Геометрия самосогласованных барьеров
I. Методы внутренней точки

Roland Hildebrand

LJK, Université Grenoble Alpes / CNRS

Летняя школа "Управление, Информация, Оптимизация"
Вороново, 18 июня 2019 г.

Выпуклость

Определение

Множество $X \subset \mathbb{R}^n$ *называется* **выпуклым**, *если для любых точек* $x, y \in X$ *и числа* $\lambda \in [0,1]$ *имеем*

$$\lambda x + (1-\lambda)y \in X.$$

Функция $f : X \to \mathbb{R}$, *определенная на выпуклом множестве* X, *называется* **выпуклой**, *если для любых точек* $x, y \in X$ *и числа* $\lambda \in [0,1]$ *имеем*

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y).$$

Выпуклые задачи оптимизации

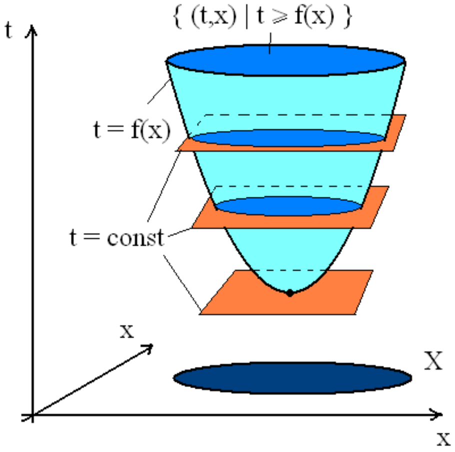
рассмотрим задачу вида

$$\min_{x \in X} f(x)$$

- $X \subset \mathbb{R}^n$ – непустое выпуклое множество
- $f : X \to \mathbb{R}$ – выпуклая функция цены, полунепрерывная снизу

полунепрерывная снизу $\Leftrightarrow$ замкнутый надграфик

Линейная функция цены

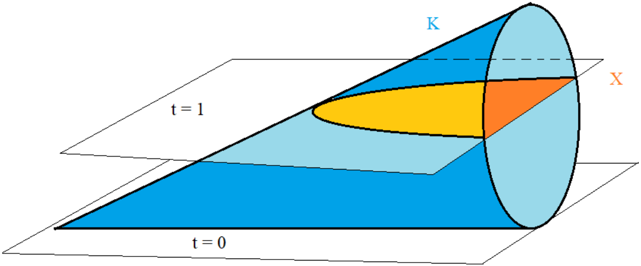


$f(x)$ можно предположить линейной

X можно предположить замкнутым

в противном случае минимизируем t по надграфику

Представление допустимого множества



замкнутое выпуклое множество X можно представить в виде пересечения замкнутого выпуклого конуса K с аффинным подпространством

Невырожденность конуса

соотношения K и X

$$K = \mathrm{cl}\{(t, tx) \mid t \geq 0,\ x \in X\}$$

$$X = \{x \mid (t, x) \in K,\ t = 1\} = K \cap \{(t, x) \mid t = 1\}$$

- если X не содержит прямых, то и K не содержит прямых
- перейдя к линейной оболочке, можно положить $K^o \neq \emptyset$

Конические программы

Определение

Выступающий замкнутый выпуклый конус $K \subset \mathbb{R}^n$ с непустой внутренностью называется регулярным.

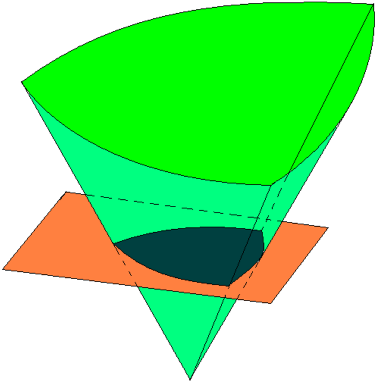
Определение

Коническая программа над регулярным конусом $K \subset \mathbb{R}^n$ – это задача оптимизации вида

$$\min_{x \in K} \langle c, x \rangle : \quad Ax = b.$$

любая задача выпуклой оптимизации может быть приведена к конической программе

Геометрическая интерпретация



допустимое множество представляется в виде пересечения конуса K с аффинным подпространством

