Полу-определённое программирование
Методы внутренней точки

Roland Hildebrand

LJK, Université Grenoble Alpes / CNRS

Методы оптимизации, ФУПМ МФТИ, апрель 2021 г.

Определение

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

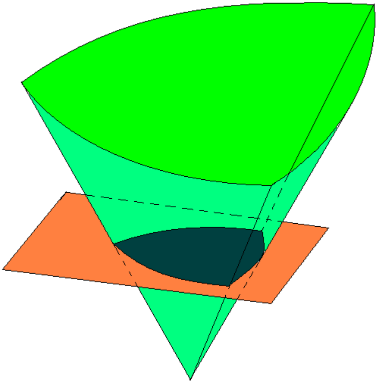
Определение

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

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

переменная параметризует не конус, а допустимое множество

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)
- степенные конуса (ограничения на p-норму)

Симметрические конуса

симметрические конуса, использующиеся в оптимизации:

- ортант $\mathbb{R}^n_+ = \left\{ (x_1, \dots, x_n)^T \mid x_i \geq 0 \right\}$
- 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 определённых матриц
- матричный конус $\mathcal{H}^n_+$ комплексных эрмитовых неотрицательно определённых матриц
- их прямые произведения

симметрические конуса обладают дополнительной структурой, позволяющей применять более эффективные алгоритмы

Классы конических программ

программы над симметрическими конусами

- 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-определённые программы (SDP) над $\mathcal{S}^n_+$: $\sim 10^3$ переменных

наличие структуры позволяет решать бóльшие задачи

$LP \subset SOCP \subset SDP$

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

Примеры несимметрических конусов

экспоненциальный конус

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

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

степенной конус

$$K_p = \{(x,y,z) \mid |z| \leq x^{1/p}y^{1/q},\ x \geq 0,\ y \geq 0\}$$

где $\frac{1}{p} + \frac{1}{q} = 1,\ p,q \in [1,+\infty]$
ограничение $x \geq |z|^p$ для $p \geq 1$ записывается в виде
$(x,1,z) \in K_p$

Определение

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

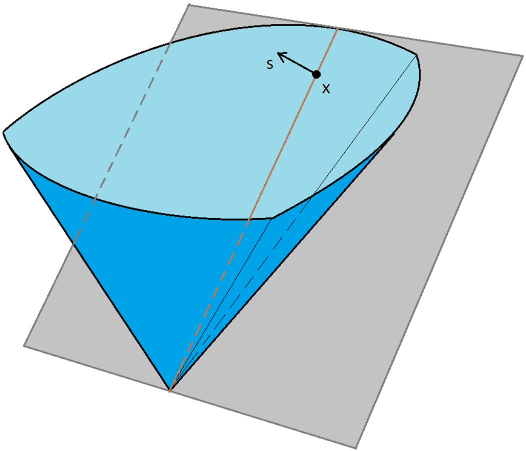
$$K^* = \{s \in \mathbb{R}_n = (\mathbb{R}^n)^* \,|\, \langle s, x\rangle \geq 0 \quad \forall\, x \in K\}\,.$$

двойственный конус определён в *двойственном* векторном пространстве

если в пространстве задано *скалярное произведение*, то двойственное можно отождествить с прямым пространством

для регулярного K конус K^* также является регулярным, и $(K^*)^* = K$

Как вычислять двойственный конус



элементы границы ∂K^* двойственного конуса являются нормалями к подпирающим K плоскостям

Примеры (само)двойственных конусов

отождествляя $\mathbb{R}^n$ с $\mathbb{R}_n$ через скалярное произведение $\langle x, y \rangle = x^T y$, получаем

- $(\mathbb{R}^n_+)^* = \mathbb{R}^n_+$
- $L_n^* = L_n$

отождествляя $\mathcal{S}^n$ с $(\mathcal{S}^n)^*$ $(\mathcal{H}^n$ с $(\mathcal{H}^n)^*)$ через скалярное произведение $\langle A, B \rangle = \mathrm{tr}(AB)$, получаем

- $(\mathcal{S}^n_+)^* = \mathcal{S}^n_+$
- $(\mathcal{H}^n_+)^* = \mathcal{H}^n_+$

