O pacote computacional IPOPT

On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming

Andreas Wächter & Lorenz T. Biegler

Mathematical Programming 106, 25–57 (2006) | Cite this article

Problema a ser resolvido:

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & f(x) \\ \text{s.t.} \quad & c(x) = 0 \\ & x_L \leq x \leq x_U, \end{aligned}$$

where $x_L \in [-\infty, \infty)^n$ and $x_U \in (-\infty, \infty]^n$, with $x_L^{(i)} \leq x_U^{(i)}$,

f, c duas vezes continuamente diferenciáveis.

- restrições $g(x) \leq 0 \longrightarrow g(x) + w = 0, \; w \geq 0$

Apresentação da teoria para o problema simplificado

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & f(x) \\ \text{s.t.} \quad & c(x) = 0 \\ & x \geq 0. \end{aligned}$$

IPOPT implementa barreira logarítmica. Seus subproblemas são

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & \varphi_\mu(x) := f(x) - \mu \sum_{i=1}^{n} \ln(x^{(i)}) \\ \text{s.t.} \quad & c(x) = 0 \end{aligned}$$

onde $\mu \to 0^+$ (teoria vista em aula)

Resolve-se o subproblema acima por suas condições KKT:

veja as contas no PDF anexo

$$\begin{cases} \nabla f(x) + \nabla c(x)\lambda - z = 0 \\ c(x) = 0 \\ XZe - \mu e = 0, \end{cases} \qquad \left(e = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \right)$$

onde, como visto em aula,

$$X = \operatorname{diag}(x_1, \dots, x_n) = \begin{bmatrix} x_1 & & 0 \\ & \ddots & \\ 0 & & x_n \end{bmatrix}$$

(análogo para Z).

veja as contas no PDF anexo

- Note que KKT para o problema original é o sistema acima com $\mu = 0$.

(a complementaridade $x_i z_i = 0$, $\forall i$, pode ser escrita como $XZe = 0$. Note que z_i é o multiplicador de Lagrange para $x_i \geq 0$).

Por isso queremos $\mu \to 0^+$.

Critério de parada

- Baseado no cumprimento aproximado das condições KKT do problema original.
- Medida de qualidade de KKT do subproblema

$$E_\mu(x, \lambda, z) := \max \left\{ \frac{\|\nabla f(x) + \nabla c(x)\lambda - z\|_\infty}{s_d}, \|c(x)\|_\infty, \frac{\|XZe - \mu e\|_\infty}{s_c} \right\}$$

↳ $\|\nabla f(x) + \nabla c(x)\lambda - z\|_\infty$ é o resíduo da 1ª equação do sistema KKT.

↳ $\|c(x)\|_\infty$ é a medida de inviabilidade

↳ $\|XZe - \mu e\|_\infty$ é o resíduo da complementaridade.

↳ $s_d \geqslant 1$ e $s_c \geqslant 1$ são parâmetros de escalamento. Eles visam afrouxar a parada em problemas mal-condicionados.

↳ Veja que $\|c(x)\|_\infty$ não é escalada. Ou seja, obter um ponto viável é prioridade, IPOPT não flexibiliza isso.

Parada com sucesso

$$E_0(\tilde{x}_*, \tilde{\lambda}_*, \tilde{z}_*) \leq \epsilon_{\text{tol}}$$

- $\varepsilon_{tol} > 0$: Tolerância fornecida pelo usário (padrão 10^{-8})
- $\mu = 0 \Rightarrow \approx$ KKT do problema original

Solução do subproblema

- Passos Newtonianos:

$$\begin{bmatrix} W_k & A_k & -I \\ A_k^T & 0 & 0 \\ Z_k & 0 & X_k \end{bmatrix} \begin{pmatrix} d_k^x \\ d_k^\lambda \\ d_k^z \end{pmatrix} = -\begin{pmatrix} \nabla f(x_k) + A_k\lambda_k - z_k \\ c(x_k) \\ X_k Z_k e - \mu_j e \end{pmatrix} \quad (*)$$

onde $W_k = \nabla^2_{xx} L(x_k, \lambda_k, z_k)$ e $A_k = \nabla c(x_k)$

- d = direção de Newton (visto em aula)
- IPOPT não resolve $(*)$ diretamente. Ao invés disso, resolve o seguinte sistema

