Mélykúti Bence

Egy polinomiális, az entrópia-programozási feladatot megoldó algoritmus bemutatása

2005. április 21.

Az előadás témája Florian Potra, Yinyu Ye: *A quadratically convergent polynomial algorithm for solving entropy optimization problems*, SIAM J. Optimization Vol. 3, No. 4, p 843-860, November 1993 cikk vázlatos bemutatása.

---

1. Bevezetés

Legyen $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $\forall j \in \{1, 2, \ldots, n\}: \quad f_j :]0, \infty[\to \mathbb{R}$.

$$\text{(EP)} \qquad \begin{array}{rcl} \min f(x) & = & \sum_{j=1}^{n} f_j(x_j) \\ x & \in & \Omega_p = \{x \in \mathbb{R}^n \mid Ax = b, x \geq 0\} \end{array}$$

$$\text{(ED)} \qquad \begin{array}{rcl} \max g(x,y) & = & b^T y - (x^T \nabla f(x) - f(x)) \\ (x,y) & \in & \Omega_d = \{(x,y) \in \mathbb{R}^{n+m} \mid x \in \Omega_p,\ \nabla f(x) - A^T y \geq 0\} \end{array}$$

$s := \nabla f(x) - A^T y$ slack vektor.

A Karush-Kuhn-Tucker-tétel alapján:
x^* optimális megoldás $\Longleftrightarrow$
1) primál megengedettség: $x^* \in \Omega_p$
2) duál megengedettség: $\exists y^* : \ (x^*, y^*) \in \Omega_d$
3) komplementaritás: $X^*(\nabla f(x^*) - A^T y^*) = 0$

Entrópia-programozás, *entropy optimization problem*, (EOP)

$\forall j \in \{1, 2, \ldots, n\} : \quad f_j \in C^2(\mathbb{R}_+)$ és

$$f_j''(x_j) = \mu_j(x_j^{d_j} + q_j),$$

ahol $\mu_j \geq 0$, $q_j \geq 0$, $d_j \in \mathbb{R}$.

Példák: x^2, $x \log x$, $x \log x - x$, $-\sqrt{x}$, $\frac{1}{x}$, $-\frac{x^\alpha}{\alpha} + x \,(\alpha < 1, \alpha \neq 0)$.

Az entrópia-programozás az egyik legnépszerűbb konvex, nemlineáris optimalizálási probléma (pl. képfeldolgozás, szállítási problémák).

Polinomialitás

Primál-duál belsőpontos algoritmusoknál minden lépésben egy $(x^k, y^k) \in \mathrm{Int}(\Omega_d)$ pontot kapunk, és az algoritmust akkor állítjuk le, ha $(x^k)^T s^k < \varepsilon$, ahol ε-t a felhasználó adja meg.

Általában véges sok iterációval nem kaphatunk pontos megoldást.
Új definíciót adunk a polinomialitásra.

Def. Az algoritmus polinomiális, ha $\exists p, q > 0, \quad \forall \varepsilon > 0$
$\exists K(\varepsilon) \in \mathbb{N}: \quad K(\varepsilon) = O(n^p |\log \varepsilon|^q); \quad k \geq K(\varepsilon) \Longrightarrow (x^k)^T s^k < \varepsilon;$
és valamilyen $r > 0$ mellett $O(n^r)$ műveletet igényel minden iteráció.

Két célunk lesz:
(1) Az EOP-ra polinomiális potenciál-csökkentő algoritmust adni.
(2) A polinomialitás megőrzésével gyors lokális konvergenciát biztosítani.

Egy belsőpontos algoritmust ismertetünk az EOP megoldására.
Feltéve, hogy ismert egy (x^0, y^0) megengedett pár, $O(\sqrt{n}|\log \varepsilon|)$ iterációval kaphatunk egy (x^k, y^k) megengedett párt, amelyre

$$\frac{(x^k)^T s^k}{(x^0)^T s^0} \leq \varepsilon.$$

A keresési irány, mint az szokásos, egy centráló- és egy csökkenési irány kombinációja lesz.
Másrészt adunk egy számolható feltételt, amelynek teljesülése esetén a tiszta Newton-eljárás alkalmazható az iterációs eljárás további részében, lokális kvadratikus konvergenciát biztosítva.

