Métodos de Penalidade
Método da Lagrangeana Aumentada

M. Fernanda P. Costa

Departamento de Matemática
Universidade do Minho

Outline

Método da Lagrangeana Aumentada

- Método da Lagrangeana Aumentada ou Método dos Multiplicadores - é uma modificação do Método de Penalidade de Quadrática para reduzir a possibilidade de mau condicionamento através da inclusão de estimativas dos multiplicadores de Lagrange na função a minimizar em cada iteração: a função Lagrangeana Aumentada.
- A função Lagrangeana Aumentada é uma combinação da função Lagrangeana e a função de penalidade quadrática.
- Em contraste com as funções de penalidade não suaves apresentadas, a função Lagrangeana Aumentada preserva, na grande maioria dos casos, a suavidade. Assim, o método pode ser implementada usando métodos baseados em derivadas para otimização sem restrições ou com restrições de limites simples.

Problemas com restrições de igualdade

Considere novamente a formulação geral do problema com restrições de igualdade:

$$\begin{array}{ll} \underset{w \in \mathbb{R}^{I}}{\text{minimizar}} & F(w) \\ \text{sujeito a} & c_n(w) = 0, \quad n \in \mathcal{E} \end{array} \tag{P2}$$

- No método de penalidade quadrática, como podemos ver do Teorema2, o minimizante $w^{(k)}$ de $Q(w;_k)$ não satisfaz exatamente as condições de admissibilidade $c_n(w) = 0$, $\forall n \in \mathcal{E}$:

$$c_n(w^{(k)}) \approx -\frac{\lambda_n^*}{\mu_k}, \forall n \in \mathcal{E}. \tag{1}$$

  Para se ter a certeza, temos $c_n(w^{(k)}) \to 0$, $\forall n \in \mathcal{E}$, quando $\mu_k \to \infty$.

- **Ideia:** redefinir a função $Q(w; \mu_k)$ de modo a que o seu minimizante $w^{(k)}$ satisfaça melhor as restrições de igualdade: $c_n(w^{(k)}) \approx 0, \forall n \in \mathcal{E}$, mesmo para valores moderados de μ_k.

- A função Lagrangeana Aumentada atinge este objetivo através da inclusão de estimativas dos multiplicadores de Lagrange λ na função a minimizar em cada iteração.

A função Lagrangeana Aumentada para (P2), é definida por:

$$\mathcal{L}_A(w,\lambda;\mu) = \underbrace{F(w) - \sum_{n\in\mathcal{E}} \lambda_n c_n(w)}_{\text{função Lagrangeana}} + \underbrace{\frac{\mu}{2}\sum_{n\in\mathcal{E}}[c_n(w)]^2}_{\substack{\text{termos de penalidade:}\\ \text{o quadrado da violação das restrições}}} \tag{L2}$$

(**nota:** A função Lagrangeana Aumentada difere da função Lagrangeana pela presença dos termos ao quadrado, daí, na designação do método utilizar-se o termo "Aumentada". E difere da função de penalidade quadrática $Q(w;\mu)$ pelo termo que considera os multiplicadores de Lagrange. Neste sentido, é uma combinação da função Lagrangeana e a função de penalidade quadrática.)

- O algoritmo, em cada iteração k, fixa o valor do parâmetro de penalidade em $\mu_k > 0$, fixa o vetor λ na estimativa atual $\lambda^{(k)}$, e minimiza a função Lagrangeana Aumentada em relação a w:

$$\underset{w \in \mathbb{R}^{\mathrm{I}}}{\text{minimizar}} \quad \mathcal{L}_A(w, \lambda^{(k)}; \mu_k)$$

- Denotando por $w^{(k)}$ o minimizante (aproximado) de $\mathcal{L}_A(w, \lambda^{(k)}; \mu_k)$, temos pela 1ª condição necessária de otimalidade sem restrição:

$$\begin{aligned} 0 &\approx \nabla_w \mathcal{L}_A(w^{(k)}, \lambda^{(k)}; \mu_k) \\ &= \nabla F(w^{(k)}) - \sum_{n \in \mathcal{E}} \lambda_n^{(k)} \nabla c_n(w^{(k)}) + \mu_k \sum_{n \in \mathcal{E}} c_n(w^{(k)}) \nabla c_n(w^{(k)}) \\ &= \nabla F(w^{(k)}) - \sum_{n \in \mathcal{E}} [\lambda_n^{(k)} - \mu_k c_n(w^{(k)})] \nabla c_n(w^{(k)}) \end{aligned} \tag{2}$$

