapidamente.

Introduz, explicitamente, uma aproximação para os multiplicadores, na função a minimizar (**função Lagrangeana aumentada**).

Conhece-se já a aproximação

$$\lambda_i^* \approx -c_i(x_k)\mu_k, \quad \forall i \in \mathcal{E}$$

para os multiplicadores.

Define-se, assim, a função Lagrangeana aumentada

$$L_A(x;\lambda,\mu) \overbrace{=}^{\text{def}} f(x) - \sum_{i\in\mathcal{E}} \lambda_i c_i(x) + \frac{\mu}{2}\sum_{i\in\mathcal{E}} (c_i(x))^2$$

Exemplo:

Consideremos o problema

$$\begin{array}{ll} \displaystyle\min_{(x_1,x_2)} & x_1 + x_2 \\ \text{sujeito a} & x_1^2 + x_2^2 - 2 = 0 \end{array}$$

em que $x_* = (-1,-1)^\top$ é a solução do problema (único minimizante local).

A função Lagrangeana aumentada é dada por

$$L_A(x;\lambda,\mu) = x_1 + x_2 - \lambda\left(x_1^2 + x_2^2 - 2\right) + \frac{\mu}{2}\left(x_1^2 + x_2^2 - 2\right)^2$$

Observe-se que $\lambda_* = -0.5$ é o multiplicador associado a x_* (CN1O).

Seja x_k um minimizante, aproximado, para $L_A(x; \lambda^k, \mu_k)$. Assim, temos

$$
\begin{aligned}
\nabla_x L_A(x_k; \lambda^k, \mu_k) &= \nabla f(x_k) - \sum_{i \in \mathcal{E}} \left(\lambda_i^k - \mu_k c_i(x_k) \right) \nabla c_i(x_k) \\
&= 0
\end{aligned}
$$

Comparando esta última expressão com

$$
\nabla_x L(x_k, \lambda^{\text{ideal}}) = \nabla f(x_k) - \sum_{i \in \mathcal{E}} \lambda^{\text{ideal}} \nabla c_i(x_k) = 0
$$

facilmente se deduz uma fórmula para melhorar a aproximação actual λ^k

$$
\lambda_i^{k+1} = \lambda_i^k - \mu_k c_i(x_k), \quad \forall i \in \mathcal{E}
$$

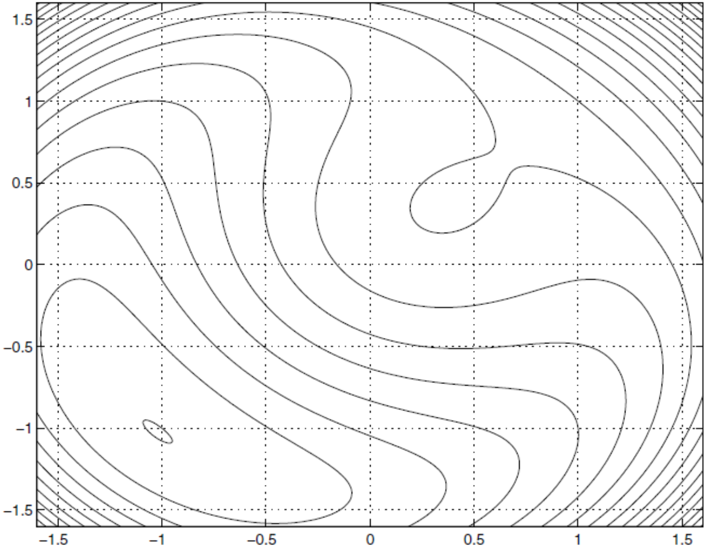

Figure: **curvas de nível de** $L_A(x; -0.4, 1)$

Algoritmo (**Método do Lagrangeano aumentado**)

Inicialização

Dados $x_0 = x_0^s$, λ^0 pontos iniciais e $\mu_0 > 0$, $\{\tau_k\}$ em que $\tau_k \geq 0$ e $\tau_k \downarrow 0$.

Para $k = 0, 1, 2, \ldots$

1. Minimizar $L_A(\,\cdot\,; \lambda^k, \mu_k)$ partindo de x_k^s e terminando quando $\left\|\nabla_x L_A(x_k; \lambda^k, \mu_k)\right\| \;\leq\; \tau_k$ (resultado em x_k).
2. Actualizar λ^{k+1}.
3. Escolher $\mu_{k+1} \geq \mu_k$.
4. Escolher x_{k+1}^s (e.g. $x_{k+1}^s = x_k$).