отождествляя $\mathbb{R}^3$ с $\mathbb{R}_3$ через скалярное произведение $\langle x, y \rangle = x^T y$, получаем

- $K_{\exp}^* = \mathrm{cl}\,\{(-tx, -t, te^{-1}y) \mid t \geq 0,\ y \geq e^x\}$
- $K_p^* = \{(x/p, y/q, z) \mid |z| \leq x^{1/p} y^{1/q},\ x \geq 0,\ y \geq 0\}$

Определение симметрических конусов

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

Определение

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

симметрические конуса полностью классифицированы [Винберг, 1960; Koecher, 1962]

- ортант $\mathbb{R}^n_+$
- конус Лоренца L_n
- матричные конуса $\mathcal{S}^n_+, \mathcal{H}^n_+, \mathcal{Q}^n_+, \mathcal{O}^3_+$ (вещественные, комплексные, кватернионные, октонионный)
- прямые произведения $\prod_{i=1}^m K_i$

дана 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 множества имеем

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

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 $r + L$

$L = \ker A \subset \mathbb{R}^n$ — линейное подпространство, $Ar = b$

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

прямая программа

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

тогда

$$\langle b, y \rangle = \langle Ar, y \rangle = \langle r, A^T y \rangle = \langle r, c - s \rangle = -\langle r, s \rangle, \quad L^\perp = Im\, A^T$$

двойственная программа

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

Принцип метода внутренней точки

сложность конической программы в ограничении $x \in K$

устраняем это условие посредством добавления строго выпуклой барьерной функции $F : K^o \to \mathbb{R}$ к функции цены $\lim_{x \to \partial K} F(x) = +\infty$ — свойство барьера

вместо исходной задачи

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

получаем 1-параметрическое семейство задач

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

$\tau > 0$ — параметр семейства

Принцип метода внутренней точки

пусть решение исходной задачи существует, тогда

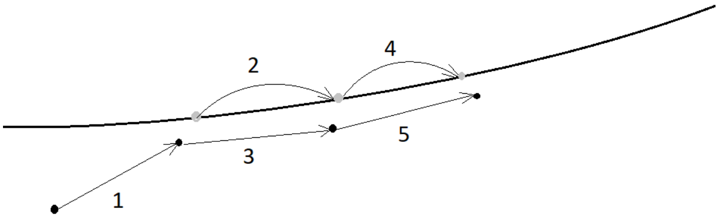
- при достаточно больших τ точки минимума $x^*(\tau)$ вспомогательной задачи существуют и единственны
- при $\tau \to +\infty$ решение $x^*(\tau)$ стремится к некоторому x^* в относительной внутренности множества решений

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

прямой метод следования центральному пути перемежает шаг по направлению к минимуму $x^*(\tau)$ вспомогательной задачи и увеличение переметра τ этой задачи

для шага минимизации используют метод Ньютона

Принцип метода внутренней точки



серые: целевые точки на центральном пути
чёрные: итерации в прямом пространстве

обновление целевой точки перемежается с шагом Ньютона по направлению к текущей целевой точке

про целевую точку известно, что она минимизирует вспомогательную функцию цены $F(x) + \tau\langle c, x\rangle$ на $\{x \mid Ax = b\}$

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

на k-ом шаге: аппроксимируем

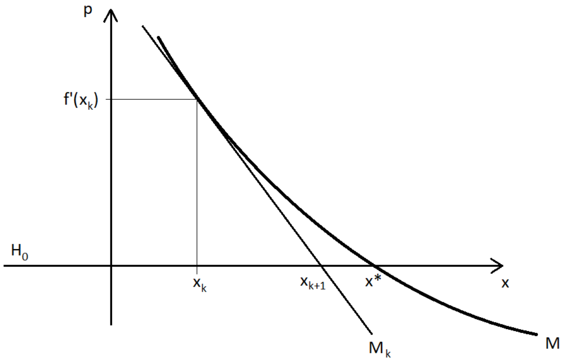
$$f(x) \approx q_k(x) = f(x_k) + \langle f'(x_k), x - x_k \rangle + \frac{1}{2}\langle f''(x_k)(x - x_k), x - x_k \rangle$$

минимизируем квадратичную аппроксимацию q_k

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

γ_k — коэффициент затухания (damping coefficient)
точный минимум функции q_k достигается при $\gamma_k = 1$