альтернативная формулировка

$$\min_z \langle c', z\rangle : \ A'z + b' \in K$$

Классы регулярных конусов

в оптимизации встречаются

- симметрические конуса (классические конические задачи)
- конуса положительных отображений (робастная оптимизация)
- конуса моментов (полиномиальная оптимизация)
- конуса положительных полиномов (полиномиальная оптимизация)
- конуса сумм квадратов (выпуклые релаксации)
- коположительные конуса (невыпуклые задачи)
- экспоненциальный конус (геометрические программы)

Примеры конусов

симметрические конуса

- ортант $\mathbb{R}^n_+ = \{(x_1, \dots, x_n)^T \mid x_i \geq 0\}$
- конус Лоренца
$L_n = \left\{(x_0, \dots, x_{n-1}) \mid x_0 \geq \sqrt{x_1^2 + \dots + x_{n-1}^2}\right\}$
- матричный конус $\mathcal{S}^n_+$ вещественных симметрических неотрицательно определенных матриц

экспоненциальный конус

$$\begin{aligned} K_{\exp} &= \mathrm{cl}\,\{(t, tx, ty) \mid t \geq 0,\ y \geq e^x\} \\ &= \{(t, tx, ty) \mid t > 0,\ y \geq e^x\} \cap \{(0, tx, ty) \mid x \leq 0,\ y \geq 0\} \end{aligned}$$

ограничение $y \geq e^x$ записывается в виде $(1, x, y) \in K_{\exp}$

Примеры конических программ

- линейные программы (LP) над $\mathbb{R}_+^n$: $\sim 10^7$ переменных
- квадратично-конические программы (SOCP) над $\prod_j L_{n_j}$: $\sim 10^5$ переменных
- полу-определенные программы (SDP) над $\mathcal{S}_+^n$: $\sim 10^3$ переменных

наличие структуры позволяет решать бóльшие задачи

$LP \subset SOCP \subset SDP$

разработаны сольверы (CLP, LiPS, SDPT3, SeDuMi, CPLEX, MOSEK, ...)

Метод Ньютона

рассмотрим задачу минимизации локально строго выпуклой C^3 функции $f(x)$ на открытом выпуклом множестве D

на k-ом шаге: аппроксимируем

$$f(x) \approx q_k(x) = f(x_k) + \langle f'(x_k), x - x_k \rangle + \frac{1}{2}\langle f''(x_k)(x - x_k), x - x_k \rangle$$

минимизируем квадратичную аппроксимацию

$$x_{k+1} = x_k - \gamma_k (f''(x_k))^{-1} f'(x_k)$$

точный минимум достигается при $\gamma_k = 1$

метод аффинно инвариантен

как измерить прогресс, сделанный на данном шаге?

- локальная метрика $f''(x_k)$
- длина шага $\rho_k = \sqrt{(x_{k+1} - x_k)^T f''(x_k)(x_{k+1} - x_k)}$
- норма градиента $\rho_k = \sqrt{f'(x_k)^T (f''(x_k))^{-1} f'(x_k)}$
- прогнозируемый прогресс $f(x_k) - q_k(x_{k+1}) = \frac{\rho_k^2}{2}$

ρ_k называется Ньютоновским декрементом

истинный прогресс контролируется разницей $q_{k+1} - q_k$ между старой и новой аппроксимацией

итерация может вывести из множества D

Самосогласованные функции

Определение (Нестеров, Немировский)

Локально строго выпуклая C^3 функция $f : D \to \mathbb{R}$ называется **самосогласованной** *если для любого $x \in D$ и любого касательного вектора $h \in T_x D$ имеет место неравенство*

$$|f'''(x)[h,h,h]| \leq 2(f''(x)[h,h])^{3/2}.$$

Функция f называется **сильно самосогласованной** *если вдобавок имеет место*

$$\lim_{x \in \partial D} f(x) = +\infty.$$

Эллипсоид Дикина

для сильно самосогласованной функции на области D
единичный шар в локальной метрике (эллипсоид Дикина)

$$E_x = \{y \mid \langle f''(x)(y-x), y-x\rangle \leq 1\}$$

содержится в D

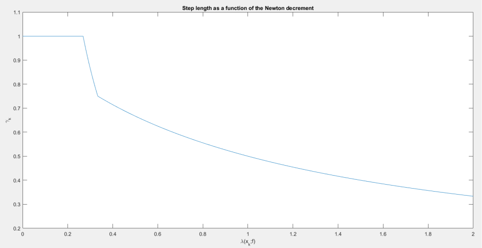
если $\rho_k < 1$, то полный шаг Ньютона не выведет из области D