Teorema

Seja x_* um minimizante local de $P_{\mathcal{E}}$ a satisfazer LICQ (e numa vizinhança do qual todas as funções são duas vezes continuamente diferenciáveis).

Seja λ_* a satisfazer as CS2O.

Então

$$\exists \bar{\mu},\ \forall \mu \geq \bar{\mu} :\ x_* \text{ é um minimizante local estrito de } L_A(x; \lambda_*, \mu).$$

Notas:

1. Conhecendo λ_*, a função Lagrangeana aumentada é uma função de penalização "exacta" (no sentido em que não é necessário fazer μ tender para $+\infty$).

2. Existirá menos mal condicionamento (pois não será preciso aumentar muito μ) e como tal a escolha de x_{k+1}^s é menos crítica.

Teorema

Nas hipóteses do teorema anterior, existem escalares δ *e* M *positivos tais que:*

(A) *Se* $\dfrac{\|\lambda^k - \lambda_*\|}{\mu_k} \leq \delta$ *e* $\mu_k \geq \bar{\mu}$, *então o subproblema*

$$\min_{x \in \mathbb{R}^n} L_A(x; \lambda^k, \mu_k)$$

tem um minimizante local estrito (x_k) *suf. perto de* x_* *(onde as CS2O do subproblema são satisfeitas). Além disso, temos*

$$\|x_k - x_*\| \leq \frac{M}{\mu_k}\|\lambda^k - \lambda_*\|$$

(B) *Nas condições de (A)*

$$\|\lambda^{k+1} - \lambda_*\| \leq \frac{M}{\mu_k}\|\lambda^k - \lambda_*\|$$

Notas:

1. Neste teorema considera-se o cenário mais realista em que $\lambda^k \neq \lambda_*$.

2. O método do Lagrangeano aumentado dá-nos duas possibilidades de aproximar x_k de x_* (aumentar μ_k ou aproximar λ^k de λ_*).

3. (B) garante-nos uma taxa de convergência local **linear** para $\{\lambda^k\}$, quando μ_k for suf. grande.

4. (A) diz-nos, também, que o subproblema do algoritmo satisfaz localmente (perto de x_*) as CS2O.

Presentation outline

1. Métodos numéricos para optimização não linear sem restrições
   - Métodos de procura directa
   - Métodos de procura unidireccional (geral)
   - Métodos de procura unidireccional (descida máxima, Newton)
   - Métodos de região de confiança
2. Teoria da optimização não linear com restrições
   - Introdução e qualificações de restrições
   - Condições de primeira ordem
   - Condições de segunda ordem
   - Teoria da dualidade
3. Métodos numéricos para optimização não linear com restrições
   - Método da penalização quadrática
   - Método do Lagrangeano aumentado
   - Programação Sequencial Quadrática (SQP)
   - Métodos de pontos interiores

Consideremos um PNL com $\mathcal{I} = \emptyset$

$$\boxed{\begin{array}{rl} \min\limits_{x \in \mathbb{R}^n} & f(x) \\ \text{sujeito a} & c_i(x) \ = \ 0, \quad i \in \mathcal{E} \end{array}} \qquad \mathsf{P}_{\mathcal{E}}$$

Seja $c(x) = [c_i(x)]_{i \in \mathcal{E}}$ e $A(x) = \left[\nabla c_i(x)^\top\right]_{i \in \mathcal{E}}$.

Sabemos que a função Lagrangeana associada a $\mathsf{P}_{\mathcal{E}}$ é dada por

$$L(x, \lambda) = f(x) - \lambda^\top c(x)$$

As CN1O-KKT para $\mathrm{P}_{\mathcal{E}}$ consistem num sistema de equações não lineares $F(x,\lambda)=0$, em que $F:\mathbb{R}^{n+m}\rightarrow\mathbb{R}^{n+m}$,

$$F(x,\lambda)=\left[\begin{array}{c}\nabla f(x)-A(x)^\top\lambda\\ c(x)\end{array}\right]=0$$

O Jacobiano, em relação a x e a λ, deste sistema é dado por

$$F'(x,\lambda)=\left[\begin{array}{cc}\nabla^2_{xx}L(x,\lambda) & -A(x)^\top\\ A(x) & 0\end{array}\right]$$