Геометрическая интерпретация



M — граф градиента f, M_k — граф градиента q_k
аппроксимируем M касательной плоскостью M_k в точке $(x, f'(x))$

Аффинная инвариантность

пусть $A \in \mathbb{R}^{n\times n}$ обратима, $b \in \mathbb{R}^n$, и

$$x \mapsto \tilde{x} = Ax + b, \quad x = A^{-1}(\tilde{x} - b)$$

минимизируем $\tilde{f}(\tilde{x}) = f(x(\tilde{x})) = f(A^{-1}(\tilde{x} - b))$ шагом Ньютона

- $\tilde{f}'(\tilde{x}) = A^{-T} f'(x)$
- $\tilde{f}''(\tilde{x}) = A^{-T} f''(x) A^{-1}$

если $\tilde{x}_k = Ax_k + b$, то следующая точка

$$\begin{aligned}
\tilde{x}_{k+1} &= \tilde{x}_k - \gamma_k (\tilde{f}''(\tilde{x}_k))^{-1} \tilde{f}'(\tilde{x}_k) \\
&= Ax_k + b - \gamma_k (A^{-T} f''(x_k) A^{-1})^{-1} A^{-T} f'(x_k) \\
&= Ax_k + b - \gamma_k A (f''(x_k))^{-1} f'(x_k) = Ax_{k+1} + b
\end{aligned}$$

и последовательность итераций эквивариантна по отношению к аффинным преобразованиям

Ньютоновский декремент

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

ρ_k измеряет, насколько далеко точка x_k находится от минимума, $\rho_k = 0$ в точке минимума

ρ_k аффинно инвариантен

в идеальном случае, когда $f = q_k$, имеем $\rho_{k+1} = 0$

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

итерация может вывести из множества D

Самосогласованные функции

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

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*

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

Свойства

операции, сохраняющие (сильную) самосогласованность

- $f(x)$ с.-с. $\Rightarrow g(x) = f(Ax + b)$ с.-с.
- f с.-с. $\Rightarrow f + l$ с.-с. для линейных l
- f с.-с., L аффинное подпр-во $\Rightarrow f|_L$ с.-с.
- f с.-с. $\Rightarrow \alpha f$ с.-с. для $\alpha \geq 1$
- f, g с.-с. $\Rightarrow f + g$ с.-с.
- f с.-с. $\Rightarrow$ сопряжённая по Лежандру f^* с.-с.

сопряжённая по Лежандру определена как

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

можно ослабить условие на третью производную и потребовать только $f \in C^2(D)$ и

$$\limsup_{t \to 0} \frac{|f''(x+th)[h,h] - f''(x)[h,h]|}{t} \leq 2(f''(x)[h,h])^{3/2}$$

Эллипсоид Дикина

пусть $x \in D$, h — касательный вектор единичной длины в $\|\cdot\|_x$
обозначим $\sigma(t) = f''(x+th)[h,h]$, тогда $\sigma(0) = 1$
по условию самосогласованности $|\dot{\sigma}| \leq 2\sigma^{3/2}$, откуда
$\frac{1}{(1+t)^2} \leq \sigma(t) \leq \frac{1}{(1-t)^2}$

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

Пусть f — сильно самосогласованная функция на области D. Тогда D содержит открытый единичный шар в локальной метрике (эллипсоид Дикина)

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

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

иначе необходимо укоротить шаг: $\gamma_k < \rho_k^{-1}$

Поведение декремента

в методах внутренней точки используются *самосогласованные* барьеры F

все вспомогательные функции $F(x) + \tau\langle c, x\rangle$ также самосогласованы на множестве допустимых точек

анализ сходимости основан на изучении эволюции декремента

- декремент зависит не только от точки $x \in D$, но и от τ
- шаг Ньютона приближает текущую точку к минимуму и *уменьшает* декремент
- обновление τ удаляет минимум от текущей точки и *увеличивает* декремент
- больший выигрыш на шаге позволяет сильнее увеличить τ
- декремент колеблется в диапазоне, позволяющем увеличивать τ с максимальной скоростью