equivalente:

$$\begin{bmatrix} W_k + \Sigma_k & A_k \\ A_k^T & 0 \end{bmatrix} \begin{pmatrix} d_k^x \\ d_k^\lambda \end{pmatrix} = -\begin{pmatrix} \nabla \varphi_{\mu_j}(x_k) + A_k \lambda_k \\ c(x_k) \end{pmatrix} \qquad (**)$$

onde $\Sigma_k := X_k^{-1} Z_k$. Com isso, a componente z da direção de Newton é calculada por

$$d_k^z = \mu_j X_k^{-1} e - z_k - \Sigma_k d_k^x$$

Veja que o sistema acima é menor e sua matriz é simétrica (W_k e Σ_k são simétricas). Isso facilita a resolução, pois existem boas técnicas para sistemas simétricos (Cholesky etc).

↳ Problema: se $A_k = \nabla c(x_k)$ não tiver posto completo (i.e., x_k não for regular), então a matriz do sistema (**) pode ser singular, e uma solução pode não existir. Para contornar este problema, em

IPOPT resolve-se o sistema "perturbado"

$$\begin{bmatrix} W_k + \Sigma_k + \delta_w I & A_k \\ A_k^T & -\delta_c I \end{bmatrix} \begin{pmatrix} d_k^x \\ d_k^\lambda \end{pmatrix} = -\begin{pmatrix} \nabla \varphi_{\mu_j}(x_k) + A_k \lambda_k \\ c(x_k) \end{pmatrix},$$

onde $\delta_w, \delta_c \geq 0$ são calculados a cada iteração de IPOPT. Este sistema sempre possui solução para certos $\delta_w, \delta_c \geq 0$, mesmo que $A_k = \nabla c(x_k)$ não tenha posto completo.

Novo iterando do método

Calculado (x_k, λ_k, z_k) e a direção de Newton d_k, o novo iterando consiste em um passo a partir de (x_k, λ_k, z_k) na direção d_k:

$$\begin{aligned} x_{k+1} &:= x_k + \alpha_k d_k^x \\ \lambda_{k+1} &:= \lambda_k + \alpha_k d_k^\lambda \\ z_{k+1} &:= z_k + \alpha_k^z d_k^z. \end{aligned}$$

onde os passos $\alpha_k, \alpha_k^z \in (0,1]$ são calculados para manter positividade de x e z.

Sendo $x_k > 0$ e $z_k > 0$, IPOPT garante os próximos $x_{k+1} > 0$ e $z_{k+1} > 0$ dando passos que mantenham uma fração de x_k e z_k:

$$\alpha_k^{\max} := \max\left\{\alpha \in (0,1] : x_k + \alpha d_k^x \geq (1-\tau_j)x_k\right\}$$
$$\alpha_k^z := \max\left\{\alpha \in (0,1] : z_k + \alpha d_k^z \geq (1-\tau_j)z_k\right\}$$

onde

$$\tau_j = \max\{\tau_{\min}, 1-\mu_j\}, \quad \tau_{\min} \in (0,1).$$ (pense que $j=k$)

Note que $(1-\tau_k)x_k > 0$ e que $1-\tau_k \to \tau_{\min}$ à medida que $\mu_k \to 0^+$.
A ideia é permitir $x_k \approx 0$ apenas no fim do método, quando $\mu_k \approx 0$.

Nota: IPOPT implementa muitas outras estratégias, não vistas em aula, que o tornam muito eficaz numericamente. A referência do pacote explica as principais.

Uma delas diz respeito ao tratamento de problemas com "interior vazio", ou seja, aqueles em que não existe x tal que

$$x_L < x < x_U$$

(relativo ao problema original

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{s.t.} \quad & c(x) = 0 \\ & x_L \leq x \leq x_U \end{aligned}$$

).

Antes de iniciar a resolução, por padrão IPOPT relaxa/alarga os limitantes x_L, x_U trocando-os da seguinte forma:

$$x_L^{(i)} \leftarrow x_L^{(i)} - \epsilon_{\text{tol}} \max\{1, |x_L^{(i)}|\}$$

onde $\varepsilon_{tol} > 0$ é a precisão de parada. Note que o lado direito é menor que x_L. Estratégia similar é usada para aumentar x_U.

↳ o exercício 4 da lista 2 discute essa estratégia