Comparando (2) com a 1ª condição necessária de otimalidade para o problema (P2) $(\nabla_w L(w^*, \lambda^*) = 0 \Leftrightarrow \nabla F(w^*) - \sum_{n=1}^{N} \lambda_n^* \nabla c_n(w^*) = 0)$, deduz-se que

$$\lambda^* \approx \lambda_n^{(k)} - \mu_k c_n(w^{(k)}), \;\; \text{para todo } n \in \mathcal{E} \tag{3}$$

ou seja, perto da solução w^*, o minimizante $w^{(k)}$ de $\mathcal{L}_A(w, \lambda^{(k)}; \mu_k)$ satisfaz

$$c_n(w^{(k)}) \approx -\frac{1}{\mu_k}(\lambda_n^* - \lambda_n^{(k)}) \tag{4}$$

- Conclui-se de (4) que se $\lambda^{(k)}$ estiver próximo do vetor de multiplicadores ótimo λ^*, então a medida de não admissibilidade $\|c_{\mathcal{E}}(w^{(k)})\|$ será muito menor do que $\frac{1}{\mu_k}$, em vez de ser proporcional a $\frac{1}{\mu_k}$ como em (1).
- A relação (3) sugere uma fórmula de atualização para melhorar a estimativa atual do vetor de multiplicadores de Lagrange $\lambda^{(k)}$, usando o minimizante (aproximado) $w^{(k)}$ que acabámos de calcular:

$$\lambda_n^{(k+1)} = \lambda_n^{(k)} - \mu_k c_n(w^{(k)}), \quad \text{para todo } n \in \mathcal{E} \tag{5}$$

nota:

- funções de restrição de igualdade foram agrupadas na função vetorial $c_{\mathcal{E}} : \mathbb{R}^{\mathrm{I}} \rightarrow \mathbb{R}^{|\mathcal{E}|}$
- $-\mu_k c_n(w^{(k)}) \rightarrow \begin{cases} \lambda_n^* & \text{no método de penalidade quadrática} \\ 0 & \text{no método da Langrangeana Aumentada} \end{cases}$

Algoritmo3: Método da Lagrangeana Aumentada-Restrições de Igualdade

- Dar: μ_0, tolerância τ_0, pontos iniciais $w_s^{(0)}$ e $\lambda^{(0)}$
- **Para** $k = 0, 1, \ldots$

1. Encontrar um minimizante (aproximado) $w^{(k)}$ de $\mathcal{L}_A(w, \lambda^{(k)}; \mu_k)$, iniciando em $w_s^{(k)}$ e parar quando $\|\nabla_w \mathcal{L}_A(w, \lambda^{(k)}; \mu_k)\| \leq \tau_k$
2. **Se** $w^{(k)}$ satisfaz o critério de paragem para o problema (P2)
   Parar com a solução (aproximada) $w^{(k)}$
   fim se
3. Atualizar multiplicadores de Lagrange usando (5) para obter $\lambda^{(k+1)}$
4. Escolher novo parâmetro de penalidade $\mu_{k+1} > \mu_k$
5. Fazer o ponto inicial $w_s^{(k+1)} = w^{(k)}$
6. Escolher tolerância τ_{k+1}

- A convergência deste método pode ser assegurada sem aumentar μ indefinidamente. O mau condicionamento é, portanto, menos problemático do que método de penalidade quadrática (Algoritmo1), pelo que a escolha do ponto inicial $w_s^{(k+1)}$ no Algoritmo3 é menos crítica (fazemos simplesmente $w_s^{(k+1)} = w^{(k)}$).
- A tolerância τ_k pode ser escolhida de modo a depender da medida de não admissibilidade $\sum_{n\in\mathcal{E}} |c_n(w^{(k)}|$, e o parâmetro μ_k pode ser incrementado se a redução desta medida de não admissibilidade for insuficiente na iteração atual.