2. Polinomiális potenciál-csökkentő algoritmus

Tegyük fel, hogy $\forall j \in \{1, 2, \dots, n\}$: $\quad f_j''(x_j) = x_j^d$.
Ellenőrizhető, hogy az erre kapott eredmények igazak az elején definiált általánosabb esetben is.
Def. primál-duál potenciálfüggvény: legyen $\varrho \geq n + \sqrt{n}$.

$$\begin{aligned} \phi(x, s) &= \varrho \log(x^T s) - \sum_{j=1}^{n} \log(x_j s_j) \\ &= (\varrho - n) \log(x^T s) - \sum_{j=1}^{n} \log\left(\frac{x_j s_j}{x^T s}\right). \end{aligned}$$

A számtani és mértani közepek közti egyenlőtlenséget $x_1 s_1, \dots, x_n s_n$-re felírva kaphatjuk:

$$-\sum_{j=1}^{n} \log\left(\frac{x_j s_j}{x^T s}\right) \geq n \log n.$$

$$(\varrho - n) \log(x^T s) \leq \phi(x, s) - n \log n.$$

ϕ-t pontosan $(\varrho - n)|\log \varepsilon|$-kel kell csökkenteni, hogy $x^T s \leq \varepsilon$-t kapjunk.

Adott $0 < x_k \in \Omega_p$ és $s_k = \nabla f(x^k) - A^T y^k > 0$ esetén oldjuk meg a következő nemlineáris egyenletrendszert Δx, Δy változókra:

$$
\begin{aligned}
X^k \Delta s + S^k \Delta x &= \theta \underbrace{\left(\frac{(x^k)^T s^k}{\varrho} e - X^k S^k e \right)}_{=:p^k}, \\
A \Delta x &= 0, \\
\Delta s &= \nabla f(x^k + \Delta x) - \nabla f(x^k) - A^T \Delta y.
\end{aligned}
$$

Legyen $x^{k+1} = x^k + \Delta x$, $y^{k+1} = y^k + \Delta y$, és $s^{k+1} = \nabla f(x^{k+1}) - A^T y^{k+1}$.

Ekkor valamely $\beta \in]0, 1[$-mal definiált

$$
\theta := \frac{\beta \min \left\{ \sqrt{x_j^k s_j^k} \middle| 1 \leq j \leq n \right\}}{\|(X^k S^k)^{-1/2} p^k\|}
$$

mellett egy $\gamma > 0$ konstanssal

$$
\phi(x^{k+1}, s^{k+1}) \leq \phi(x^k, s^k) - \gamma.
$$

Igazából az első egyenletet nem kell pontosan megoldani. ▶

1. Állítás Legyen $n + \sqrt{n} \leq \varrho \leq 2n$, Δx és Δy pedig megoldása a következőnek:

$$\begin{aligned} X^k \Delta s + S^k \Delta x &= \theta\Big(\frac{(x^k)^T s^k}{\varrho} e - X^k S^k e\Big) + z^k = \theta p^k + z^k, \\ A\Delta x &= 0, \\ \Delta s &= \nabla f(x^k + \Delta x) - \nabla f(x^k) - A^T \Delta y. \end{aligned}$$

Ha valamilyen C konstans mellett $\|z^k\| \leq C\beta^2 \min\left\{x_j^k s_j^k \middle| 1 \leq j \leq n\right\}$, akkor olyan $\beta > 0$-t választva, amelyre $\beta(1+C\beta) \leq 1/2$ és $1 - C\beta \geq 0$, kapjuk, hogy

$$\phi(x^{k+1}, s^{k+1}) \leq \phi(x^k, s^k) - \gamma,$$

ahol $\gamma = (-\sqrt{3}/2)\beta(1 - C\beta) + \beta^2(1 + C\beta)^2$.

A z^k maradékra tett kikötés biztosítható egy Newton-lépés alkalmazásával. ▶

Tegyük fel, hogy Δx és Δy megoldja a következő lineáris egyenletrendszert:

$$\begin{aligned} X^k(\nabla^2 f(x^k)\Delta x - A^T\Delta y) + S^k\Delta x &= \theta p^k, \\ A\Delta x &= 0. \end{aligned}$$

2. Lemma Legyen $\bar{d}(\alpha) = \max\left\{1, |d| \max\{(1+\alpha)^{d-1}, (1-\alpha)^{d-1}\}\right\}$. Válasszunk olyan $\alpha \in]0,1[$-t, amelyre $\alpha\bar{d}(\alpha) = 1/2$. Ekkor

$$\|z^k\| \leq \frac{\bar{d}(\alpha)\beta^2 \min\{x_j^k s_j^k | 1 \leq j \leq n\}}{8}.$$

