1 Методы внутренней точки

1.1 Задача конической оптимизации

В этом курсе мы рассматриваем *выпуклые* конечно-мерные задачи оптимизации. Такая задача может быть сформулирована в виде

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

где $X \subset \mathbb{R}^n$ – непустое выпуклое множество допустимых точек, а $f : X \to \mathbb{R}$ – выпуклая функция цены на X. Во избежание патологических ситуаций мы также будем предполагать функцию f полунепрерывной снизу.

Формально выпуклость определяется следующим образом.

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

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

т.е. интервал $[x, y]$, соединяющий точки x, y, также принадлежит 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*

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

Приведем вышеописанную задачу к некоторому стандартному виду. Без ограничения общности можно считать, что в задаче выпуклой оптимизации выпуклая полунепрерывная снизу функция цены f является линейной, а множество X допустимых точек замкнуто. Действительно, определим

$$\tilde{X} = \mathrm{epi}\ f = \{(t, x) \,|\, x \in X,\ t \geq f(x)\}, \qquad \tilde{f}(t, x) = t.$$

Тогда исходная задача эквивалентна задаче

$$\min_{(t,x) \in \tilde{X}} \tilde{f}(t, x),$$

множество $\tilde{X}$ – выпукло и замкнуто, а функция $\tilde{f}$ – линейна. Приведение f к линейной функции стоило нам повышения размерности множества X на единицу.

Любое замкнутое выпуклое множество X можно представить в виде пересечения замкнутого выпуклого конуса с аффинным подпространством. Действительно, определим множество

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

являющееся замкнутым выпуклым конусом. Операция замыкания приводит только к добавлению точек в подпространстве $\{(t, x) \,|\, t = 0\}$. Поэтому

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

Заменив объемлющее пространство на линейную оболочку конуса K, можно считать, что K имеет непустую внутренность. Если множество X не содержит прямых, то конус K также не будет содержать прямых.

Определение 2. *Выпуклый конус $K \subset \mathbb{R}^n$ называется* регулярным*, если он замкнутый, имеет непустую внутренность, и не содержит прямых.*

Определение 3. *Пусть $K \subset \mathbb{R}^n$ – регулярный выпуклый конус.* Конической программой *над K называется задача оптимизации вида*

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

где $\langle c, x\rangle$ – линейная функция цены, а $Ax = b$ – линейные ограничения типа равенства.

Таким образом, коническая программа является задачей минимизации линейной функции на пересечении регулярного выпуклого конуса с аффинным подпространством. Выше мы показали, что любую задачу выпуклой оптимизации, если она не является вырожденной, можно свести к конической программе над некоторым регулярным выпуклым конусом.

На практике коническая программа часто возникает в виде задачи типа

$$\min_x \langle c, x\rangle \ : \qquad Ax = b, \quad \mathcal{A}(x) \in K,$$

где $\mathcal{A}$ – аффинное отображение. Современные сольверы принимают описание задачи в этой исходной форме и сами переписывают ее в стандартном виде, описанном в Определении 3.

Сложность конической программы зависит исключительно от соответствующего конуса K, так как он заключает в себя все нетривиальные ограничения. Однако, даже некоторые *невыпуклые* задачи можно записать в виде конической программы, если скрыть невыпуклые ограничения в описании конуса K.

Резюме: Конической программой называется задача минимизации линейной функции на регулярном выпуклом конусе с линейными ограничениями типа равенства. Это очень широкий класс задач, практически включающий в себя все выпуклые задачи оптимизации.

1.2 Классы конических программ

Рассмотрим самые распространенные на практике классы конических программ.

Линейные программы (Linear programs, LP): Линейной программой называется задача оптимизации с линейной функцией цены и линейными ограничениями типа равенства и неравенства. Вводом вектора дополнительных неотрицательных переменных $s \geq 0$ можно трансформировать векторно-значное линейное неравенство $Ax \leq b$ в равенство $Ax + s = b$. Используя равенства для выражения исходных переменных x через s мы добиваемся того, что единственными ограничениями типа неравенства остаются соотношения $s \geq 0$. Они эквивалентны коническому ограничению $s \in \mathbb{R}^n_+$, где n – размерность вектора s.

Таким образом, в линейных программах в качестве конуса выступает ортант $\mathbb{R}^n_+$. В стандартной форме линейная программа имеет вид

$$\min_{x \in \mathbb{R}^n_+} \langle c, x\rangle \ : \qquad Ax = b.$$

Ее множеством допустимых точек является полиэдр, а оптимальное значение функции цены, если оно конечное, достигается в вершине этого полиэдра.

Линейные программы являются самым первым классом задач в исследовании операций, для которого был разработан эффективный на практике алгоритм решения, а именно *симплекс-метод* [2]. Этот метод выдает последовательность вершин полиэдра допустимых точек, в которых значение функции цены монотонно убывает, и не относится к классу рассматриваемых в данном курсе методов внутренней точки. Симплекс-метод решает задачу за конечное число шагов. Однако, это число в теории ограничено лишь экспоненциальной функцией от числа ограничений [5].

Квадратично-конические программы (Second order cone programs, SOCP): Квадратично-конической программой называется коническая программа над конусом вида $K = \prod_{i=1}^m L_{n_i}$, где *конус Лоренца* L_n имеет вид

$$L_n = \{(x_0, x_1, \dots, x_{n-1})^T \in \mathbb{R}^n \,|\, x_0 \geq \sqrt{x_1^2 + \dots + x_{n-1}^2}\}.$$

Любую задачу оптимизации с выпуклыми квадратичными ограничениями можно свести к квадратично-конической программе.

Полу-определенные программы (Semi-definite programs, SDP): Полу-определенной программой называется задача минимизации линейной функции цены при ограничениях типа линейных матричных неравенств (linear matrix inequlaity, LMI). Такое ограничение на вектор переменных x имеет вид

$$A(x) \succeq 0,$$

где A – симметрическая матрица, элементы которой являются аффинными функциями от x, а само условие означает, что матрица A неотрицательно определенная, т.е. все ее собственные значения неотрицательны.