иначе необходимо укоротить шаг, $\gamma_k \lesssim \rho_k^{-1}$

Метод Ньютона с укороченным шагом

для глобальной сходимости метода Ньютона следует положить

$$\gamma_k = \begin{cases} \frac{1}{1+\rho_k}, & \rho_k \geq \frac{1}{3} \\ \frac{1-\rho_k}{\rho_k(3-\rho_k)}, & \rho_k \in (2-\sqrt{3}, \frac{1}{3}) \\ 1, & \rho_k \leq 2-\sqrt{3} \end{cases}$$



Свойства метода

[Нестеров, Немировский 1994]

- если $\rho_k \geq \frac{1}{3}$, то $f(x_{k+1}) \leq f(x_k) - (\rho_k - \log(1+\rho_k))$
- значение ρ_k в промежуточном интервале $(2-\sqrt{3}, \frac{1}{3})$ принимается не более одного раза
- если $\rho_k \leq 2-\sqrt{3}$, то $\rho_{k+1} \leq \left(\frac{\rho_k}{1-\rho_k}\right)^2$
- если всюду $\rho > 1$, то функция неограничена снизу

вдалеке от минимума гарантировано понижение значения функции на ≈ 1

в окрестности минимума находимся в быстром (квадратичном) режиме сходимости

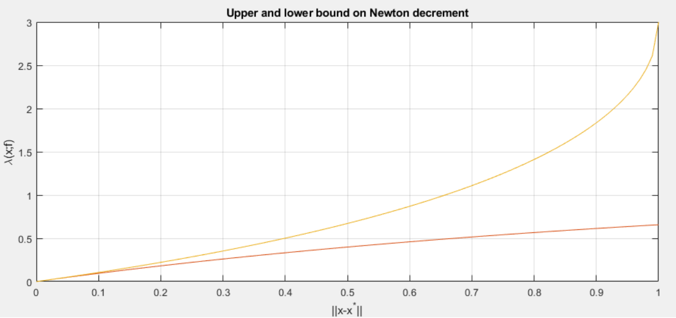
Радиус зоны быстрой сходимости

для любой точки x в эллипсоиде Дикина с центром в x^* справедливы неравенства

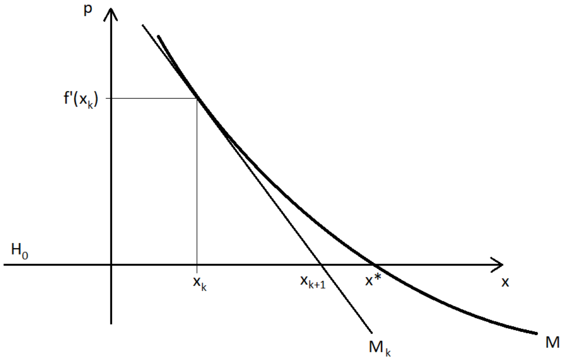
$$(\sqrt{1+r_x}-1)(3-\sqrt{1+r_x}) \leq \rho \leq (1-\sqrt{1-r_x})(3-\sqrt{1-r_x})$$

где $r_x = ||x - x^*||_{x^*}$

шар с радиусом $1-(2-\sqrt{3-\sqrt{3}})^2 \approx 0.2362$ находится в зоне квадратической сходимости оценки на декремент как функция от расстояния до минимума



Геометрическая интерпретация



M – граф градиента f, M_k – граф градиента q_k

Принцип метода внутренней точки

сложность конической программы в ограничении $x \in K$

устраняем это условие посредством добавления барьерной функции $F : K^o \to \mathbb{R}$ к функции цены

вместо исходной задачи

$$\min_{x \in K} \langle c, x \rangle \quad : \qquad Ax = b$$

получаем 1-параметрическое семейство задач

$$\min_{x} \tau \langle c, x \rangle + F(x) \quad : \qquad Ax = b$$

$\tau > 0$ – параметр семейства

разумно предъявить следующие требования к F

- $F(x)$ выпукла (проблема минимизации суммы должна оставаться выпуклой)
- $\lim_{x \to \partial K} F(x) = +\infty$ (барьерное свойство)
- $F(x)$ самосогласованная (чтобы использовать метод Ньютона с высокой локальной скоростью сходимости)