1. Algoritmus

Adott $Ax^0 = b,\ x^0 > 0$ és $s^0 = \nabla f(x^0) - A^T y^0 > 0$.
$k := 0$;
while $(x^k)^T s^k \geq \varepsilon$ **do**
  kiszámolni a legutóbbi egyenletrendszer $\Delta x,\ \Delta y$ megoldását;
  $x^{k+1} := x^k + \Delta x$;
  $y^{k+1} := y^k + \Delta y$;
  $s^{k+1} := \nabla f(x^{k+1}) - A^T y^{k+1}$;
  $k := k + 1$;
end while

1. Tétel Legyen $n + \sqrt{n} \leq \varrho \leq 2n$. Ekkor $x^k > 0,\ s^k > 0$, és kielégítik (EP) és (ED) primál-duál feladatpárt.
Az 1. Algoritmus $O\big(\phi(x^0, s^0) + (\varrho - n)|\log \varepsilon|\big)$ iterációt hajt végre.

Megjegyzés Generálható olyan (x^0, s^0) kezdeti megengedett pontpár, amellyel $\phi(x^0, s^0) \le (\varrho - n)|\log((x^0)^T s^0)|$.

Megjegyzés A gyakorlatban a lépéshosszt egy egydimenziós kereséssel is választhatjuk:

$$\bar{\eta} := \arg\min_{\eta \ge 0} \phi(x^k + \eta\Delta x, s^k + \eta\Delta s),$$

$$x^{k+1} := x^k + \bar{\eta}\Delta x,$$
$$y^{k+1} := y^k + \bar{\eta}\Delta y.$$

3. Lokális kvadratikus konvergencia

Egy kiszámolható feltételt fogunk adni a négyzetes konvergenciát garantáló Newton-eljárás alkalmazhatóságára.
Tegyük fel, hogy valamilyen k-ra

$$\|(X^k \nabla^2 f(x^k) X^k)^{-1}\| \|X^k s^k\| < 1.$$

Feltehető, hogy $k = 0$.
A Newton-algoritmust $\nabla f(x) - A^T y = 0,\ Ax = b$-re alkalmazzuk:

$$\begin{aligned} \nabla^2 f(x^0)\Delta x - A^T \Delta y &= -s^0, \\ A\Delta x &= 0. \end{aligned}$$

rendszert kell megoldani, és a megoldással

$$\begin{aligned} x^1 &:= x^0 + \Delta x, \\ y^1 &:= y^0 + \Delta y, \\ s^1 &:= \nabla f(x^1) - A^T y^1. \end{aligned}$$

2. Tétel Legyen $\alpha = 1/2\bar{d}(\alpha) \leq \frac{1}{2}$, és

$$\widehat{d} := \frac{\bar{d}(\alpha)}{(1-\alpha)^2(1-\alpha\bar{d}(\alpha))}.$$

Ekkor bármely $\beta \leq \alpha$-ra a

$$\widehat{d}\|(X^0\nabla^2 f(x^0)X^0)^{-1}\|\|X^0 s^0\| \leq \beta \tag{1}$$

egyenlőtlenség teljesülése maga után vonja, hogy

$$x^1 > 0, \quad \|(X^0)^{-1}(x^1 - x^0)\| \leq \beta,$$

$$\widehat{d}\|(X^1\nabla^2 f(x^1)X^1)^{-1}\|\|X^1 s^1\| \leq \beta^2.$$

2. Algoritmus

Adott $Ax^0 = b,\ x^0 > 0$ és $s^0 = \nabla f(x^0) - A^T y^0 > 0$.

Tegyük fel, hogy (x^0, s^0)-ra teljesül az 1,
azaz $\widehat{d}\|(X^0\nabla^2 f(x^0)X^0)^{-1}\|\|X^0 s^0\| \leq \beta$.

$k := 0$;

while $\|(X^k\nabla^2 f(x^k)X^k)^{-1}\|\|X^k s^k\| \geq \varepsilon$ **do**
  kiszámolni a Newton-rendszer $\Delta x,\ \Delta y$ megoldását (x^k, s^k)-ban;
  $x^{k+1} := x^k + \Delta x$;
  $y^{k+1} := y^k + \Delta y$;
  $s^{k+1} := \nabla f(x^{k+1}) - A^T y^{k+1}$;
  $k := k + 1$;
end while

Megjegyzés A 2. Algoritmus olyan $x^k > 0$, $Ax^k = b$ és s^k sorozatot állít elő, amelyre $\|(X^k\nabla^2 f(x^k)X^k)^{-1}\|\|X^k s^k\|$ négyzetesen tart nullához.

