

Коническая оптимизация и самосогласованные барьеры

Roland Hildebrand

LJK, Université Grenoble Alpes / CNRS

Мат. кружок МФТИ
19 апреля 2019 г.

1 Введение

- Мотивация
- Принципы решения выпуклых задач
- Методы внутренней точки

2 Конические программы

- Двойственность
- Логарифмично однородные барьеры

3 Геометрическая интерпретация

Приложения оптимизации



задачи оптимизации возникают в самых различных областях

- оптимизация формы,
- планирование траекторий движения,
- дизайн экспериментов,
- оценка параметров, ...

нужно принять обусловленное решение для минимизации цены

Формализация проблемы оптимизации

выбрать значения **параметров** минимизирующих **целевую функцию** при **ограничениях**

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

- переменные $x \in \mathbb{R}^n$
- целевая функция $f : \mathbb{R}^n \rightarrow \mathbb{R}$
- допустимое множество $X \subset \mathbb{R}^n$

часто формализация также нетривиальная проблема
необходимо иметь в виду решаемость получаемой задачи
оптимизации

Представления и вычислимость

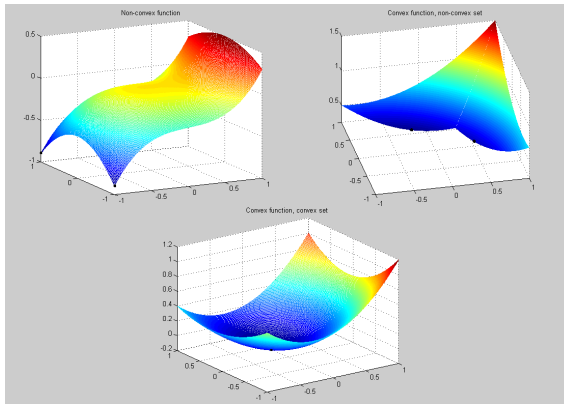
допустимое множество может быть представлено в виде

- равенств $g_i(x) = 0$,
- неравенств $h_j(x) \leq 0$,
- матричных неравенств $A_k(x) \succeq 0$,
- оракулом (черный / серый ящик), ...

целевая функция может быть задана

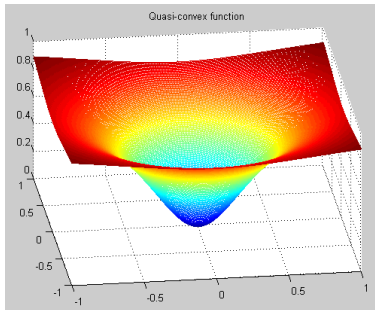
- аналитически,
- оракулом,
- с (суб-)градиентом,
- с гессианом, ...
- данные могут быть с погрешностью / случайные

Выпуклые vs. невыпуклые задачи



если f и X выпуклы, то задача оптимизации называется *выпуклой*

Скрытая (не-)выпуклость



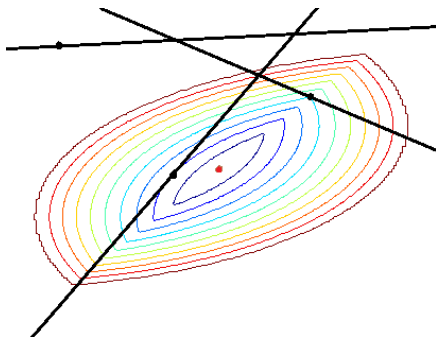
функция, поверхности уровня
которой ограничивают
выпуклые множества,
называется *квази-выпуклой*

выпуклое множество со
скрытой невыпуклостью:
коположительный конус

$$\mathcal{C}_n = \{A \in \mathcal{S}_n \mid x^T A x \geq 0 \ \forall x \geq 0\}$$

[Murty, Kabadi, 1987]:
определить принадлежность к
коположительному конусу
является **ко-NP-полной**
проблемой

Отделяющая гиперплоскость



выпуклая задача: для любой точки можно на основе *локальной* информации вычислить гиперплоскость, которая отделяет точку от *глобального* минимума

пример:

$$\hat{X} \not\geq 0, u^T \hat{X} u < 0$$

$\Rightarrow \langle uu^T, X \rangle > \langle uu^T, \hat{X} \rangle$ для всех $X \succeq 0$