Приняв элементы матриц A, фигурирующих в ограничениях, за новые переменные и исключив x с помощью линейных связей между A и x, полу-определенную программу можно записать как коническую программу над конусом вида $K = \prod_{i=1}^m \mathcal{S}_+^{n_i}$, где матричный конус $\mathcal{S}_+^n$ имеет вид

$$\mathcal{S}_+^n = \{A \in \mathcal{S}^n \,|\, A \succeq 0\},$$

где $\mathcal{S}^n$ – векторное пространство вещественных симметрических матриц порядка $n \times n$.

Вышеописанные классы задач удовлетворяют соотношениям включения

$$LP \subset SOCP \subset SDP,$$

т.е. линейную программу можно записать в виде квадратично-конической, а квадратично-коническую – в виде полу-определенной программы.

Действительно, ортант записывается в виде прямого произведения $\mathbb{R}_+^n = L_1^n$ одномерных конусов Лоренца, а n-мерный конус Лоренца L_n можно представить в виде пересечения матричного конуса порядка $n-2$ с линейным подпространством,

$$L_n = \left\{ (x_0, \dots, x_{n-1})^T \;\middle|\; \begin{pmatrix} x_0 + x_1 & x_2 & x_3 & \cdots & x_{n-1} \\ x_2 & x_0 - x_1 & 0 & \cdots & 0 \\ x_3 & 0 & x_0 - x_1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ x_{n-1} & 0 & \cdots & 0 & x_0 - x_1 \end{pmatrix} \succeq 0 \right\}.$$

Ортант можно также непосредственно представить в виде подпространства неотрицательно определенных диагональных матриц, $\mathbb{R}_+^n = \{x \in \mathbb{R}^n \,|\, D(x) = \operatorname{diag}(x) \succeq 0\}$.

Ниже мы увидим, что эти классы задач эффективно решаются. В то же время они покрывают очень широкий спектр оптимизационных проблем, встречающихся на практике. Однако, найти способ переписать данную задачу в одной из вышеописанных стандартных форм часто нетривиально.

Конуса, фигурирующие в перечисленных классах конических задач, имеют границу, описываемую алгебраическим уравнением. Поэтому задачи с трансцендентными ограничениями не принадлежат к этим классам. Приведем пример трансцендентного конуса, с помощью которого можно описать ограничения, включающие экспоненциальную функцию.

Экспоненциальным конусом называют подмножество $\mathbb{R}^3$, задаваемое соотношением

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

С помощью экспоненциального конуса выпуклое ограничение $y \geq e^x$ записывается в виде $(1, x, y) \in K_{\exp}$. Такие ограничения встречаются в геометрических программах (Geometric program, GP).

Упражнение 1. *Найдите группу автоморфизмов конуса $K_{\exp}$.*

Резюме: Самыми распространенными классами конических программ являются линейные программы (LP), квадратично-конические программы (SOCP) и полу-определенные программы (SDP), определенные над ортантом, произведением конусов Лоренца и матричным конусом, соответственно. Эти классы программ эффективно решаются и покрывают широкий спектр задач.

1.3 Самосогласованные функции

Самосогласованные функции были введены Ю.Е. Нестеровым и А.С. Немировским при изучении поведения метода Ньютона. Связывающим звеном здесь является аффинная инвариантность. Последовательность итераций метода Ньютона на данной функции переходит в себя при аффинном преобразовании координат, т.е. метод обладает аффинной инвариантностью. Поэтому естественно изучать поведение метода на классах функций, также являющимися аффинно инвариантными. Самосогласованные функции естественным образом возникают как аффинно инвариантный аналог функций с липшицевым гессианом.

В этом разделе мы рассмотрим поведение метода Ньютона на классе самосогласованных функций. Материал почерпнут из книги Ю.Е. Нестерова и А.С. Немировского [10].

Рассмотрим задачу минимизации выпуклой функции $f : D \to \mathbb{R}$ класса C^3, определенной на открытом выпуклом множестве $D \subset A$, где A – некое аффинное пространство. Для простоты предположим, что гессиан f'' функции положительно определен всюду на D. Исходя из данной итерации $x_k \in D$, *метод Ньютона* выдает следующую итерацию по формуле

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

В точке x_{k+1} строго выпуклый полином Тейлора функции f второго порядка вокруг точки x_k достигает минимума,

$$x_{k+1} = \arg\min_x q_k(x) = \arg\min_x f(x_k) + \langle f'(x_k), x - x_k\rangle + \frac{1}{2}(x - x_k)^T f''(x_k)(x - x_k).$$

Вышестоящее выражение требует уточнения. Градиент $f'(x_k)$ и гессиан $f''(x_k)$ являются линейной и квадраточной формой, соответственно, на касательном пространстве к A в точке x_k, которое может быть отождествлено с векторным пространством V, лежащем в основе аффинного пространства A. Разность $x - x_k$ есть вектор из V, и вышеназванные формы применимы к этому вектору. Выражение $(f''(x_k))^{-1} f'(x_k)$ также является вектором из V, поскольку обратную гессиана можно интерпретировать как квадратичную форму на кокасательном пространстве к A. Поэтому это выражение можно вычесть из точки x_k аффинного пространства, и результатом будет другая точка из A.