A fő eredmény, amely hosszabb számolással bizonyítható:

3. Tétel A 2. Algoritmus olyan megengedett pontsorozatot ad, amelynek tagjai kvadratikusan konvergálnak az optimális megoldáshoz:

$$\begin{aligned}
\|s^k\| &\le \frac{3^{|d|+1}}{\widehat{d}}\|(X^0)^{d+1}\|\beta^{2^k}, \\
\|X^k s^k\| &\le \frac{3^{|d|+2}}{\widehat{d}}\|(X^0)^{d+2}\|\beta^{2^k}, \\
\|x^k - x^*\| &\le 6\|X^0\|\beta^{2^k}.
\end{aligned}$$

Megjegyzés A cikkben (s^k), $(X^k s^k)$ és (x^k) sorozatok R-kvadratikus konvergenciáját bizonyítják. A szerzők állítják, hogy kicsit több erőfeszítéssel Q-kvadratikus konvergencia is bizonyítható.

4. További bonyolultság-elemzés

Legyen x^* az entrópia-programozási feladat optimális belső megoldása, továbbá tegyük fel, hogy a $\nabla^2 f(x)$ Hesse-mátrix pozitív definit a megengedettségi tartomány belsejében.

Megvizsgáljuk, hogy az 1. Algoritmusnak legfeljebb hány iterációra van szüksége az 1 $(\widehat{d}\|(X^k\nabla^2 f(x^k)X^k)^{-1}\|\|X^k s^k\| \leq \beta)$ teljesítéséhez.
Ha ez már teljesül, akkor elindíthatjuk a 2. Algoritmust.

A számításokból következik, hogy

$$\|(X^k\nabla^2 f(x^k)X^k)^{-1}\| \le \frac{1}{(1-\alpha)^2(1-\alpha\bar{d}(\alpha))}\|(X^*\nabla^2 f(x^*)X^*)^{-1}\|.$$

És vele együtt

$$\|(X^k\nabla^2 f(x^k)X^k)^{-1}\|\|X^k s^k\| \le \frac{1}{(1-\alpha)^2(1-\alpha\bar{d}(\alpha))}\|(X^*\nabla^2 f(x^*)X^*)^{-1}\|\|X^k s^k\|.$$

Könnyen látható, hogy

$$\|(X^*\nabla^2 f(x^*)X^*)^{-1}\|\|X^k s^k\| \le \|(X^*\nabla^2 f(x^*)X^*)^{-1}\|(x^k)^T s^k.$$

Tehát, ha az 1. Algoritmusban elértük, hogy

$$(x^k)^T s^k \le \frac{\delta^2}{\|(X^*\nabla^2 f(x^*)X^*)^{-1}\|},$$

akkor

$$\|(X^*\nabla^2 f(x^*)X^*)^{-1}\|(x^k)^T s^k \le \delta^2,$$

és ez a következőt adja:

$$\|(X^k\nabla^2 f(x^k)X^k)^{-1}\|\|X^k s^k\| \le \frac{\delta^2}{(1-\alpha)^2(1-\alpha\bar{d}(\alpha))}.$$

Ekkor (x^k, s^k)-t (x^0, s^0)-ra átkeresztelhetjük, és ezzel a párral elindíthatjuk a 2. Algoritmust, ugyanis a kvadratikus konvergenciát biztosító 1: $\widehat{d}\|(X^0\nabla^2 f(x^0)X^0)^{-1}\|\|X^0 s^0\| \le \beta$ teljesül a megfelelő δ konstanssal. Ez most még éppen nem látszik, de átszorzással, a szereplő mennyiségek becslésével könnyen adódik.

Összefoglalva, a következő komplexitási korlátot lehet bizonyítani:

4. Tétel Legyen x^* az entrópia-programozási feladat optimális belső megoldása, továbbá tegyük fel, hogy a $\nabla^2 f(x)$ Hesse-mátrix pozitív definit a megengedettségi tartomány belsejében.
Ez esetben az 1. Algoritmus az önmaga leállítását lehetővé tévő (x^k, s^k) párt generál

$$O\big(\phi(x^0, s^0) + \sqrt{n}(1 + |\log \|(X^* \nabla^2 f(x^*) X^*)^{-1}\||)\big)$$

lépésben.
További megfontolásokkal igazolható, hogy az ekkor indított 2. Algoritmus $f(x) - f(x^*) \leq \varepsilon$-t kielégítő, megengedett x megoldást állít elő legfeljebb

$$\log\left(\log\left(\frac{\sqrt{n}\|X^* \nabla^2 f(x^*) X^*\|}{\varepsilon}\right)\right)$$

lépésben.