Метод эллипсоидов

задача минимизации выпуклой функции на компактном выпуклом множестве C

начальные данные: эллипсоид

$$E_0 = \{x \in \mathbb{R}^n \mid (x - x_0)^T P_0^{-1} (x - x_0) \leq 1\}$$

содержащий множество C

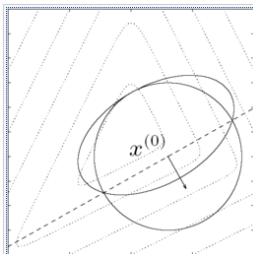
на k -ом шаге:

вычисляем отделяющую гиперплоскость с нормалью g_{k+1}

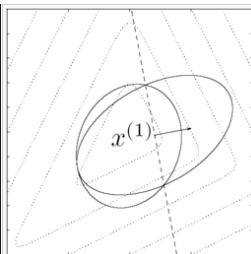
такую, что $\langle g_{k+1}, x_k - x^* \rangle \geq 0$

вычисляем новый эллипсоид E_{k+1} по формулам

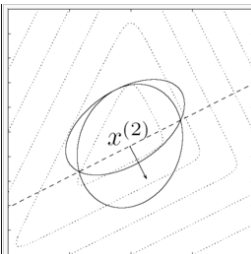
$$x_{k+1} = x_k - \frac{1}{n+1} \frac{P_k g_{k+1}}{\sqrt{g_{k+1}^T P_k g_{k+1}}}, \quad P_{k+1} = \frac{n^2}{n^2 - 1} \left(P_k - \frac{2}{n+1} \frac{P_k g_{k+1} (P_k g_{k+1})^T}{g_{k+1}^T P_k g_{k+1}} \right)$$



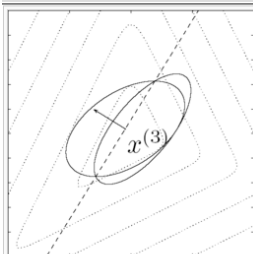
$k = 0$



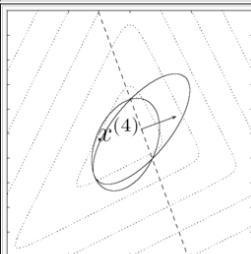
$k = 1$



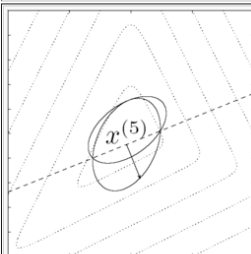
$k = 2$



$k = 3$



$k = 4$



$k = 5$

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

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

точный минимум достигается при $\gamma = 1$

метод аффинно инвариантен

как измерить прогресс, сделанный на данном шаге?

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

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 $f : D \rightarrow \mathbb{R}$ 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 $\rho < 1$, то полный шаг Ньютона не выведет из области D
иначе необходимо укоротить шаг, $\gamma \lesssim \rho^{-1}$

Поведение метода Ньютона

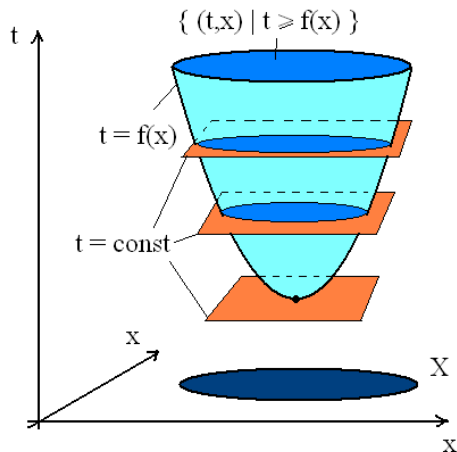
для глобальной сходимости метода Ньютона следует положить

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

- вдалеке от минимума гарантировано понижение значения функции на ≈ 1
- в окрестности минимума находимся в быстром (квадратичном) режиме сходимости

[Нестеров, Немировский 1994]

Линейная функция цены



$f(x)$ можно
предположить линейной

в противном случае
минимизируем t по
надграфику

Самосогласованные барьеры

пусть $X \subset \mathbb{R}^n$ – допустимое множество, $\langle c, x \rangle$ – целевая функция

самосогласованным барьером на выпуклом множестве X называется C^3 функция $F : X^\circ \rightarrow \mathbb{R}$, 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 $x \in X^\circ$, $h \in T_x X$
- $F'(x)[h] \leq \sqrt{\nu F''(x)[h, h]}$ для всех $x \in X^\circ$, $h \in T_x X$