Другими словами, метод Ньютона является *аффинно инвариантным*, т.е. последовательность итераций не зависит от выбора системы координат в аффинном пространстве A. При анализе шага Ньютона естественно использовать евклидову норму $||\cdot||_{x_k}$, задаваемую гессианом $f''(x_k)$ на касательном пространстве в точке x_k. В этой локальной норме подмножества уровня $\{x \in A \,|\, q_k(x) \le c\}$ квадратичной аппроксимации являются шарами с центром в точке минимума x_{k+1} функции $q_k(x)$. Длина шага в этой норме равна

$$\rho = \sqrt{(x_{k+1} - x_k)^T f''(x_k)(x_{k+1} - x_k)} = \sqrt{f'(x_k)^T (f''(x_k))^{-1} f'(x_k)}. \tag{1}$$

Это выражение также задает длину градиента $f'(x_k)$ в локальной норме, и через него выражается разница между значением функции f в текущей точке и значением минимума аппроксимации q_k,

$$f(x_k) - q_k(x_{k+1}) = -\langle f'(x_k), x_{k+1} - x_k\rangle - \frac{1}{2}(x_{k+1} - x_k)^T f''(x_k)(x_{k+1} - x_k) = \frac{1}{2} f'(x_k)^T (f''(x_k))^{-1} f'(x_k) = \frac{\rho^2}{2}.$$

Очевидно, без дополнительных условий на функцию f невозможно сделать какие-либо утверждения о поведении этой функции в окрестности новой точки x_{k+1}, в частности, насколько уменьшилось или уменьшилось ли вообще значение функции f в новой точке по сравнению с предыдущей. Для этого необходимо, чтобы аппроксимация q_k функции f, построенная в точке x_k, не слишком сильно отличалась от f в окрестности новой точки x_{k+1}. Другими словами, локальные нормы $||\cdot||_{x_k}, ||\cdot||_{x_{k+1}}$ не должны слишком сильно отличаться друг от друга. Это требование соответствует некоему условию типа Липшица на гессиан $f''(x)$. С другой стороны, аффинная инвариантность предпологает, что сравнивать изменение гессиана в окрестности данной точки x можно только с длиной вариации, измеренной в локальной норме $||\cdot||_x$. Это приводит нас к следующему определению.

Определение 4. *Пусть $a > 0$. Выпуклая функция $f : D \to \mathbb{R}$ класса C^3, определенная на выпуклой области $D \subset A$ некоего аффинного пространства называется* a-самосогласованной *если она удовлетворяет неравенству*

$$|f'''(x)[h, h, h]| \le 2a^{-1/2}(f''(x)[h, h])^{3/2}$$

для всех $x \in D$ и всех касательных векторов $h \in T_x D$.

Функция называется сильно *a-самосогласованной если дополнительно выполняется условие*

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

Напомним, что производные функции f интерпретируются как мульти-линейные формы на касательном пространстве, т.е.

$$f''(x)[h,h] = \sum_{i,j=1}^{n} \frac{\partial^2 f(x)}{\partial x_i \partial x_j} h_i h_j, \qquad f'''(x)[h,h,h] = \sum_{i,j,k=1}^{n} \frac{\partial^3 f(x)}{\partial x_i \partial x_j \partial x_k} h_i h_j h_k,$$

где n – размерность пространства. Показатель $\frac{3}{2}$ на правой стороне неравенства необходим для того, чтобы неравенство было однородным по h.

Множитель $2a^{-1/2}$ можно рассматривать как константу Липшица гессиана f'' по отношению к локальной норме. Чем больше a, тем сильнее условие на f. Однако, a-самосогласованную функцию всегда можно нормальзовать так, чтобы она стала 1-самосогласованной (или просто самосогласованной), умножив на константу a^{-1}.

Граничное условие означает, что f стремится к $+\infty$ на любой последовательности точек в D, стремящейся к точке на границе ∂D. Это условие обеспечивает замкнутость подмножеств уровня $\{x \in D \,|\, f(x) \leq c\}$ для любых констант $c \in \mathbb{R}$.

Следующий результат гарантирует, что метод Ньютона в применении к сильно a-самосогласованной функции f может безопасно делать шаги конечной длины, где "безопасно"означает, что все итерации являются элементами D и значения функции f на последовательности итераций монотонно убывают.

Лемма 1. *Пусть $f : D \to \mathbb{R}$ – сильно a-самосогласованная функция. Тогда для любого $x^* \in D$ центрированный на x^** эллипсоид Дикина*, задающийся*

$$E_{x^*} = \{x \,|\, (x - x^*)^T f''(x^*)(x - x^*) < a\},$$

содержится в области D.

Из этого следует, что шаг Ньютона не выводит из области D если градиент $f'(x_k)$ в текущей точке имеет локальную норму, не превышающую $\sqrt{a}$. В противном случае необходимо делать более короткие шаги, т.е. использовать метод Ньютона с укороченным шагом.

Определение 5. *Пусть $f : D \to \mathbb{R}$ – сильно a-самосогласованная функция.* Ньютоновский декремент *функции f в точке $x \in D$ определяется величиной*

$$\lambda(x; f) = a^{-1/2}\rho = a^{-1/2}\sqrt{f'(x_k)^T (f''(x_k))^{-1} f'(x_k)}.$$

Отметим, что Ньютоновский декремент инвариантен относительно умножения функции f на положительные константы. Поэтому без ограничения общности можно ограничиться минимизацией сильно 1-самосогласованных функций. Нестеров и Немировский в [10] предложили следующий метод с укороченным шагом для минимизации сильно 1-самосогласованной функции $f : D \to \mathbb{R}$.

`выберем начальную точку` $x_0 \in D$`, порог точности` $\epsilon > 0$ `и положим` $k = 0$
`while` $\lambda(x_k; f) \geq \epsilon$`,`

$$\gamma_k = \begin{cases} \frac{1}{1+\lambda(x_k;f)}, & \lambda(x_k; f) \geq \frac{1}{3}, \\ \frac{1-\lambda(x_k;f)}{\lambda(x_k;f)(3-\lambda(x_k;f))}, & \lambda(x_k; f) \in (2-\sqrt{3}, \frac{1}{3}), \\ 1, & \lambda(x_k; f) \leq 2-\sqrt{3}. \end{cases}$$

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

`end`

Пока Ньютоновский декремент $\lambda(x; f)$ большой, алгоритм делает короткие шаги порядка $\lambda(x; f)^{-1}$. Как только $\lambda(x; f)$ становится меньше, чем $2 - \sqrt{3}$, алгоритм делает полные шаги. Алгоритм обладает следующими свойствами:

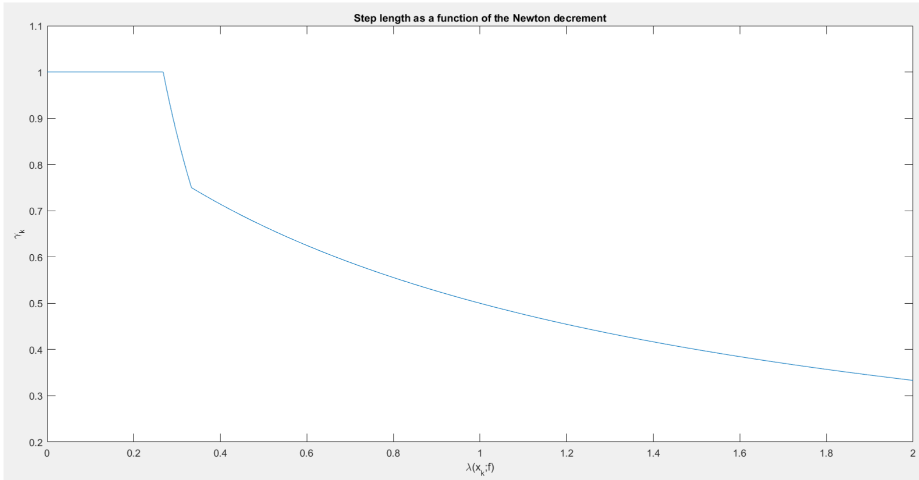

Рис. 1: Зависимость длины шага в методе Ньютона от декремента

- если $\lambda(x_k; f) \geq \frac{1}{3}$, то $f(x_{k+1}) \leq f(x_k) - (\lambda(x_k; f) - \log(1 + \lambda(x_k; f)))$,
- значение $\lambda(x_k; f)$ в промежуточном интервале $(2 - \sqrt{3}, \frac{1}{3})$ принимается не более одного раза,
- если $\lambda(x_k; f) \leq 2 - \sqrt{3}$, то $\lambda(x_{k+1}; f) \leq \left(\frac{\lambda(x_k;f)}{1-\lambda(x_k;f)}\right)^2$.

Таким образом, при $\lambda(x_k; f) \leq 2 - \sqrt{3}$ алгоритм сходится с *квадратичной* скоростью (на каждом шаге число значимых цифр после запятой удваивается).

Однако, последовательность итераций не обязана заходить в зону квадратичной сходимости. Этого не произойдет тогда и только тогда, когда функция f неограничена снизу. Тогда Ньютоновский декремент не опускается ниже единицы и последовательность $\{f(x_k)\}$ значений функции стремится к $-\infty$. Другими словами, если на каком-то шаге декремент становится меньше единицы, то функция ограничена снизу и метод сойдется к минимуму с квадратичной скоростью.

Заметим, что если f – ограниченная снизу 1-самосогласованная функция, то для любой точки x в эллипсоиде Дикина с центром в x^* справедливы неравенства (см. также Рис. 1.3)

$$(\sqrt{1 + ||x - x^*||_{x^*}} - 1)(3 - \sqrt{1 + ||x - x^*||_{x^*}}) \leq \lambda(x; f) \leq (1 - \sqrt{1 - ||x - x^*||_{x^*}})(3 - \sqrt{1 - ||x - x^*||_{x^*}}). \quad (2)$$

В частности, все точки x, удовлетворяющие $||x - x^*||_{x^*} \leq 1 - (2 - \sqrt{3 - \sqrt{3}})^2 \approx 0.2362$, находятся в зоне квадратической сходимости.

Дадим геометрическую интерпретацию метода Ньютона. Точка минимума x^* функции f характеризуется условием оптимальности первого порядка $f'(x^*) = 0$. Это позволяет нам игнорировать значения f и ограничиться поиском корня этого векторного уравнения.

Пусть V – векторное пространство, лежащее в основе аффинного пространства A. Тогда градиент $f'(x_k)$ является элементом двойственного векторного пространства V^*. Граф градиентного отображения $\nabla f : x \mapsto f'(x)$ тогда является n-мерным подмногообразием M $2n$-мерного произведения $A \times V^*$. Точка минимума x^* соответствует паре $(x^*, 0) \in A \times V^*$, которая одновременно является пересечением многообразия M с горизонтальным подпространством $H_0 = \{(x, 0) \,|\, x \in A\}$.

Тогда итерация метода Ньютона может быть интерпретирована следующим образом. Напомним, что в текущей точке x_k мы строим квадратичную аппроксимацию q_k функции f. Граф $M_k \subset A \times V^*$ градиентного отображения ∇q_k также является n-мерным подмногообразием. Более того, градиент ∇q_k *линейный*, и поэтому M_k является даже подпространством произведения $A \times V^*$. Так как градиент и гессиан q_k в точке x_k совпадает с градиентом и гессианом f, соответственно, это подпространство не что иное, чем *касательная плоскость* к M в точке $(x_k, f'(x_k))$. Следующая точка x_{k+1} определяется пересечением $(x_{k+1}, 0)$ подпространств M_k и H_0 (см. Рис. 1.3).