Известные оценки: полный шаг

оценки основаны на соотношении [Ю.Е. Нестеров, А.С. Немировский 1994]

$$(1 - ||x - x_k||_{x_k})^2 F''(x_k) \preceq F''(x) \preceq (1 - ||x - x_k||_{x_k})^{-2} F''(x_k)$$

для самосогласованных функций

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

Пусть F — сильно самосогласованная функция на D. Если $\rho_k < \lambda^* = \frac{3-\sqrt{5}}{2} \approx 0.3820$, то после полного шага Ньютона имеем

$$\rho_{k+1} \leq \left(\frac{\rho_k}{1 - \rho_k}\right)^2 < \rho_k.$$

Известные оценки: укороченный шаг

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

Пусть F *— сильно самосогласованная функция на* D*. Если* $\rho_k < \lambda^* = \frac{\sqrt{5}-1}{2} \approx 0.6180$*, то после укороченного шага Ньютона с коэффициентом* $\gamma_k = \frac{1}{1+\rho_k}$ *имеем*

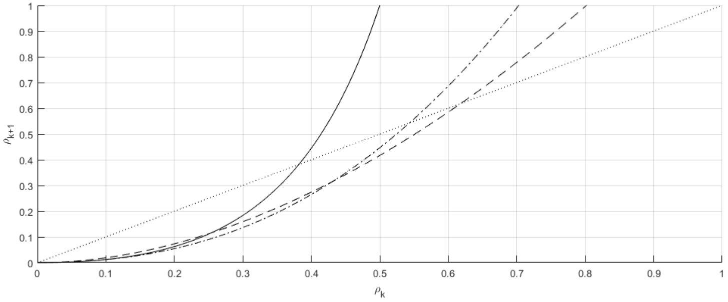
$$\rho_{k+1} \leq \frac{\rho_k^2(2+\rho_k)}{1+\rho_k} < \rho_k.$$

Теорема (Ю.Е. Нестеров 2018)

Пусть F *— сильно самосогласованная функция на* D*. Если* $\rho_k < \lambda^* = roots(\lambda^3 + \lambda^2 + \lambda - 1) \approx 0.5437$*, то после укороченного шага Ньютона с коэффициентом* $\gamma_k = \frac{1+\rho_k}{1+\rho_k+\rho_k^2}$ *имеем*

$$\rho_{k+1} \leq \rho_k^2\left(1+\rho_k+\frac{\rho_k}{1+\rho_k+\rho_k^2}\right) < \rho_k.$$

верхние оценки на ρ_{k+1}



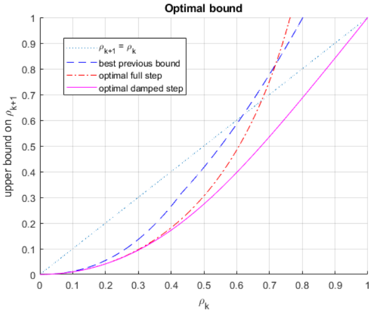
сплошная — полный шаг $\gamma_k = 1$

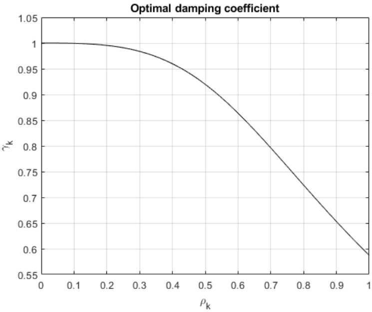
штриховая — укороченный шаг $\gamma_k = \frac{1}{1+\rho_k}$

штрих-пунктирная — укороченный шаг $\gamma_k = \frac{1+\rho_k}{1+\rho_k+\rho_k^2}$

Оптимальные оценки

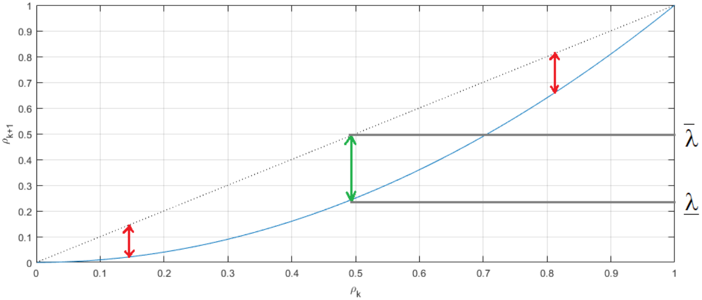
оптимальные оценки можно получить с помощью теории оптимального управления