условия гарантируют $\rho \leq \sqrt{\nu}$

Методы внутренней точки

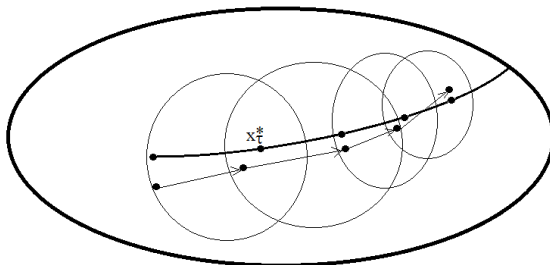
пусть $X \subset \mathbb{R}^n$ – допустимое множество, $\langle c, x \rangle$ – целевая функция

пусть F самосогласованный барьер на X

для $\tau \geq 0$ задача $\min_x (F(x) + \tau \langle c, x \rangle)$ *без ограничений*
неограничена или имеет решение x_τ^*

при $\tau \rightarrow +\infty$ решение x_τ^* стремится к x^*

вокруг каждого x_τ^* имеется зона квадратической сходимости



метод попеременно делает шаг Ньютона и повышает τ , все время находясь в зоне квадратической сходимости вокруг x_T^*
итерации не выходят из окрестности *центрального пути*
радиус окрестности и длина шага $\sim \nu^{-1/2}$

Конические программы

Определение

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

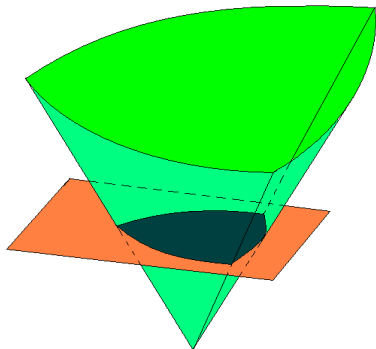
Определение

***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

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

Двойственная задача

Определение

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

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

прямая задача

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

дуализируем условие $Ax = b$

$$\min_{x \in K} \max_z (\langle c, x \rangle - \langle z, Ax - b \rangle)$$

переставляем минимум и максимум

$$\max_z \min_{x \in K} (-\langle A^T z - c, x \rangle + \langle b, z \rangle)$$

минимизация по x приводит к **двойственной** задаче

$$\max_{s = -(A^T z - c) \in K^*} \langle b, z \rangle$$

значение двойственной задачи не больше значения прямой

$$\min_x \max_s f(x, s) \geq \max_s \min_x f(x, s)$$

переставляем минимум и максимум

$$\max_z \min_{x \in K} (-\langle A^T z - c, x \rangle + \langle b, z \rangle)$$

минимизация по x приводит к **двойственной** задаче

$$\max_{s = -(A^T z - c) \in K^*} \langle b, z \rangle$$

значение двойственной задачи не больше значения прямой

$$\min_x \max_s f(x, s) \geq \max_s \min_x f(x, s)$$

Аффинные подпространства

прямое аффинное подпространство

$$P_A = \{x \mid Ax = b\} \subset \mathbb{R}^n$$

$\dim P_A = k$, $n - k$ кол-во строк A

двойственное аффинное подпространство

$$D_A = \{s \mid \exists z : s = -(A^T z - c)\} \subset \mathbb{R}_n$$

$$\dim D_A = n - k$$

$$\dim P_A + \dim D_A = n, \quad V(D_A) = (V(P_A))^\perp$$

введем аффинное подпространство $\mathcal{A} = P_A \times D_A$, $\dim \mathcal{A} = n$

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 $F : K^\circ \rightarrow \mathbb{R}$ на внутренности K , удовлетворяющая условиям

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

для всех касательных векторов h в каждой точке $x \in K^\circ$.

Параметр однородности ν называется **параметром** барьера.

Двойственный барьер

пусть $f : D \rightarrow \mathbb{R}$ – 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 $-K^$.*

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.

Классы регулярных конусов

в оптимизации встречаются

- симметрические конуса
- конуса положительных отображений
- конуса моментов
- конуса положительных полиномов
- конуса сумм квадратов
- коположительные конуса

Симметрические конуса

Определение

*Самодвойственный однородный конус называется
симметрическим.*

[Винберг, 1960; Koecher, 1962] любой симметрический конус является произведением конечного числа следующих неприводимых симметрических конусов:

- конус Лоренца
$$L_n = \left\{ (x_0, \dots, x_{n-1}) \mid x_0 \geq \sqrt{x_1^2 + \dots + x_{n-1}^2} \right\}, \quad n \neq 2$$
- матричный конус $S_+(n)$, $H_+(n)$, $Q_+(n)$ вещественных, комплексных или кватернионных эрмитовых положительно определенных матриц, $n \geq 3$
- конус Альберта $O_+(3)$ октонионных эрмитовых положительно определенных матриц размера 3×3