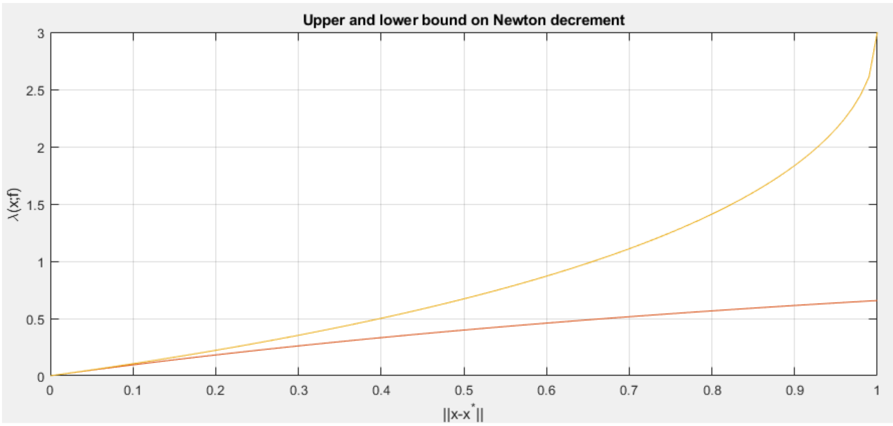

Рис. 2: Оценки на декремент как функция от расстояния до минимума

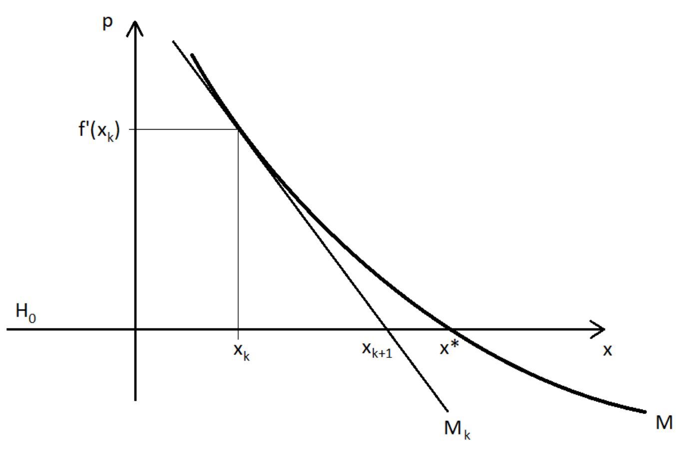

Рис. 3: Геометрическая интерпретация метода Ньютона

Резюме: Самосогласованные функции возникают как естественный аффинно инвариантный класс выпуклых функций для минимизации с помощью метода Ньютона. Если у самосогласованной функции есть минимум, то вокруг него есть область конечного радиуса в локальной норме, в которой метод Ньютона сходится с квадратичной скоростью.

1.4 Прямой метод внутренный точки

В этом разделе мы рассмотрим принцип методов внутренней точки на самой простой версии, прямом методе следования центральному пути.

Вся сложность конической программы заключена в нелинейном ограничении $x \in K$. Основная идея методов внутренней точки заключается в устранении нелинейного условия включения посредством добавления *барьерной функции* $F : K^o \to \mathbb{R}$ к линейной функции цены. Для удобства положим $F(x) = +\infty$ для всех $x \notin \operatorname{int} K$. К барьерной функции разумно предъявить следующие требования:

- $F(x)$ выпукла (проблема минимизации суммы должна оставаться выпуклой),
- $F(x)$ достаточно гладкая (чтобы можно было использовать методы более высокого порядка с высокой локальной скоростью сходимости),
- $\lim_{x\to\partial K} F(x) = +\infty$ (барьерное свойство),
- F является 1-самосогласованной.

Последнее требование необходимо для того, чтобы контролировать поведение метода Ньютона при минимизации. Заметим также, что перечисленные выше свойства остаются в силе, если прибавить к F линейную функцию.

На данном этапе неважно, что нелинейное ограничение коническое. Пусть $X \subset \mathbb{R}^n$ – произвольное замкнутое выпуклое множество с непустой внутренностью, не содержащее прямой, а $F : X^o \to \mathbb{R}$ – сильно 1-самосогласованная функция на X. Наша цель решить выпуклую задачу оптимизации

$$\inf_{x\in X} c^T x \quad : \quad Ax = b,$$

существование решения $x^* \in \partial X$ которой мы предполагаем.

Вместо исходной задачи мы рассмотрим семейство задач

$$\min_x(\tau \cdot c^T x + F(x)) \quad : \qquad Ax = b,$$

параметризованное вещественным параметром $\tau \geq 0$. В этих вспомогательных задачах нелинейное условие $x \in X$ отсутствует, но зато мы вместо линейной функции цены минимизируем сильно 1-самосогласованные. В предыдущем разделе мы увидели, что для этого можно использовать метод Ньютона, причем наиболее эффективно, если мы находимся в зоне квадратичной сходимости. Заметим, что линейные ограничения типа равенства для этого не являются препятствием, поскольку можно перейти в задаваемое этими ограничениями аффинное подпространство.

Для достаточно больших τ точки минимума $x^*(\tau)$ вспомогательной задачи существуют и единственны. Более того, они являются дифференцируемыми по параметру τ и образуют кривую, называемую *центральным путем.* При $\tau \to +\infty$ центральный путь стремится к решению $x^* = x^*(+\infty)$ исходной задачи. Если у исходной задачи решение не единственно, то пределом центрального пути будет служить точка в относительной внутренности множества решений.

Опишем схему метода внутренней точки. Метод стартует с пары (x_0, τ_0), где x_0 – точка в зоне квадратичной сходимости решения $x^*(\tau_0)$ вспомогательной задачи с параметром τ_0. Делая полный шаг Ньютона по направлению к решению $x^*(\tau_0)$, метод выдает новую точку x_1, которая находится ближе к $x^*(\tau_0)$ и тем самым к центральному пути, чем предыдущая. После этого значение параметра обновляется на $\tau_1 > \tau_0$. При этом τ_1 настолько близко к τ_0, что точка x_1 все еще находится в зоне квадратичной сходимости вокруг решения $x^*(\tau_1)$ вспомогательной задачи с параметром τ_1. После этого происходит переход к следующей итерации.

Алгоритм устроен таким образом, что на каждом шаге выполняются неравенства

$$\lambda(x_k; f_k) \le \overline{\lambda}, \qquad \lambda(x_{k+1}; f_k) \le \underline{\lambda},$$

где $f_k(x) = F(x) + \tau_k \langle c, x\rangle$ – функция цены вспомогательной задачи с параметром τ_k, а константы $0 < \underline{\lambda} < \overline{\lambda} < 2 - \sqrt{3}$ выбраны так, что первое из приведенных выше неравенств влечет за собой второе. Из оценок (2) следует, что $||x_{k+1} - x^*(\tau_k)||_{x^*(\tau_k)} \le \underline{d}$, а неравенство $||x_{k+1} - x^*(\tau_{k+1})||_{x^*(\tau_{k+1})} \le \overline{d}$ влечет за собой условие $\lambda(x_{k+1}; f_{k+1}) \le \overline{\lambda}$ для некоторых положительных констант $\underline{d} < \overline{d}$. Из этого следует, что 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 приведенные выше неравенства всегда будут выполняться.