длина оптимального коэффициента затухания зависит от ρ_k
достаточно находиться в эллипсоиде Дикина, чтобы гарантировано понизить декремент

Диапазон быстрейшей сходимости



слишком маленький или большой декремент в исходной точке шага Ньютона приводит к малому прогрессу

оптимальное значение $\overline{\lambda}$ максимизирует $\rho_{k+1} - \rho_k$

обновляем параметр τ таким образом, чтобы декремент стал равным $\overline{\lambda}$

Алгоритм

- стартуем с пары $(x_0, \tau_0) \in X \times \mathbb{R}_{++}$, так что $\rho_0^0 = \overline{\lambda}$
- делаем шаг Ньютона по направлению к решению $x^*(\tau_0)$, получаем x_1, декремент принимает значение ρ_1^0
- обновляем значение параметра на $\tau_1 > \tau_0$, так что $\rho_1^1 = \overline{\lambda}$
- переход к следующей итерации

на каждом шаге выполняются соотношения

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

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

потребуем ещё одно условие от барьера:
пусть существует константа ν такая, что

$$||F'(x)||_x^2 = (F'(x))^T (F''(x))^{-1} F'(x) \leq \nu \qquad \forall\, x \in D$$

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* $x \in X^o$, $h \in T_x X^o$
- $F'(x)[h] \leq \sqrt{\nu F''(x)[h,h]}$ *для всех* $x \in X^o$, $h \in T_x X^o$

Обновление τ

в текущей точке x_k декремент как функция от τ задаётся гиперболой

$$\rho(\tau) = ||F'(x_k)+\tau c||_{x_k} = \sqrt{(F'(x_k) + \tau c)^T (F''(x_k))^{-1} (F'(x_k) + \tau c)}$$

отсюда получаем оценку на производную

$$\frac{d\rho}{d\tau} \leq \sqrt{c^T (F''(x_k))^{-1} c} = ||c||_{x_k}$$

эквивалентно

$$\frac{d\rho}{d \log \tau} \leq \tau ||c||_{x_k}$$

оценка шага по τ сверху

$$\tau_{k+1} - \tau_k \geq \frac{\rho_{k+1}^{k+1} - \rho_{k+1}^{k}}{||c||_{x_k}} \geq \frac{\bar{\lambda} - \underline{\lambda}}{||c||_{x_k}}$$

для больших τ получаем

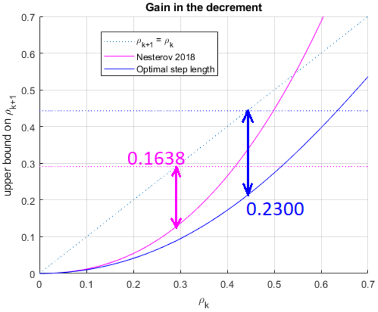
$$F'(x_k) \approx -c\tau_k, \quad ||c||_{x_k} \approx \frac{||F'(x_k)||_{x_k}}{\tau_k} \leq \frac{\sqrt{\nu}}{\tau_k}$$

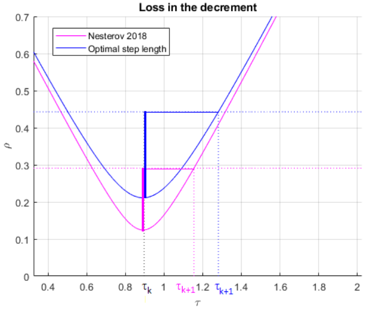
отсюда $\frac{\tau_{k+1}-\tau_k}{\tau_k} \approx \log \frac{\tau_{k+1}}{\tau_k} \sim \frac{\bar{\lambda}-\underline{\lambda}}{\sqrt{\nu}}$

меньший параметр барьера соответствует большей скорости сходимости

генерируемые точки находятся в $O(1)$ окрестности центрального пути (в локальной норме)

Настройка параметров метода





больший выигрыш в ρ на шаге Ньютона позволяет

- делать больший шаг $\tau_{k+1} - \tau_k$ вдоль центрального пути
- увеличить окрестность центрального пути, в которой генерируются точки