свойства остаются в силе

- для функций $\tau\langle c, x\rangle + F(x)$
- при ограничении x на допустимое множество X

Центральный путь

пусть решение исходной задачи существует, тогда

- при достаточно больших τ точки минимума $x^*(\tau)$ вспомогательной задачи существуют и единственны
- при $\tau \to +\infty$ решение $x^*(\tau)$ стремится к некоторому x^* в относительной внутренности множества решений

дифференцируемую кривую $x^*(\tau)$ называют центральным путем

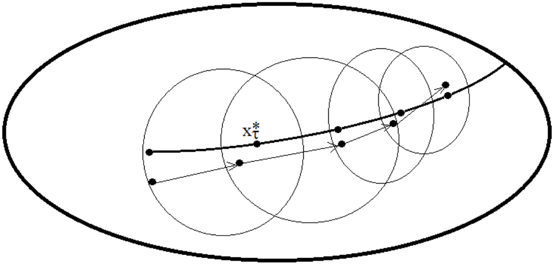
Алгоритм

- стартуем с пары $(x_0, \tau_0) \in X \times \mathbb{R}_{++}$, где x_0 находится в зоне квадратичной сходимости вокруг $x^*(\tau_0)$
- делаем полный шаг Ньютона по направлению к решению $x^*(\tau_0)$, получаем x_1
- обновляем значение параметра на $\tau_1 > \tau_0$, так что x_1 все еще в зоне квадратичной сходимости вокруг $x^*(\tau_1)$
- переход к следующей итерации

на каждом шаге выполняются неравенства

$$\rho_k^k \leq \overline{\lambda}, \qquad \rho_{k+1}^k \leq \underline{\lambda}$$

ρ_k^l – декремент в x_k по отношению к функции $\tau_l \langle c, x \rangle + F(x)$



метод попеременно делает шаг Ньютона и повышает τ, все время находясь в зоне квадратической сходимости вокруг x_τ^*

итерации не выходят из окрестности центрального пути константы $0 < \underline{\lambda} < \overline{\lambda} < 2 - \sqrt{3}$ выбраны так, что

$$\rho_k^k \leq \overline{\lambda} \quad \Rightarrow \quad \rho_{k+1}^k \leq \underline{\lambda}$$

для некоторых $0 < \underline{d} < \overline{d}$ имеем

- $\rho_{k+1}^k \leq \underline{\lambda} \Rightarrow ||x_{k+1} - x^*(\tau_k)||_{x^*(\tau_k)} \leq \underline{d}$
- $||x_{k+1} - x^*(\tau_{k+1})||_{x^*(\tau_{k+1})} \leq \overline{d} \Rightarrow \rho_{k+1}^{k+1} \leq \overline{\lambda}$

если расстояние между точками $x^*(\tau_k)$ и $x^*(\tau_{k+1})$ не превышает некоторую константу порядка $\overline{d} - \underline{d}$ (надо учесть поправку на то, что эти расстояния измеряются в разных, но вследствие самосогласованности не сильно отличающихся друг от друга метриках), то все неравенства будут выполняться на каждом шаге

Расстояния на центральном пути

насколько можно увеличивать параметр τ на каждом шаге?

на центральном пути необходимо сопоставить параметр τ с параметром длины в локальной метрике

Лемма

Для всех значений $\tau > 0$, отвечающих точке на центральном пути, справедливо соотношение $||\frac{dx^*(\tau)}{d\log\tau}||_{x^*(\tau)} = \sqrt{F'(x^*(\tau))^T(F''(x^*(\tau)))^{-1}F'(x^*(\tau))} = ||F'(x^*(\tau))||_{x^*(\tau)}$.