Assim, a aplicação do método de Newton resulta em

$$\begin{bmatrix} x_{k+1} \\ \lambda_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ \lambda_k \end{bmatrix} + \begin{bmatrix} p_k \\ \Delta\lambda_k \end{bmatrix}$$

em que p_k e $\Delta\lambda_k$ são a solução do sistema (Newton-KKT) de equações lineares

$$\underbrace{\begin{bmatrix} \nabla^2_{xx}L(x_k,\lambda_k) & -A(x_k)^\top \\ A(x_k) & 0 \end{bmatrix}}_{F'(x_k,\lambda_k)} \begin{bmatrix} p_k \\ \Delta\lambda_k \end{bmatrix} = \underbrace{\begin{bmatrix} -\nabla f(x_k) + A(x_k)^\top \lambda_k \\ -c(x_k) \end{bmatrix}}_{-F(x_k,\lambda_k)}$$

Notas:

1. Observe-se que $F'(x, \lambda)$ é não-singular se car$(A(x)) = m$ e $\nabla^2_{xx} L(x, \lambda)$ for DP para todo o $d \neq 0$ tal que $A(x)d = 0$ — ver final do capítulo anterior...

2. Sob estas condições em (x_*, λ_*), o método de Newton apresenta uma taxa de convergência local quadrática.

Existe uma forma alternativa de se chegar ao mesmo sistema linear. Essa forma consiste em considerar o seguinte problema quadrático (PQ)

$$
\begin{array}{rl}
\min\limits_{p} & f(x_k) + \nabla f(x_k)^\top p + \frac{1}{2} p^\top \nabla^2_{xx} L(x_k, \lambda_k) p \\
\text{sujeito a} & A(x_k) p + c(x_k) \; = \; 0
\end{array}
$$

ou seja, tomando um modelo quadrático de f (com a curvatura dada pelo Lagrangeano) e linearizando as restrições.

De facto sob as hipóteses da nota 1, o PQ anterior tem solução única (x_k, l_k) tal que

$$
\begin{aligned}
\nabla^2_{xx} L(x_k, \lambda_k) p_k + \nabla f(x_k) - A(x_k)^\top l_k &= 0 \\
A(x_k) p_k + c(x_k) &= 0
\end{aligned}
$$

Desta forma, (p_k, λ_{k+1}), em que $\lambda_{k+1} = l_k$, satisfaz

$$\left[\begin{array}{cc} \nabla^2_{xx} L(x_k, \lambda_k) & -A(x_k)^\top \\ A(x_k) & 0 \end{array}\right] \left[\begin{array}{c} p_k \\ \lambda_{k+1} \end{array}\right] = \left[\begin{array}{c} -\nabla f(x_k) \\ -c(x_k) \end{array}\right]$$

e tem-se que λ_{k+1} é, de facto, igual a $\lambda_k + \Delta\lambda_k$.

Ambas as abordagens são úteis:

- a abordagem pela programação sequencial quadrática (SQP) permite algoritmos práticos neste e noutros cenários,
- Newton facilita a análise local.

Algoritmo (Método SQP (local))

Inicialização

Sejam (x_0, λ_0).

Para $k = 0, 1, 2, \ldots$

1. Calcular os valores de $f(x_k)$, $\nabla f(x_k)$, $\nabla^2_{xx} L(x_k, \lambda_k)$, $c(x_k)$ e $A(x_k)$.
2. Resolver o PQ para obter p_k e l_k.
3. Fazer $x_{k+1} = x_k + p_k$ e $\lambda_{k+1} = l_k$.

Notas:

3. Observe-se que

$$
\begin{aligned}
& L(x_k, \lambda_k) + \nabla_x L(x_k, \lambda_k)^\top p \\
= \quad & f(x_k) - \lambda_k^\top c(x_k) + \nabla f(x_k)^\top p - \underbrace{\Big(A(x_k)^\top \lambda_k \Big)^\top p}_{= -\lambda_k^\top c(x_k) \ (\text{se } A(x_k) p_k = -c(x_k))} \\
= \quad & f(x_k) + \nabla f(x_k)^\top p
\end{aligned}
$$

e, assim sendo, a função objectivo do PQ poderia ter sido escrita como um modelo quadrático do Lagrangeano.