Exercício1: Considere novamente o problema

$$\underset{w\in\mathbb{R}^2}{\text{minimizar}} \quad F(w) = w_1 + w_2 \ \text{sujeito a} \ w_1^2 + w_2^2 - 2 = 0$$

cujo solução é $w^* = (-1, -1)^T$ e $\lambda^* = -0.5$. Minimize a função Langrangeana Aumentada $\mathcal{L}_A(w, \lambda; \mu)$ do problema, para $\mu = 1$ e $\lambda = -0.4$. Faça o gráfico dos contornos de $\mathcal{L}_A$.

Solução: $\mathcal{L}_A(w, \lambda; \mu) = w_1 + w_2 - \lambda(w_1^2 + w_2^2 - 2) + \frac{\mu}{2}(w_1^2 + w_2^2 - 2)^2$.

- Para $\mu = 1$ e $\lambda = -0.4$, a função $\mathcal{L}_A(w, -0.4; 1)$ tem um minimizante no ponto $(-1.02, -1.02)^T$, que está muito mais próximo da solução $w^* = (-1, -1)^T$ do que o minimizante de $Q(w; 1)$, que é $(-1.1072, -1.1072)^T$.

(**nota:** este exercício mostra que a inclusão do termo do multiplicador de Lagrange na função $\mathcal{L}_A(w, \lambda; \mu)$ pode resultar numa melhoria significativa em relação ao método de penalidade quadrática, como forma de reformular o problema de otimização com restrições.)

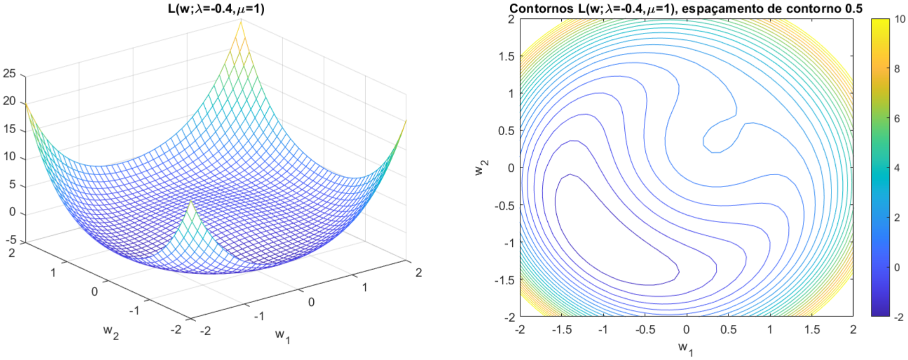

Fig. Função Lagrangeana Aumentada para $\lambda = -0.4$ e $\mu = 1$

Exercício2: Considere novamente o problema

$$\min_{w\in\mathbb{R}^2} \quad F(w) = w_1 + w_2 \text{ sujeito a } w_1^2 + w_2^2 - 2 = 0$$

a) Implemente o Método da Lagrangeana Aumentada (Algoritmo3) para resolver o problema, com $w_s^{(0)} = (0,0)^T$, $\lambda^{(0)} = 1$ e $\mu_0 = 1$. Faça $\mu_{k+1} = 10\mu_k$, $\tau_k = 1/\mu_k$ e escolha para ponto inicial $w_s^{(k+1)}$ a solução de $\mathcal{L}(w, \lambda_k; \mu_k)$. Para critério de paragem final use as condições:

$$\|c_{\mathcal{E}}(w^{(k)})\|_\infty \leq \varepsilon_1 \quad \text{(medida da violação)}$$

$$\frac{|F(w^{(k)}) - F(w^{(k-1)})|}{|F(w^{(k)})|} \leq \varepsilon_2, \quad \frac{\|w^{(k)} - w^{(k-1)}\|_\infty}{\|w^{(k)}\|_\infty} \leq \varepsilon_3$$

com $\varepsilon_1 = 10^{-6}, \varepsilon_2 = \varepsilon_3 = 10^{-4}$.
Indique a solução (aproximada) de cada função $\mathcal{L}_A(w, \lambda_k; \mu_k)$.

b) Compare a solução ótima obtida em a) com a solução obtida usando o método de penalidade quadrática.