Остается выяснить, насколько можно увеличивать параметр τ на каждом шаге. Для этого на центральном пути необходимо сопоставить параметр τ с параметром длины в римановой метрике, задаваемой гессианом F''.

Лемма 2. *Для всех значений $\tau > 0$, отвечающих точке на центральном пути, справедливо соотношение* $||\frac{dx^*(\tau)}{d\log\tau}||_{x^*(\tau)} = \sqrt{F'(x^*(\tau))^T(F''(x^*(\tau)))^{-1}F'(x^*(\tau))} = ||F'(x^*(\tau))||_{x^*(\tau)}$.

Упражнение 2. *Докажите лемму 2.*

Из леммы 2 следует, что на каждом шаге можно увеличивать $\log\tau$ на величину порядка $\frac{\overline{d} - \underline{d}}{||F'(x^*(\tau))||_{x^*(\tau)}}$. Чтобы была гарантирована линейная скорость сходимости, необходимо ограничить норму $||F'(x^*(\tau))||_{x^*(\tau)}$ сверху. Это мотивирует следующее определение.

Определение 6. *Сильно 1-самосогласованная функция $F : X \to \mathbb{R}$, определенная на некоторой выпуклой области X, называется* самосогласованным барьером *на X с параметром ν, если для всех $x \in X$ выполнено соотношение*

$$||F'(x)||^2_{x^*} \le \nu.$$

Таким образом, если функция F является самосогласованным барьером с параметром ν, то вышеописанный прямой метод следования центральному пути позволяет на каждом шаге увеличивать $\log\tau$ на величину порядка $\nu^{-1/2}$, или, что эквивалентно, умножать τ на константу $\theta = 1 + O(\nu^{-1/2})$. Чем больше параметр барьера ν, тем меньше множитель θ и тем медленнее метод будет сходиться. Поэтому на каждом данном выпуклом множестве желательно иметь в наличии барьеры с как можно меньшим параметром. Сложность задачи зависит от наличия эффективно вычислимого самосогласованного барьера с небольшим значением параметра.

В коническом программировании барьер определяется на конусе K, допускающем действие группы растяжений. Разумно потребовать также инвариантность барьера на K по отношению к этой группе симметрий. Постоянство функции F на внутренних лучах конуса приводит к противоречию с остальными условиями. Однако, в методе Ньютона используются только производные минимизируемой функции. Поэтому достаточно потребовать инвариантность производных F по отношению к растяжениям. Это естественным образом приводит к условию логарифмичной однородности.

Определение 7. *Пусть $K \subset \mathbb{R}^n$ – выпуклый регулярный конус. Выпуклая функция $F : K^o \to \mathbb{R}$ класса C^3 с положительно определенным гессианом называется* логарифмично однородным самосогласованным барьером *с* параметром ν *если F является сильно 1-самосогласованной функцией на K^o и F логарифмично однородно порядка $-\nu$, т.е.*

$$F(\alpha x) = -\nu\log\alpha + F(x) \tag{3}$$

для всех $x \in K^o$ и $\alpha > 0$.

Таким образом, растяжения действуют прибавлением констант к логарифмично однородной функции F.

Данное для случая конусов определение самосогласованного барьера совместимо с приведенным выше общим определением, поскольку логарифмичная однородность также ограничивает Ньютоновский декремент, и локальная норма градиента барьера с параметром ν равна константе $\sqrt{\nu}$. Действительно, дифференцируя соотношение (3) по x, мы получим $\alpha F'(\alpha x) = F'(x)$. Дифференцируя это соотношение и (3) по α при $\alpha = 1$, мы получим

$$F'(x) + F''(x) \cdot x = 0, \qquad \langle F'(x), x\rangle = -\nu. \tag{4}$$

Из этого следует, что $(F''(x))^{-1}F'(x) = -x$ и следовательно $(F'(x))^T(F''(x))^{-1}F'(x) = -\langle F'(x), x\rangle = \nu$.

Резюме: Для решения конической программы с помощью метода внутренней точки необходимо наличие эффективно вычислимого логарифмично однородного самосогласованного барьера с небольшим значением параметра ν на соответствующем выпуклом конусе. В методах следования центральному пути нелинейное коническое ограничение устраняется переходом к однопараметрическому семейству вспомогательных задач, каждая из которых состоит в минимизации самосогласованной функции. Метод чередует шаг Ньютона для решения вспомогательной задачи с умножением параметра семейства на величину $1+O(\nu^{-1/2})$, при этом последовательность итераций сходится с линейной скоростью к решению исходной задачи. Скорость сходимости увеличивается с уменьшением параметра барьера ν.

1.5 Двойственность

В этом разделе мы рассмотрим коническую двойственность. Она распространяется как на регулярные выпуклые конуса, так и на самосогласованные барьеры и на сами конические программы.

Определение 8. *Пусть $K \subset \mathbb{R}^n$ – произвольный конус.* Двойственным *к K конусом называется множество*

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

Очевидно K^* является замкнутым и выпуклым. Нетрудно видеть, что имеет место включение $K \subset (K^*)^*$. Если конус K регулярный, то двойственный конус K^* также будет регулярным. Более того, в этом случае мы имеем соотношение $(K^*)^* = K$.