можно увеличивать $\log\tau$ на величину порядка $\frac{\bar{d}-\underline{d}}{||F'(x^*(\tau))||_{x^*(\tau)}}$

для линейной скорости сходимости необходимо условие $||F'(x^*(\tau))||_{x^*(\tau)} < const$

Самосогласованные барьеры

Определение

Самосогласованным барьером с параметром ν на выпуклом множестве X называется C^3 функция $F : X^o \to \mathbb{R}$, удовлетворяющая условиям

- $F'' \succ 0$ (локально строгая выпуклость)
- $F|_{\partial X} = +\infty$ (свойство барьера)
- $F'''(x)[h,h,h] \leq 2(F''(x)[h,h])^{3/2}$ для всех $x \in X^o$, $h \in T_x X$
- $F'(x)[h] \leq \sqrt{\nu F''(x)[h,h]}$ для всех $x \in X^o$, $h \in T_x X$

Параметр и скорость сходимости

если F – самосогласованный барьер с параметром ν, то на каждом шаге можно увеличивать $\log \tau$ на величину порядка $\nu^{-1/2}$

чем больше ν, тем медленнее метод будет сходиться

поэтому желательно иметь в наличии барьеры с как можно меньшим параметром

сложность задачи зависит от наличия эффективно вычислимого самосогласованного барьера с небольшим значением параметра

Логарифмично однородные барьеры

Определение (Ю.Е. Нестеров, А.С. Немировский 1994)

Пусть $K \subset \mathbb{R}^n$ – регулярный выпуклый конус. Логарифмично однородным самосогласованным барьером на K называется C^3 функция $F : K^o \to \mathbb{R}$, удовлетворяющая условиям

- $F(\alpha x) = -\nu \log \alpha + F(x)$ *(логарифмичная однородность)*
- $F''(x) \succ 0$ *(выпуклость)*
- $\lim_{x \to \partial K} F(x) = +\infty$ *(барьерное свойство)*
- $|F'''(x)[h,h,h]| \leq 2(F''(x)[h,h])^{3/2}$ *(самосогласованность)*

для всех касательных векторов h в каждой точке $x \in K^o$. Параметр однородности ν называется параметром барьера.

растяжения действуют прибавлением констант к F

Ограниченность декремента

оба определения совместимы, поскольку логарифмичная однородность ограничивает Ньютоновский декремент

дифференцируя соотношение $F(\alpha x) = -\nu \log \alpha + F(x)$ по x
получим $\alpha F'(\alpha x) = F'(x)$
дифференцируя это и исходное соотношение по α при $\alpha = 1$ получим

$$F'(x) + F''(x) \cdot x = 0, \qquad \langle F'(x), x \rangle = -\nu$$

$$(F''(x))^{-1} F'(x) = -x, \; (F'(x))^T (F''(x))^{-1} F'(x) = -\langle F'(x), x \rangle = \nu$$

Двойственный конус

Определение

Пусть $K \subset \mathbb{R}^n$ – регулярный выпуклый конус. ***Двойственным*** *к K называется конус*

$$K^* = \{p \in \mathbb{R}_n = (\mathbb{R}^n)^* \mid \langle p, x \rangle \geq 0 \ \forall \ x \in K\}.$$

конус K^* также является регулярным

$(K^*)^* = K$

Симметрические конуса

Определение

Самодвойственный однородный конус называется ***симметрическим****.*

- однородный: группа линейных автоморфизмов конуса действует транзитивно на внутренности
- самодвойственный: линейно изоморфен двойственному

примеры

- ортант $\mathbb{R}^n_+$
- конус Лоренца L_n
- матричный конус $\mathcal{S}^n_+$
- прямые произведения $\prod_{i=1}^m K_i$

Двойственный барьер

пусть $f : D \to \mathbb{R}$ – выпуклая функция

преобразованием Лежандра функции f называется функция

$$f^*(p) = \sup_{x \in D} \langle p, x \rangle - f(x)$$

Лемма (Ю.Е. Нестеров, А.С. Немировский 1994)

Преобразование Лежандра логарифмично однородного барьера с параметром ν на конусе K является логарифмично однородным барьером с тем же параметром ν на $-K^*$.

$$F_*(p) = F^*(-p) = \sup_{x \in K} -\langle p, x \rangle - F(x)$$

Римановы метрики

положительно определенный гессиан барьера F на конусе K можно интерпретировать как риманову метрику

тогда внутренность конуса K принимает структуру полного риманова многообразия

Лемма (Ю.Е. Нестеров, А.С. Немировский 1994)

Преобразование Лежандра $\mathcal{D} : x \mapsto p = -F'(x)$ является изометрией между внутренностью прямого и двойственного конусов.