Exercício3:(HS6) Considere o problema

$$\underset{w\in\mathbb{R}^2}{\text{minimizar}}\ F(w) = (1-w_1)^2 \ \text{sujeito a} \quad 10(w_2 - w_1^2) = 0$$

a) Utilize o Método da Lagrangeana Aumentada para resolver o problema, com $w^{(0)} = (-1.2, 1)^T$, $\lambda^{(0)} = 1$ e $\mu_0 = 1$. Faça $\mu_{k+1} = 2\mu_k$, $\tau_k = 1/\mu_k$ e escolha para ponto inicial $w_s^{(k+1)}$ a solução de $\mathcal{L}(w, \lambda_k; \mu_k)$. Para critério de paragem final use as condições:

$$\|c_{\mathcal{E}}(w^{(k)})\|_\infty \leq \varepsilon_1 \ \text{(medida da violação)}$$

$$\frac{|F(w^{(k)}) - F(w^{(k-1)})|}{|F(w^{(k)})|} \leq \varepsilon_2, \quad \frac{\|w^{(k)} - w^{(k-1)}\|_\infty}{\|w^{(k)}\|_\infty} \leq \varepsilon_3$$

com $\varepsilon_1 = 10^{-6}, \varepsilon_2 = \varepsilon_3 = 10^{-6}$.
Indique a solução (aproximada) de cada função de penalidade $\mathcal{L}_A(w, \lambda_k; \mu_k)$.

b) Compare a solução ótima obtida em a) com a solução obtida usando o método de penalidade quadrática.

Propriedades da função Langrangeana Aumentada

- O primeiro Teorema valida o Algoritmo3, mostrando que quando temos conhecimento do vetor exato dos multiplicadores de Lagrange λ^*, a solução w^* do problema (P2) é um minimizante estrito da função $\mathcal{L}_A(w, \lambda^*; \mu)$, para todo μ suficientemente grande ($\exists\ \overline{\mu} > 0 : \forall \mu \geq \overline{\mu}$).
- Na prática, embora não saibamos λ^* exatamente, o Teorema e a sua prova sugerem que podemos obter uma boa estimativa de w^* minimizando $\mathcal{L}_A(w, \lambda; \mu)$, mesmo quando μ não é particularmente grande, desde que λ seja uma estimativa razoavelmente boa de λ^*.

Teorema 1

Seja w^* uma solução local do problema (P2) na qual os gradientes das restrições $\nabla c_n(w^*), n \in \mathcal{E}$, são linearmente independentes, e a condição suficiente de 2ª ordem é satisfeita para $\lambda = \lambda^*$. Então existe um valor limiar $\overline{\mu}$ tal que para todo $\mu \geq \overline{\mu}$, w^* é um minimizante local estrito de $\mathcal{L}_A(w, \lambda^*; \mu)$.

Demonstração: (ver [1], Teorema 17.5)

- O segundo Teorema, descreve a situação mais realista de $\lambda \neq \lambda^*$. Fornece condições sob as quais existe um minimizante de $\mathcal{L}_A(w, \lambda; \mu)$ que se situa próximo de w^* e fornece limites de erro em $w^{(k)}$ e na estimativa de $\lambda^{(k+1)}$.

Teorema 2

Suponhamos que as hipóteses do Teorema 1 são satisfeitas em w^* e λ^* e que $\overline{\mu}$ seja escolhido como nesse teorema. Então existem escalares positivos δ, ϵ e M tais que as seguintes afirmações verificam-se:

(a) *Para todo $\lambda^{(k)}$ e μ_k satisfazendo*

$$\|\lambda^{(k)} - \lambda^*\| \leq \mu_k \delta, \quad \mu_k \geq \overline{\mu} \tag{6}$$

o problema

$$\min \mathcal{L}_A(w, \lambda^{(k)}; \mu_k) \quad \text{sujeito a} \quad \|w - w^*\| \leq \epsilon$$

tem uma única solução $w^{(k)}$. Além disso, temos

$$\|w^{(k)} - w^*\| \leq \|\lambda^{(k)} - \lambda^*\| / \mu_k \tag{7}$$

(b) Para todo $\lambda^{(k)}$ e μ_k que satisfaz (6), temos

$$\|\lambda^{(k+1)} - \lambda^*\| \leq \|\lambda^{(k)} - \lambda^*\|/\mu_k \tag{8}$$

onde $\lambda^{(k+1)}$ é dado pela fórmula (5).

(c) Para todo $\lambda^{(k)}$ e μ_k que satisfaz (6), a matriz $\nabla\mathcal{L}_A(w^{(k}, \lambda^{(k)}; \mu_k)$ é definida positiva e os gradientes das restrições $\nabla c_n(w^{(k)})$, $n \in \mathcal{E}$, são linearmente independentes.