Примеры

для классических задач конического программирования
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	S_+^n	$-\log \det A$	n

наличие вычисляемого барьера на конусе K определяет
решаемость задачи

Программы над симметрическими конусами

симметрические программы эффективно решаются методами внутренней точки [Nesterov, Nemirovski, 1994]

- 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 $S_+(n)$: $\sim 10^3$ переменных

наличие структуры позволяет решать большие задачи

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

Точки шкалировки

Определение

Пусть $K \subset \mathbb{R}^n$ регулярный конус, а F – самосогласованный барьер на K с параметром ν . Барьер F называется **авто-шкалированным** если для всех $x, w \in K^\circ$ справедливо

$$s = F''(w)x \in \text{int } K^*, \quad F^*(s) = F(x) - 2F(w) - \nu.$$

Конус K , на котором существует авто-шкалированный барьер, называется **авто-шкалированным** конусом.

Hauser, Güler, Lim: авто-шкалированный $\Leftrightarrow$ симметрический

Nesterov, Todd: для любой пары $(x, s) \in (K \times K^*)^\circ$ существует единственная **точка шкалировки** $w \in K^\circ$ такая, что

$$F''(w)x = s$$

Граф преобразования Лежандра

преобразование Лежандра нелинейно, поэтому его граф

$$\Gamma = \{(x, p) \mid p = -F'(x)\}$$

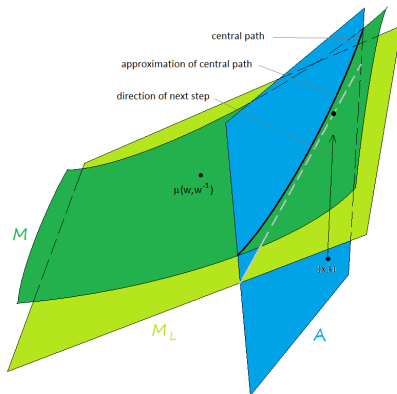
нелинейное подмногообразие произведения $\mathbb{R}^n \times \mathbb{R}_n$
 $\dim \Gamma = n$

введем еще

$$M = \{(x, \mu p) \mid p = -F'(x), \mu > 0\} = \mathbb{R}_{++} \times \Gamma$$

$$\dim M = n + 1$$

Методы с авто-шкалировкой



M_L – линейная аппроксимация M в точке шкалировки

Структура произведения пространств

ни $\mathbb{R}^n$ ни двойственное пр-во $\mathbb{R}_n$ не обладают канонической метрикой

но произведение $\mathbb{R}^n \times \mathbb{R}_n$ – носитель нескольких структур

- плоская псевдо-риманова метрика
$$G((x, p); (y, q)) = \frac{1}{2}(\langle x, q \rangle + \langle y, p \rangle)$$
- $\text{dist}((x, p); (y, q)) = \langle x - y, p - q \rangle$
- симплектическая форма $\omega((x, p); (y, q)) = \frac{1}{2}(\langle x, q \rangle - \langle y, p \rangle)$
- инволюция $J : (x, p) \mapsto (x, -p)$ с интегрируемыми собственными распределениями
- условия совместимости $\hat{\nabla} \omega = 0, G = \omega J$

$\mathbb{R}^n \times \mathbb{R}_n$ – плоское однородное пара-Кэлерово пр-во

Γ как лагранжево подмногообразие

пусть $\mathcal{D} : x \mapsto -F'(x)$ – граф преобразования Лежандра для барьера с параметром ν

- $\Gamma(\mathcal{D})$ – лагранжево подмногообразие $\mathbb{R}^n \times \mathbb{R}_n$
- метрика на $\Gamma(\mathcal{D})$ равна $\nu \times$ индуцированную на подмногообразии метрику
- (внешняя) кривизна $\Gamma(\mathcal{D})$ ограничена $\sqrt{\nu}$

Точка шкалировки как ближайшая точка

точка шкалировки, заданная условием

$$F''(w)x = s$$

является ближайшей к (x, s) точкой на M в псевдо-римановой метрике произведения:

$$\min_{w \in K^o} \langle x - \mu w, s + \mu F'(w) \rangle$$

производная по w приводит к условию

$$-\mu(s + \mu F'(w)) + \mu F''(w)(x - \mu w) = 0$$

красные члены сокращаются, поскольку $F''(w)w = -F'(w)$

похожая проблема рассмотрена в [Nesterov, Todd 1997]

Спасибо за внимание