Двойственная программа

прямая коническая программа

$$\min_{x \in K} \langle c, x \rangle \quad : \qquad Ax = b$$

рассмотрим произвольный элемент $s \in K^*$ вида $s = c - A^T y$

для любого x из допустимого множеств имеем

$$\langle s, x \rangle = \langle c, x \rangle - \langle A^T y, x \rangle = \langle c, x \rangle - \langle y, Ax \rangle = \langle c, x \rangle - \langle y, b \rangle \geq 0.$$

мы получили нижнюю оценку $\langle b, y \rangle$ на оптимальное значение исходной (прямой) программы двойственная коническая программа над двойственным конусом K^* формулируется как задача *максимизации* этой оценки

$$\max_y \langle b, y\rangle \quad : \qquad c - A^T y \in K^*$$

для каждой допустимой точки x величина $\langle c, x\rangle$ является верхней оценкой оптимального значения двойственной программы

разница между оптимальными значениями – разрыв двойственности

Симметричная формулировка

аффинную оболочку допустимого множества исходной программы можно представить в виде суммы $b + L$

$L \subset \mathbb{R}^n$ – линейное подпространство

вектор $b \in \mathbb{R}^n$ выбран так, что $\langle c, b \rangle = 0$

прямая программа

$$\min_{x \in K} \langle c, x \rangle \quad : \qquad x \in b + L$$

двойственная программа

$$\max_{s \in K^*} -\langle b, s \rangle \quad : \qquad s \in c + L^\perp$$

для допустимых (x, s) имеем $\langle x, s \rangle = \langle c, x \rangle + \langle b, s \rangle \geq 0$

Прямо-двойственные методы

прямо-двойственные методы решают прямую и двойственную программы одновременно и генерируют пары $(x_k, s_k) \in \operatorname{int} K \times \operatorname{int} K^*$

последовательность зазоров $\langle c, x_k \rangle + \langle b, s_k \rangle$ монотонно убывает

некоторые методы минимизируют потенциал, например

$$V(x, s) = F(x) + F_*(s) + const \cdot \log\langle x, s \rangle$$

константа зависит от параметра ν барьера F

Симметрические конуса

для классических задач используются следующие барьеры

класс	K	F	ν
LP	$\mathbb{R}_+^n$	$-\sum_{i=1}^n \log x_i$	n
SOCP	$\prod_{j=1}^J L_{n_j}$	$-\frac{1}{2}\sum_j \log((x_0^j)^2 - (x_1^j)^2 - \dots - (x_{n_j-1}^j)^2)$	$2J$
SDP	$\mathcal{S}_+^n$	$-\log\det A$	n

на $\prod_{i=1}^m K_i$ используется $F(x_1,\dots,x_m) = \sum_{i=1}^m F_i(x_i)$ с параметром $\nu = \sum_{i=1}^m \nu_i$

- параметр этих барьеров оптимальный
- барьеры *авто-шкалированные* $\Rightarrow$ методы с длинным шагом

Экспоненциальный конус

$$K_{\exp} = \{(x,y,0) \mid x \leq 0,\ \ y \geq 0\} \cup \{(x,y,z) \mid z > 0,\ \ y \geq z e^{x/z}\}$$

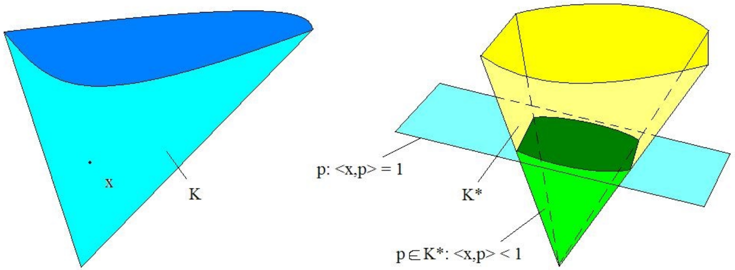
на экспоненциальном конусе имеется барьер

$$F(x,y,z) = -\log(z \log \frac{y}{z} - x) - \log y - \log z,$$

значение параметра $\nu = 3$ также оптимально

Универсальный барьер

пусть K – произвольный регулярный выпуклый конус



функция объема $V : K^o \ni x \mapsto Vol\{p \in K^* \mid \langle x,p \rangle < 1\}$

Универсальный барьер

Теорема (Нестеров, Немировский 1994)

Существует константа $c > 0$ такая, что для произвольного регулярного выпуклого конуса $K \subset \mathbb{R}^n$

$$F(x) = c \log V(x)$$

является самосогласованным барьером на K с параметром $(c \cdot n)$. Этот барьер называется универсальным.