Consideremos, agora, o caso em que temos restrições de desigualdade

$$\boxed{\begin{array}{rrcll} \min\limits_{x \in \mathbb{R}^n} & f(x) & & & \\ \text{sujeito a} & c_i(x) & = & 0, & i \in \mathcal{E} \\ & c_i(x) & \geq & 0, & i \in \mathcal{I} \end{array}} \qquad \text{PNL}$$

Linearizando as restrições, o PQ tem a forma

$$\boxed{\begin{array}{rrcll} \min\limits_{p} & f(x_k) + \nabla f(x_k)^\top p + \frac{1}{2} p^\top \nabla^2_{xx} L(x_k, \lambda_k) p & & & \\ \text{sujeito a} & \nabla c_i(x_k)^\top p + c_i(x_k) & = & 0, & i \in \mathcal{E} \\ & \nabla c_i(x_k)^\top p + c_i(x_k) & \geq & 0, & i \in \mathcal{I} \end{array}} \qquad \text{IPQ}$$

A versão local do método SQP seria fácil de esquematizar.

As restrições activas do IPQ podem constituir uma aproximação das restrições activas do PNL.

O resultado seguinte valida localmente esta aproximação.

Teorema

Seja (x_*, λ_*) a satisfazer LICQ, CS2O e complementaridade estrita.

Então,

se (x_k, λ_k) estiver suf. próximo de (x_*, λ_*),

há uma solução local do IPQ cujo conjunto das restrições activas $\underbrace{\mathcal{A}(x_k)}_{\text{IPQ}}$ coincide com $\underbrace{\mathcal{A}(x_*)}_{\text{PNL}}$.

Desta forma, o método SQP baseado em IPQ pode ser visto, localmente e a partir de uma certa ordem, como um método de Newton aplicado a

$$
\begin{array}{lrcll}
\displaystyle\min_{x\in\mathbb{R}^n} & f(x) & & & \\
\text{sujeito a} & c_i(x) & = & 0, & i \in \mathcal{E} \\
 & c_i(x) & = & 0, & i \in \mathcal{I} \cap \mathcal{A}(x_*)
\end{array}
$$

As estratégias de procura unidireccional e regiões de confiança podem ser utilizadas para globalizar os métodos SQP.

Porém, coloca-se a questão de como avaliar o passo nas CW (no caso da procura unidireccional) e decréscimos previstos e actuais (no caso das regiões de confiança).

Surgem, assim, as **funções mérito**.

A título ilustrativo, reformula-se primeiro o PNL do seguinte modo:

- substituindo $c_{\mathcal{I}}(x) \geq 0$ por $c_{\mathcal{I}}(x) - s = 0,\ s \geq 0$,
- fazendo $c(x, s) = \left[\begin{array}{c} c_{\mathcal{E}}(x) \\ c_{\mathcal{I}}(x) - s \end{array} \right]$,
- definindo $x = \left[\begin{array}{c} x \\ s \end{array} \right]$.

Chega-se, assim, a

$$\boxed{\begin{array}{rl} \displaystyle\min_{x\in\mathbb{R}^n} & f(x) \\ \text{sujeito a} & c_i(x) \; = \; 0, \quad i \in \mathcal{E} \text{ novo} \\ & x \; \geq \; 0 \end{array}} \qquad \text{PNL2}$$

Considera-se, então, a função mérito ℓ_1

$$\phi_1(x;\mu) = f(x) + \mu\|c(x)\|_1$$

(ignorando as restrições $x \geq 0$).

A CDS, por exemplo, assumiria a forma

$$\phi_1(x_k + \alpha_k p_k; \mu_k) \;\leq\; \phi_1(x_k; \mu_k) + \eta \alpha_k D(\phi_1(x_k; \mu_k); p_k), \quad \eta \in (0,1)$$

onde $D(\phi_1(x_k; \mu_k); p_k)$ é a derivada direccional de ϕ_1 em x_k ao longo de p_k.

Quando é que esta derivada **é negativa**?

Teorema

Seja (p_k, λ_{k+1}) *o passo de SQP (PQ e não IPQ).*

Então a derivada direccional de ϕ_1 *em* x_k*, ao longo de* p_k*, satisfaz*

$$D(\phi_1(x_k; \mu_k); p_k) = \nabla f(x_k)^\top p_k - \mu_k \|c(x_k)\|_1$$

e

$$D(\phi_1(x_k; \mu_k); p_k) \ \leq \ -p_k^\top \nabla^2_{xx} L(x_k, \lambda_k) p_k - (\mu_k - \|\lambda_{k+1}\|_\infty)\|c(x_k)\|_1$$

Logo, se $p_k \neq 0$ e $\nabla^2_{xx} L(x_k, \lambda_k)$ for DP, p_k é direcção de descida para ϕ_1 se $\mu_k > \|\lambda_{k+1}\|_\infty$.