Demonstração: (ver [1], Teorema 17.6)

O Teorema 2 ilustra algumas propriedades importantes do método da Lagrangeana Aumentada:

- O limite (7) mostra que $w^{(k)}$ se situará próximo de w^* se $\lambda^{(k)}$ for exato ou se μ_k for grande. Portanto, este método dá-nos duas formas de melhorar a precisão de $w^{(k)}$, enquanto o método de penalidade quadrática dá-nos apenas uma opção: aumentar μ_k.
- O limite (8) diz que, localmente, podemos garantir uma melhoria na exatidão dos multiplicadores escolhendo um valor suficientemente grande de μ_k.
- A observação final do teorema mostra que as condições suficientes de segunda ordem para a minimização sem restrições, também são satisfeitas pelo k-ésimo problema da iteração sob as condições dadas. Portanto, pode-se esperar um bom desempenho aplicando técnicas de minimização sem restrições.

Problemas com restrições de desigualdade

Considere novamente a formulação geral do problema com restrições

$$\begin{array}{rll} \underset{w\in\mathbb{R}^{\mathrm{I}}}{\text{minimizar}} & F(w) & \\ \text{sujeito a} & c_n(w) = 0, & n \in \mathcal{E} \\ & c_n(w) \geq 0, & n \in \mathcal{I} \end{array} \qquad \text{(P1)}$$

- Existe na literatura procedimentos da Lagrangeana Aumentada práticos, para lidar com restrições de desigualdade.
- Existem três tipos abordagens baseadas, respetivamente, em:
  - formulação com restrições de limite simples;
  - formulação com restrições lineares;
  - formulação sem restrições.

Vamos aqui apenas apresentar o procedimento da Lagrangeana Aumentada baseado em formulação sem restrições.

Dado o problema com restrições geral (P1), podemos converte-lo num problema com restrições de igualdade e restrições de limites simples, introduzindo variáveis de folga s_n e substituindo as desigualdades $c_n(w) \geq 0,\ n \in \mathcal{I}$, por:

$$c_n(w) - s_n = 0,\ \ s_n \geq 0,\ \ \text{para todo}\ \ n \in \mathcal{I}$$

Vamos considerar, para simplificar a exposição, que não existem restrições de igualdade no problema (P1) ($\mathcal{E} = \emptyset$).

Reformulando desta forma, o problema (P1) é reescrito como um problema com restrições de igualdade e restrições de limites simples, da forma:

$$\begin{array}{ll} \underset{w,s}{\text{minimizar}} & F(w) \\ \text{sujeito a} & c_n(w) - s_n = 0,\ \ s_n \geq 0, \quad \text{para todo}\ n \in \mathcal{I} \end{array} \tag{P1b}$$

Definindo a função Lagrangeana Aumentada em termos apenas das restrições de igualdade de (P1b) e sujeita às restrições de limites simples, obtemos o seguinte problema a resolver em cada iteração k do Algoritmo3:

$$\begin{array}{ll} \min\limits_{w,s} & \mathcal{L}_A(w,s,\lambda^{(k)};\mu_k) = F(w) - \sum\limits_{n\in\mathcal{AI}} \lambda_n^{(k)}(c_n(w)-s_n) + \frac{\mu_k}{2}\sum\limits_{n\in\mathcal{I}}(c_n(w)-s_n)^2 \\ \text{s.a} & s_n \geq 0, \ \text{para todo } n\in\mathcal{I} \end{array} \qquad \text{(L1b)}$$

Esta função, é uma função quadrática convexa e separável relativamente a cada uma das variáveis de folga s_n, $n \in \mathcal{I}$. Assim, podemos minimizar (L1b) em relação a cada uma das variáveis s_n separadamente:

$$\min_{w,s_n} \quad \mathcal{L}_A(w,s_n,\lambda^{(k)};\mu_k) = F(w) - \lambda_n^{(k)}(c_n(w)-s_n) + \frac{\mu_k}{2}(c_n(w)-s_n)^2$$