Итоги

метод следования центральному пути с коротким шагом

- линейная скорость сходимости
- на каждом шаге можно увеличивать $\log \tau$ на величину порядка $\nu^{-1/2}$
- чем больше параметр ν, тем медленнее метод будет сходиться
- сложность задачи зависит от наличия эффективно вычислимого самосогласованного барьера с небольшим значением параметра
- возможность отступать дальше от центрального пути увеличивает скорость

методы с *длинным* шагом используют дополнительную структуру

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

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

- $F(\alpha x) = -\nu \log \alpha + F(x)$ *(логарифмичная однородность)*
- $F''(x) \succ 0$ *(локально строгая 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 барьер

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

преобразованием Лежандра функции f называется функция

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

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 $-K^*$.

двойственный барьер определим как
$F_*(s) = F^*(-s) = \sup_{x \in K} (-\langle s, x \rangle - F(x))$
имеем $F_*(\tau x) = \tau^{-1} F_*(x)$ для всех $\tau > 0$, $x \in K^o$

положительно определённый 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. Изометрия действует на тензор третьих производных барьера умножением на -1.

для классических задач используются следующие барьеры

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

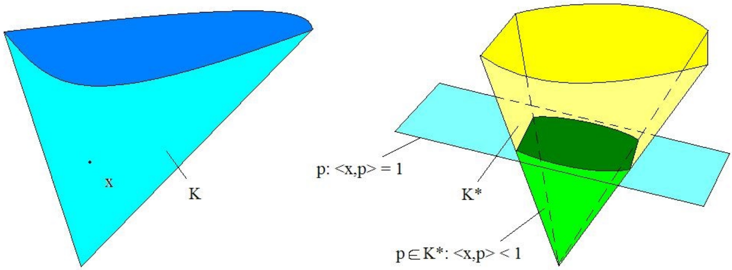
Экспоненциальный конус

$$K_{\exp} = \left\{(x,y,0) \,|\, x \le 0,\ \ y \ge 0\right\} \cup \left\{(x,y,z) \,|\, z > 0,\ \ y \ge z e^{x/z}\right\}$$

на экспоненциальном конусе имеется барьер

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

значение параметра $\nu = 3$ также оптимально пусть K — произвольный регулярный выпуклый конус



функция объёма $V : K^o \ni x \mapsto Vol\{p \in K^* \mid \langle x, p \rangle < 1\}$
определена с точностью до множителя

Универсальный барьер

Теорема (Ю.Е. Нестеров, А.С. 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 $\nu = c \cdot n$. Этот барьер называется универсальным.

позже было установлено, что можно выбрать $c = 1$ [Bubeck, Eldan 2015]

Свойства

универсальный барьер F

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

для неоднородных конусов трудно вычислим
не эквивариантен по отношению к двойственности

двойственный к универсальному называется *энтропическим*
[Bubeck, Eldan '15]

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

канонический барьер

- обладает теми же свойствами инвариантности что и универсальный
- эквивариантен по отношению к двойственности
- совпадает с универсальным (и энтропическим) на однородных конусах (в том числе на симметрических)
- вычислим на некоторых неоднородных конусах с большой группой симметрий

на симметрических конусах все три барьера задаются стандартным логарифмическим барьером

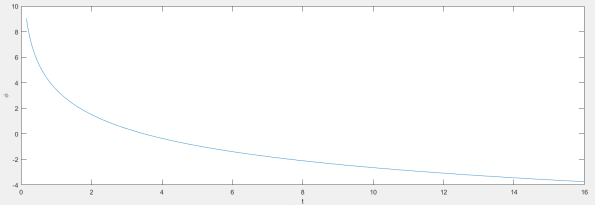
Пример канонического барьера

на экспоненциальном конусе $K_{\exp}$

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

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

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



Прямо-двойственные методы

прямая программа

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

двойственная программа

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

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

прямо-двойственные методы удобно анализировать в произведении $\mathbb{R}^n \times \mathbb{R}_n$ прямого и двойственного пространств