(Bastaria, aliás, que a Hessiana reduzida $Z_k^\top \nabla^2_{xx} L(x_k, \lambda_k) Z_k$ fosse DP.)

Presentation outline

1. Métodos numéricos para optimização não linear sem restrições
   - Métodos de procura directa
   - Métodos de procura unidireccional (geral)
   - Métodos de procura unidireccional (descida máxima, Newton)
   - Métodos de região de confiança
2. Teoria da optimização não linear com restrições
   - Introdução e qualificações de restrições
   - Condições de primeira ordem
   - Condições de segunda ordem
   - Teoria da dualidade
3. Métodos numéricos para optimização não linear com restrições
   - Método da penalização quadrática
   - Método do Lagrangeano aumentado
   - Programação Sequencial Quadrática (SQP)
   - Métodos de pontos interiores

Os **métodos de pontos interiores** podem ser vistos como **métodos de continuação ou de homotopia**.

Consideremos o problema

$$\boxed{\begin{array}{ll} \displaystyle\min_{x\in\mathbb{R}^n} & f(x) \\ \text{sujeito a} & c_i(x) \;=\; 0 \quad i \in \mathcal{E} \text{ novo} \\ & x \;\geq\; 0 \end{array}} \qquad \text{PNL}2$$

Sendo $e = (1, \ldots, 1)^\top$ e $\mathcal{X} = \mathrm{diag}(x)$, as CN1O-KKT do PNL2 são dadas por

$$\left.\begin{array}{rcl} \nabla f(x) - A(x)^\top y - z & = & 0 \\ \mathcal{X} z & = & \mu e \\ c(x) & = & 0 \end{array}\right\} (1)$$

com $x \geq 0$ e $z \geq 0$, quando $\mu = 0$.

Este tipo de método consiste em resolver o sistema de equações não lineares (1), por uma sucessão $\{\mu_k\}$ a convergir para zero, mantendo $x, z > 0$.

Tal estratégia encontra-se, localmente, bem justificada pelo teorema seguinte.

Teorema

Seja (x_*, λ_*, z_*) *uma solução do PNL*2 *que satisfaz LICQ, CS2O e complementaridade estrita.*

Então,

para μ *suf. pequeno, o sistema* (1) *tem solução única* $(x(\mu), y(\mu), z(\mu))$ *perto de* (x_*, y_*, z_*)

e

$$\lim_{\mu \to 0} (x(\mu), y(\mu), z(\mu)) = (x_*, y_*, z_*)$$

A trajectória descrita por estes pontos chama-se **caminho central primal-dual**.

Tem existência apenas **local em PNL** mas existe **globalmente em PL**.

Uma outra forma de desenvolver métodos de pontos interiores para PNL, consiste em considerar o **problema de barreira** (primal-dual)

$$\begin{array}{rl} \min & f(x) - \mu \sum_{i=1}^{n} \log(x_i) \\ \text{sujeito a} & c(x) \;=\; 0 \end{array}$$

com $\mu > 0$.

O método de barreira consiste em resolver, aproximadamente, uma sucessão de problemas de barreira para $\{\mu_k\}$ em que $\mu_k \to 0$, mantendo x positivo.

Exercícios:

- Escreva as CN1O para o problema de barreira e mostre que são equivalentes a (1), supondo $\mu > 0$ e $x > 0$.

- Prove que o caminho central primal dual existe apenas localmente em PNL, mas existe globalmente em PL.

Existe, também, o **método de barreira primal**, baseado na formulação

$$\boxed{\begin{array}{rrcll} \displaystyle\min_{x\in\mathbb{R}^n} & f(x) & & & \\ \text{sujeito a} & c_i(x) & = & 0, & i\in\mathcal{E} \\ & c_i(x) & \geq & 0, & i\in\mathcal{I} \end{array}} \qquad \text{PNL}$$

Neste caso, considera-se o problema de barreira

$$\min_{x\in\mathbb{R}^n} \; f(x) - \mu \sum_{i\in\mathcal{I}} \log(c_i(x)) \qquad (\mathcal{E}=\emptyset)$$

Uma das dificuldades desta abordagem é encontrar um ponto inicial estritamente admissível ($c_{\mathcal{I}} > 0$).