Упражнение 3. *Найдите двойственный конус к ортанту $\mathbb{R}^n_+$, конусу Лоренца L_n и матричному конусу $\mathcal{S}^n_+$.*

Для любого логарифмично однородного самосогласованного барьера $F : \operatorname{int} K \to \mathbb{R}$ на регулярном выпуклом конусе K с параметром ν можно определить *двойственный* барьер $F_* : \operatorname{int} K^* \to \mathbb{R}$ на двойственном к K конусе посредством преобразования Лежандра

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

Супремум для данного $s \in \operatorname{int} K^*$ достигается в точке $x \in K^o$, удовлетворяющей уравнению $F'(x) = -s$. В силу строгой выпуклости F и свойства $F|_{\partial K} = +\infty$ такая точка x существует и единственна для любого $s \in \operatorname{int} K^*$. Действительно, рассмотрим аффинную гиперплоскость $A = \{x \in \mathbb{R}^n \,|\, \langle x, s\rangle = \nu$. Так как s находится во внутренности двойственного конуса K^*, пересечение $X = A \cap K$ является компактным. На внутренности X функция F строго выпукла, а на границе имеем $F|_{\partial X} = +\infty$. Поэтому минимум функции F на X существует и единственен. Обозначим этот минимум через x^*. Условие оптимальности первого порядка в точке x^* сводится к существованию $\lambda \in \mathbb{R}$ такого, что $F'(x^*) = \lambda s$. В силу (4) из этого следует, что

$$-\nu = \langle F'(x^*), x^*\rangle = \lambda\langle s, x^*\rangle = \lambda\nu.$$

Поэтому $\lambda = -1$ и $F'(x^*) = -s$, чего и требовалось доказать.

Теорема 1.1. *Функция F_* является логарифмично однородным самосогласованным барьером с параметром ν на K^* [10, Теорема 2.4.4]. Отображение $\mathcal{L} : x \mapsto -F'(x)$ является биекцией между внутренностями конусов K и K^*. Если рассмотреть эти внутренности как римановы многообразия, оснащенные метриками $F''(x)$ и $F''_*(s)$, соответственно, то отображение $\mathcal{L}$ является изометрией ([10, стр. 45], см. также [9]).*

Замечание 1.2. По определению преобразование Лежандра f^* функции f задается формулой $f^*(p) = \sup_{x\in K}(\langle p,x\rangle - f(x))$. Знак минус при аргументе двойственного барьера был введен для совместимости с определением двойственного конуса, $F_*(p) = F^*(-p)$.

Двойственность также распространяется на конические программы. Пусть дана коническая программа

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

над некоторым регулярным выпуклым конусом K. Рассмотрим произвольный элемент $s \in K^*$ вида $s = c - A^T y$, где y – некоторый вектор соответствующей размерности. Тогда для любого x из допустимого множества исходной конической программы имеем

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

Таким образом, мы получили *нижнюю* оценку $\langle b,y\rangle$ на оптимальное значение исходной конической программы.

Двойственная коническая программа формулируется как задача *максимизации* этой оценки и записывается в виде

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

Она определена над двойственным конусом K^*. Очевидно для каждой допустимой точки x исходной конической программы величина $\langle c,x\rangle$ является *верхней* оценкой оптимального значения двойственной программы.

Разница между оптимальными значениями прямой и двойственной задачи называется *разрывом двойственности.*

Двойственную пару конических программ можно записать в более симметрическом виде. Аффинную оболочку допустимого множества исходной программы с функцией цены $\langle c,x\rangle$ можно представить в виде суммы $b+L$, где $L \subset \mathbb{R}^n$ – некоторое линейное подпространство, а вектор $b \in \mathbb{R}^n$ выбран так, чтобы имело место соотношение $\langle c,b\rangle = 0$. Таким образом получаем программу

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

Тогда двойственную программу можно записать в виде

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

где $L^\perp \subset \mathbb{R}_n$ – ортогональное к L линейное подпространство. Для любой пары (x,s) допустимых точек будет иметь место соотношение

$$\langle x,s\rangle = \langle c,x\rangle + \langle b,s\rangle \geq 0.$$

Прямо-двойственные методы решают обе программы одновременно и для каждой из них генерируют последовательность точек $x_k \in \operatorname{int} K$ и $s_k \in \operatorname{int} K^*$, соответственно. Эти точки можно объединить в пары (x_k,s_k) в произведении прямого и двойственных пространств. Последовательность зазоров $\langle c,x_k\rangle + \langle b,s_k\rangle$ при этом монотонно убывает.

Некоторые прямо-двойственные методы минимизируют потенциал, определенный на произведении внутренностей прямого и двойственного конуса. Примером такого потенциала может служить функция

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

где константа зависит от параметра ν барьера F.

Резюме: Каждому регулярному выпуклому конусу K можно сопоставить двойственный конус K^*, который также является регулярным и выпуклым. Каждому логарифмично однородному самосогласованному барьеру F на K можно сопоставить двойственный барьер F_* на K^* с тем же значением параметра, определяемый через преобразование Лежандра. Каждой конической программе над K можно сопоставить двойственную программу над K^*. Прямо-двойственные методы решают обе программы одновременно и генерируют последовательность пар $(x_k,s_k) \in \operatorname{int} K \times \operatorname{int} K^*$.

1.6 Примеры самосогласованных барьеров

Приведем явные выражения для самосогласованных барьеров на конусах, лежащих в основе ходовых классов конических программ, приведенных выше.

Барьером на прямом произведении $K = \prod_{i=1}^m K_i$ конусов, оснащенных барьерами F_i с параметрами ν_i, может служить сумма

$$F(x_1, \dots, x_m) = \sum_{i=1}^m F_i(x_i).$$

Она обладает параметром $\nu = \sum_{i=1}^m \nu_i$.

Рассмотрим элеметнарные конусы, лежащие в основе классов программ LP, SOCP и SDP.

конус	$\mathbb{R}_+^n$	L_n	$\mathcal{S}_+^n$	$K = \prod_j K_j$
барьер	$-\sum_{j=1}^n \log x_j$	$-\log(x_0^2 - x_1^2 - \dots - x_{n-1}^2)$	$-\log\det A$	$\sum_j F_j$
параметр ν	n	2	n	$\sum_j \nu_j$

Параметр этих барьеров оптимальный, т.е. на этих конусах нет барьера со строго меньшим параметром. Но они обладают еще одним полезным свойством, они являются *авто-шкалированными*. Это свойство допускает особенно эффективные методы внутренних точек для решения конических программ над этими конусами, а именно методы с длинным шагом [7],[8].