Объекты в $\mathbb{R}^n \times \mathbb{R}_n$

- прямое аффинное подпространство $P = \{x \,|\, Ax = b\} \subset \mathbb{R}^n$, $\dim P = k$, $n - k$ кол-во строк A
- двойственное аффинное подпространство $D = \{s \,|\, \exists\, z : \; s = -(A^T z - c)\} \subset \mathbb{R}_n$, $\dim D = n - k$
- произведение $\mathcal{A} = P \times D$, $\dim \mathcal{A} = n$
- граф лежандровой изометрии $M = \{(x, -F'(x)) \,|\, x \in K^o\}$, $\dim M = n$

прямой центральный путь представится в виде прямой компоненты пересечения $M \cap (P \times \mathbb{R} \cdot D)$

двойственный центральный путь представится в виде двойственной компоненты пересечения $M \cap (\mathbb{R} \cdot P \times D)$

Прямо-двойственный центральный путь

точку $(x, \mu s) \in M \cap (P \times \mathbb{R} \cdot D)$ можно соотнести с точкой $(\mu x, s) \in M \cap (\mathbb{R} \cdot P \times D)$

это определяет каноническую биекцию между прямым и двойственным центральными путями

определим кривую

$$\mathbb{R} \cdot M \cap \mathcal{A} = \{(\sqrt{\mu}x, \sqrt{\mu}s) \in M \mid (x, s) \in \mathcal{A}\}$$

как *прямо-двойственный центральный путь*

параметр μ определяется из параметра τ прямого пути как $\mu = \tau^{-1}$

- итерации (x_k, s_k) состоят из прямой и двойственной компонент
- перемежаем обновление параметра τ (или μ) на прямо-двойственном центральном пути и шаг Ньютона по направлению к соответствующей точке
- разные способы линеаризации нелинейной системы, определяющей центральный путь, приводят к разным *направлениям поиска* (search direction)
- последовательность зазоров $\langle x_k, s_k\rangle$ 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

Авто-шкалированные барьеры

Определение

Пусть $K \subset \mathbb{R}^n$ — регулярный выпуклый конус, K^* двойственный к нему, F — самосогласованный барьер на K с параметром ν, F_* — двойственный к нему барьер на K^*. Тогда F называется **авто-шкалированным** *если для всех $x, w \in K^o$ справедливо*

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

Конус K, допускающий авто-шкалированный барьер, называется **авто-шкалированным**.

[Ю.Е. Нестеров, М. Тодд, 1996]: для любой пары $(x, s) \in (K \times K^*)^o$ существует единственная **точка шкалировки** $w \in K^o$ такая, что

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

Классификация

Hauser, Güler, Lim, Schmieta 1998 – 2002:

- авто-шкалированный конус $\Leftrightarrow$ симметрический конус
- авто-шкалированные барьеры на произведениях конусов являются взвешенными суммами авто-шкалированных барьеров на неприводимых факторах
- авто-шкалированные барьеры на неприводимых конусах являются логарифмами детерминантов

Примеры

точка шкалировки для стандартных барьеров

LP: $K = \mathbb{R}^n_+$, $F = -\sum_{i=1}^n \log x_i$

$$F''(w) = \operatorname{diag}\left(w_1^{-2}, \dots, w_n^{-2}\right), \; w_i^{-2} x_i = s_i, \; w = \left(\sqrt{\frac{x_1}{s_1}}, \dots, \sqrt{\frac{x_n}{s_n}}\right)$$

SDP: $K = \mathcal{S}^n_+$, $F = -\log\det X$

$$F''(W)[X, \cdot] = W^{-1} X W^{-1} = S, \quad W = U \Lambda_X^{1/2} \Lambda_S^{-1/2} U^T,$$

где $X = U \Lambda_X U^T$, $S = U \Lambda_S U^T$ — диагонализирующие разложения X, S

w — геодезическая средняя между x и s^{-1}

Точка шкалировки как ближайшая точка

на произведении $\mathbb{R}^n \times \mathbb{R}_n$ можно ввести *расстояние*

$$d^2((x,s),(x',s')) = \langle x - x', s - s' \rangle$$

точка $(w, -F'(w))$ является *ближайшей* к (x,s) точкой на M:

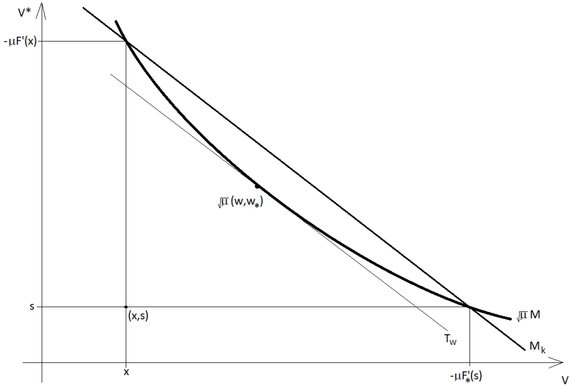
$$\langle x - w, s + F'(w) \rangle \to \max$$

производная по w приводит к условию

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

красные члены сокращаются, поскольку $F''(w)w = -F'(w)$
получаем условие шкалировки $F''(w)x = s$

Направление поиска Нестерова-Тодда



плоскость M_k параллельна к касательной плоскости в точке $(w, -F'(w))$ и проходит через точки $(x, -F'(x))$, $(-F'_*(s), s)$ (существует и единственна для симметрических конусов)

Направление поиска Нестерова-Тодда

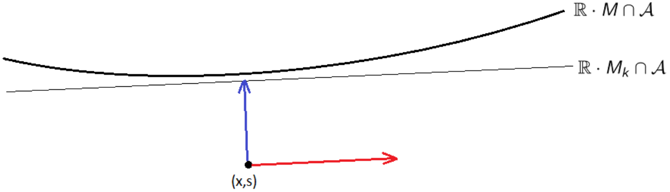
пересечение $\mathbb{R} \cdot M_k \cap \mathcal{A}$ является прямой и аппроксимирует прямо-двойственный центральный путь

направление аффинной шкалировки Нестерова-Тодда есть касательное к этой прямой направление

направление центрирующей шкалировки Нестерова-Тодда есть направление из текущей точки (x_k, s_k) на ближайшую точку этой прямой

направление аффинной шкалировки: продвигается параллельно центральному пути

направление центрирующей шкалировки: корректирует невязку с центральным путём, не улучшая параметр τ



направление аффинной шкалировки
направление центрирующей шкалировки

прямо-двойственный центральный путь $\mathbb{R} \cdot M \cap \mathcal{A}$ аппроксимируется прямой $\mathbb{R} \cdot M_k \cap \mathcal{A}$

Методы с длинным шагом

методы с коротким шагом используют комбинацию аффинной и центрирующей шкалировок

методы с длинным шагом идут по направлению аффинной шкалировки, пока не достигнут границы некоторой «большой» окрестности центрального пути

варианты методов

- предиктор-корректор: после «длинного» шага следует один или несколько корректирующих шагов по центрирующему направлению, пока итерации не приблизятся достаточно к центральному пути
- высшего порядка: аппроксимируют центральный путь полиномом вместо прямой

на практике быстрее методов с коротким шагом, хотя в теории не лучше или даже уступают

Определение «большой» окрестности

LP:

автоморфизм $\mathbb{R}^n_+$ — диагональная матрица D действует на прямо-двойственную пару по формуле $(x, p) \mapsto (Dx, D^{-1}p)$

по-компонентное произведение $y = x \cdot p$ инвариантно

на центральном пути имеем $y = \mu \cdot \mathbf{1}$, в точке решения задачи $y = \mathbf{0}$

произведение прямого и двойственного допустимого множества задачи биективно отображается на $\mathbb{R}^n_{++} \ni y$

«большую» окрестность зададим

$$N(\gamma) = \left\{ y \mid \frac{\max_j y_j}{\min_j y_j} \le \gamma \right\}$$

типичное значение $\gamma \sim 10^3$

Определение «большой» окрестности

SDP:

автоморфизм A конуса $\mathcal{S}_+^n$ действует по формуле

$$(X, P) \mapsto (AXA^T, A^{-T}PA^{-1}), \quad XP \mapsto AXPA^{-1}$$

спектр произведения XP инвариантен по отношению к сопряжению с A

«большую» окрестность зададим

$$N(\gamma) = \left\{ y \mid \frac{\lambda_{\max}(XP)}{\lambda_{\min}(XP)} \le \gamma \right\}$$

История конического программирования

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

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

Ellipsoid method
[Yudin, Nemirovski 1976]
polynomial-time

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

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