Como $\mathcal{L}_A(w,s_n,\lambda^{(k)};\mu_k)$ é quadrática convexa na variável s_n, o seu minimizante global ocorre quando a sua $\frac{\partial \mathcal{L}_A}{\partial s_n} = 0$, ou seja,

$$\frac{\partial \mathcal{L}_A(w,s_n,\lambda^{(k)};\mu_k)}{\partial s_n} = 0 \Leftrightarrow \lambda_n^{(k)} - \mu_k(c_n(w)-s_n) = 0 \Leftrightarrow s_n = c_n(w) - \frac{1}{\mu_k}\lambda_n^{(k)}$$

Se o minimizante $s_n = c_n(w) - \frac{1}{\mu_k}\lambda^{(k)} < 0$, então como $\mathcal{L}_A(w, s, \lambda^{(k)}; \mu_k)$ é convexa em s_n, o valor ótimo de s_n em (L1b) é 0. Portanto, os valores ótimos de s_n em (L1b) são:

$$s_n = \max\left(0, c_n(w) - \frac{1}{\mu_k}\lambda_n^{(k)}\right), \ \text{ para todo } n \in \mathcal{I}$$

Substituindo as soluções s_n, $n \in \mathcal{I}$, na função Lagrangeana em (L1b), obtemos uma forma equivalente do problema (L1b) em w apenas.

Observe que, considerando apenas os termos que envolvem s_n, obtemos:

$$-\lambda_n^{(k)}(c_n(w) - s_n) + \frac{\mu_k}{2}(c_n(w) - s_n)^2 =$$

$$= \begin{cases} -\lambda_n^{(k)} c_n(w) + \frac{\mu_k}{2} c_n^2(w), & \text{se } c_n(w) - \frac{1}{\mu_k}\lambda_n^{(k)} \leq 0 \\ -\frac{(\lambda_n^{(k)})^2}{2\mu_k}, & \text{caso contrário} \end{cases}$$

Portanto, substituindo em (L1b) obtemos o problema sem restrições:

$$\min_{w} \quad \mathcal{L}_A(w, \lambda^{(k)}; \mu_k) = F(w) + \sum_{n \in \mathcal{I}} \psi(c_n(w), \lambda_n^{(k)}; \mu_k) \tag{L1}$$

onde

$$\psi(t, \sigma; \mu) = \begin{cases} -\sigma t + \dfrac{\mu}{2} t^2, & \text{se } t \leq \dfrac{\sigma}{\mu} \\ -\dfrac{\sigma^2}{2\mu}, & \text{caso contrário} \end{cases}$$

- A definição de $\mathcal{L}_A(w, \lambda^{(k)}; \mu_k)$ em (L1) representa a extensão da Lagrangeana Aumentada para problemas com restrições de desigualdade. Com esta extensão, aplicamos o **Algoritmo3** ao caso com restrições de desigualdade.

- **Atualizar os multiplicadores de Lagrange** da forma:

$$\lambda_n^{(k+1)} = \max(\lambda_n^{(k)} - \mu_k c_n(w^{(k)}), 0), \ \text{ para todo } n \in \mathcal{I}$$

(nota: $\lambda_n \geq 0$, $n \in \mathcal{I}$, **para que as condições KKT se verifiquem na solução.)**

nota:

- Cada uma das funções $\psi(c_n(w), \lambda_n^{(k)}; \mu_k)$ é continuamente diferenciável em relação a w, mas existe em geral uma descontinuidade na segunda derivada em relação a w sempre que $c_n(w) = \mu\lambda_n^{(k)}$, para algum $n \in \mathcal{I}$.
- No entanto, quando a condição complementar estrita se verifica, o minimizante (aproximado) do subproblema (L1) geralmente não está próximo das regiões não suaves, portanto, não interfere no algoritmo que resolve o subproblema.

  (Quando a restrição n está ativa, então para uma iteração k suficientemente avançada tem-se $c_n(w^{(k)}) \approx 0$, enquanto $\lambda_n^{(k)}$ e $1/\mu_k$ são ambos significativamente maiores do que zero. Quando a restrição n não está ativa, temos $c_n(w^{(k)}) > 0$ enquanto $\lambda_n^{(k)} \approx 0$. Em ambos os casos, é de esperar que o ponto $w^{(k)}$ evite a região em que a condição de descontinuidade $c_n(w) = \mu\lambda_n^{(k)}$ é satisfeita.)

J. Nocedal and S. Wright.
Numerical Optimization.
Springer, 2006.