Экспоненциальный конус: Рассмотрим снова конус

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

встречающийся в геометрическом программировании. На этим конусе существует барьер

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

значение параметра которого равно $\nu = 3$. Это значение параметра также оптимально.

1.6.1 Универсальные конструкции барьеров

Наличие эффективно вычислимого логарифмично однородного самомогласованного барьера на выпуклом конусе позволяет решать конические программы над этим конусом. Естественно возникает вопрос о существовании барьеров на произвольных конусах. Ответ на этот вопрос утвердительный, и в этом разделе мы приведем два способа построения таких барьеров.

Универсальный барьер. Пусть $K \subset \mathbb{R}^n$ – регулярный выпуклый конус. Рассмотрим его *характеристическую функцию*

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

определенную для всех внутренних точек $x \in K$. Функция $F(x) = \log \varphi(x)$ является логарифмично однородным самосогласованным барьером на K со значением параметра $\nu = n$. Этот барьер называется *универсальным.* Он был введен в [10] Ю.Е. Нестеровым и А.С. Немировским и является первой универсальной конструкцией барьера. Изначально было известно, что параметр универсального барьера ограничен сверху величиной $C \cdot n$, где $C \geq 1$ – некоторая постоянная. Позже в работах [1],[6] было установлено, что можно выбрать $C = 1$.

Если A – линейный автоморфизм конуса K, то имеет место соотношение $F(Ax) = F(x) + \log\det A$ для всех $x \in K^o$. Это позволяет вычислить универсальный барьер на однородных конусах, т.е. на внутренности которых группа линейных автоморфизмов действует транзитивно [3, 4].

В общем случае универсальный барьер задается многомерным интегралом по двойственному конусу и его трудно вычислить даже для сравнительно простых неоднородных конусов.

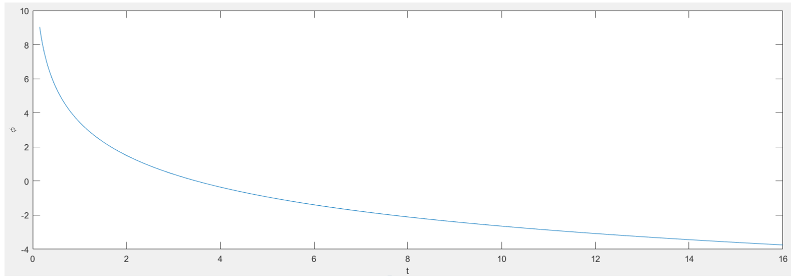

Рис. 4: Граф функции $\phi(t)$, определяющей канонический барьер на конусе $K_{\exp}$

Энтропический барьер. Этот барьер двойственный к универсальному. Его параметр также равен размерности конуса. На ограниченных выпуклых множествах энтропический барьер был рассмотрен в работе [1].

Канонический барьер. Построение этого барьера основано на глубокой теореме в теории уравнений в частных производных.

Теорема 1.3. *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* $\lim_{x \to \partial D} F(x) = +\infty$.

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

Канонический барьер обладает тем же свойством инвариантности по отношению к действию группы автоморфизмов конуса K, что и универсальный, и поэтому совпадает с ним на однородных конусах.

В общем случае этот барьер также трудно вычислить, но для некоторых неоднородных конусов имеются аналитические выражения.

Пример: На 3-мерном экспоненциальном конусе $K_{\exp}$ канонический барьер задается выражением

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

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

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

Для ортанта, конуса Лоренца и матричного конуса все три конструкции приводят к барьерам, пропорциональным авто-шкалированным барьерам, приведенным в вышестоящей таблице.

Резюме: Для конусов, лежащих в основе линейных, квадратично-конических и полу-определенных программ, имеются эффективно вычислимые самосогласованные барьеры с малым значением параметра. Для произвольных конусов такие барьеры также существуют, но в общем случае их трудно вычислить.

Список литературы

[1] Sébastien Bubeck and Ronen Eldan. The entropic barrier: a simple and optimal universal self-concordant barrier. In *Proceedings of The 28th Conference on Learning Theory*, volume 40 of *Proceedings of Machine Learning Research*, 2015.

[2] George B. Dantzig. Maximization of a linear function of variables subject to linear inequalities. In T.C. Koopmans, editor, *Activity Analysis of Production and Allocation*, pages 339–347. John Wiley & Sons, New York, 1951.

[3] Osman Güler. Barrier functions in interior point methods. *Math. Oper. Res.*, 21(4):860–885, 1996.

[4] Osman Güler and Levent Tunçel. Characterization of the barrier parameter of homogeneous convex cones. *Math. Program., Ser. B*, 81(1):55–76, 1998.

[5] Victor Klee and George J. Minty. How good is the simplex algorithm? In Oved Shisha, editor, *Inequalities III. Proceedings of the Third Symposium on Inequalities*, pages 159–175, New York, London, 1972. University of California, Academic Press. Held in Los Angeles, Sept. 1969.

[6] Yin Tat Lee and Man-Chung Yue. Universal barrier is n-self-concordant. Optimization Online 2018/09/6810, 2018.

[7] Yuri E. Nesterov and Michael J. Todd. Self-scaled barriers and interior-point methods for convex programming. *Math. Oper. Res.*, 22:1–42, 1997.

[8] Yuri E. Nesterov and Michael J. Todd. Primal-dual interior-point methods for self-scaled cones. *SIAM J. Optimiz.*, 8(2):324–364, 1998.

[9] Yuri E. Nesterov and Michael J. Todd. On the Riemannian geometry defined by self-concordant barriers and interior-point methods. *Found. Comput. Math.*, 2:333–361, 2002.

[10] Yurii Nesterov and Arkadii Nemirovskii. *Interior-point Polynomial Algorithms in Convex Programming*, volume 13 of *SIAM Stud. Appl. Math.* SIAM, Philadelphia, 1994.