позже было установлено, что можно выбрать $c = 1$ [Bubeck, Eldan '15, Lee '18]

Интегральное представление

Лемма (Güler 1996)

Универсальный барьер с точностью до аддитивной константы равен $c \log \varphi(x)$, *где*

$$\varphi(x) = \int_{K^*} e^{-\langle x,p \rangle} dp$$

характеристическая функция K.

Свойства

- $A \in \mathrm{Aut}\ K,\ F(Ax) = F(x) + \log\det\ A$
- инвариантен по отношению к действию $SL(n, \mathbb{R})$
- $F_{\prod_i K_i} = \sum_i F_{K_i}$

не инвариантен по отношению к двойственности

для неоднородных конусов трудно вычислим

Энтропический барьер

[Bubeck, Eldan '15]

двойственный к универсальному

обладает теми же свойствами

Канонический барьер

Теорема

Пусть $D \subset \mathbb{R}^n$ – выпуклая область, не содержащая прямой. Тогда существует единственное выпуклое решение $F : D \to \mathbb{R}$ уравнения в частных производных $\log\det F'' = 2F$ с граничным условием $\lim_{x \to \partial D} F(x) = +\infty$.

Теорема

Если D – внутренность выпуклого регулярного конуса K, это решение является логарифмично однородным самосогласованным барьером на K со значением параметра $\nu = n$. Этот барьер называется канонический. Двойственный барьер к каноническому барьеру на конусе K совпадает с каноническим барьером на двойственном конусе K^*.

Свойства

обладает теми же свойствами инвариантности что и универсальный

совпадает с ним на однородных конусах

вычислим на некоторых неоднородных конусах с большой группой симметрий

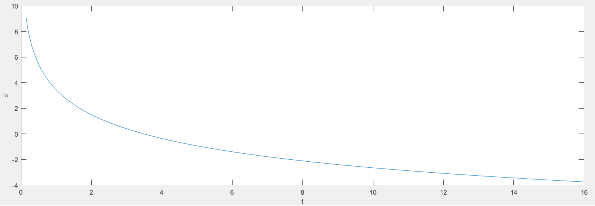
Пример

на экспоненциальном конусе $K_{\exp}$

$$F_{can}(x,y,z) = -\log y - 2\log z + \phi(\log\frac{y}{z} - \frac{x}{z})$$

$\phi : \mathbb{R}_{++} \to \mathbb{R}$ задана неявно кривой

$$\left\{ \begin{pmatrix} t \\ \phi \end{pmatrix} = \frac{1}{2}\begin{pmatrix} \log(1+\kappa) + 2\kappa \\ \log(1+\kappa) - 3\log\kappa \end{pmatrix} \;\middle|\; \kappa \in \mathbb{R}_{++} \right\}$$



История конического программирования

LP: Simplex method
[Dantzig 1951], exp. compl.

Ellipsoid method
[Yudin, Nemirovski 1976]
polynomial-time

LP: Interior-point
projective scaling
[Karmarkar 1984]
polynomial-time

LP: Interior-point
affine scaling
[Dikin 1967]
rediscovery 1986

General cones: IP
[Nesterov, Nemirovski 1988]
self-concordant barriers

LP: Primal-dual IP
[Kojima, Mizuno, Yoshise 1989]
[Monteiro, Adler 1989]
[Todd, Ye 1990]

CP: primal, primal-dual IP
[Nesterov, Nemirovski 1994]
systematic approach
Universal barrier

Symmetric cones IP
[Nesterov, Todd 1994]
self-scaled barriers

Symmetric cones IP
Euclidean Jordan algebras
[Faybusovich 1995]

Classification of self-scaled barriers
[Hauser 1999, 2000]
[Hauser, Güler 2002]
[Hauser, Lim 2002]
[Schmieta 2000]

Спасибо за внимание