В.Ю.Протасов (Москва, МГУ)

Как найти минимум выпуклой функции по ее значениям ?

$f(x)$ выпуклая функция на выпуклом множестве $G \subset \mathbf{R}^d$

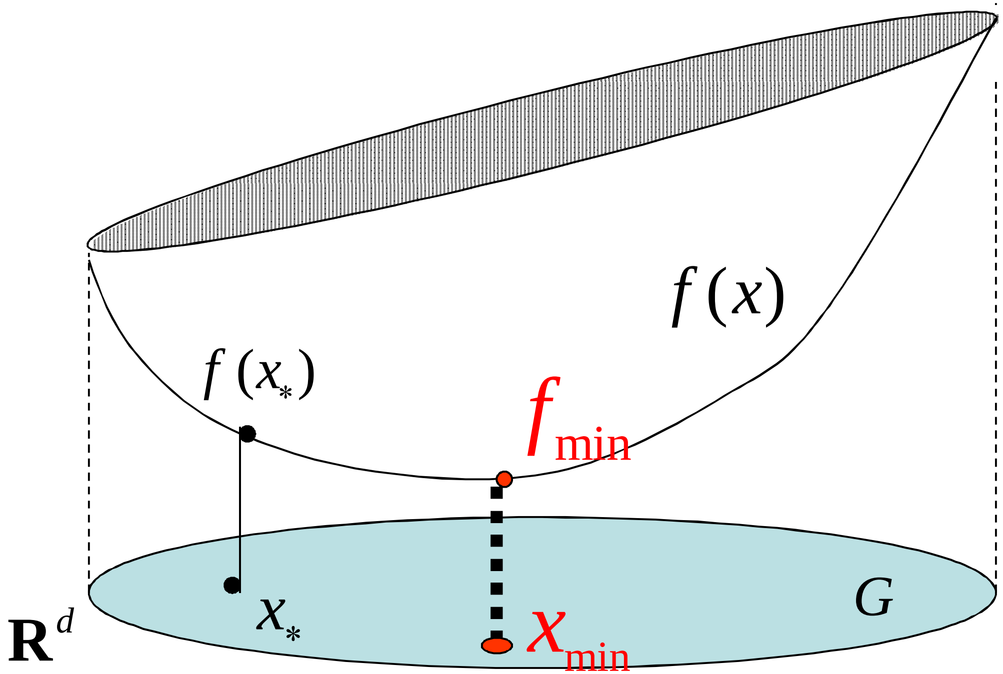
$f: G \to \mathbf{R}$

Нужно найти приближенно значение минимума f на G

solve $f(x) \to \min$

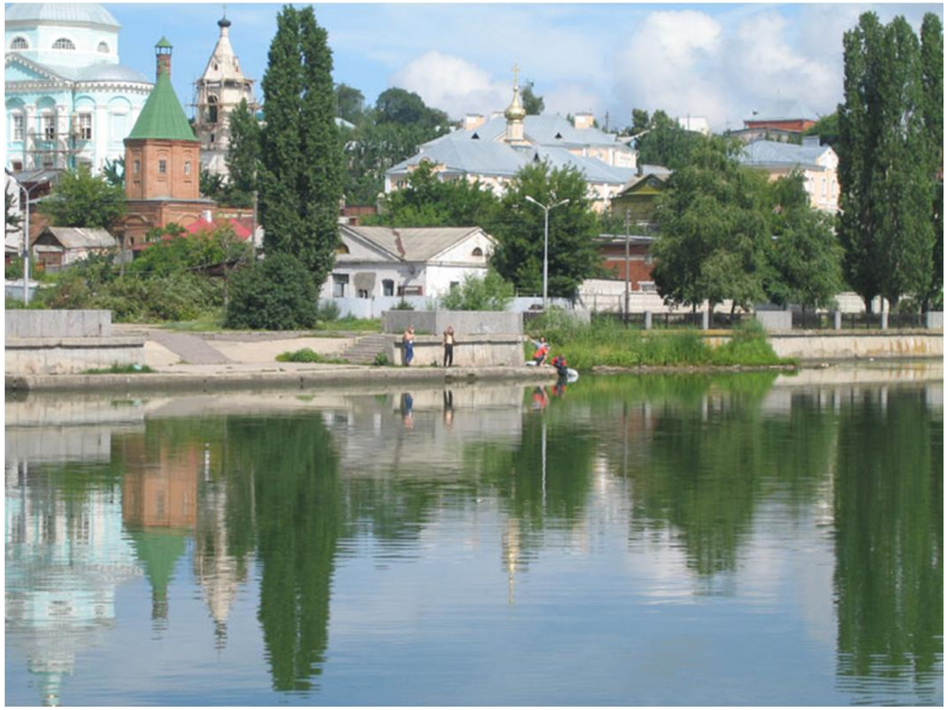
subject to $x \in G$

$$\boxed{\frac{f(x_*) - f(x_{\min})}{f(x_{\max}) - f(x_{\min})} < \varepsilon}$$



Метод центрированных сечений

А.Ю. Левин 1961-1965, D.Newman, 1965



``Маленький русский город'' Воронеж

Воронежская математическая школа
(Красносельский, Крейн, Ефимов, Рутицкий, Миролюбов, Солдатов, Семенов)

Задача, которую пытался решить Левин:

$$x = (x_1, \ldots, x_d),$$

$$f(x) = \sum_{k=1}^{N} A_k \, e^{a_1^k x_1 + a_2^k x_2 + \cdots + a_d^k x_d}$$

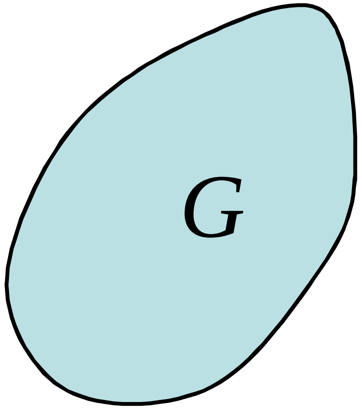
$$A_k \geq 0, \qquad a_i^k \in \mathbf{R}$$

Выпуклое тело G здавалось несколькими неравенствами

$$f_j(x) \leq 0, \qquad j = 1, \ldots, m$$

все выпуклые функции $f_j(x)$ одного и того же типа

$$f_j(x) = \sum_{k=1}^{N} A_k^j \, e^{a_{1,j}^k x_1 + a_{2,j}^k x_2 + \cdots + a_{d,j}^k x_d}$$



Число переменных (размерность) d была порядка $30 - 40$.

Как решать ?

Первый из возможных подходов

Теорема Каруша-Куна-Таккера

$f_0(x) \to \min,$

subject to $f_1(x) \le 0,\ f_2(x) \le 0,\ \ldots.\ ,\ f_m(x) \le 0$

$x = (x^1, \ldots, x^d),$ все функции $f_0, f_1, \ldots f_m$ выпуклы

Лагранжиан $\mathcal{L}(x,\lambda) = \sum_{k=0}^{m} \lambda_k f_k(x)$

если $\operatorname{int} G \ne \varnothing$

(найдётся точка x_0 для которой $f_k(x_0) < 0,\ k=1,\ldots,m$) то

$\hat{x} \in abs\min \Leftrightarrow$ существуют $\lambda_1, \lambda_2, \ldots, \lambda_m$ $(\lambda_0 = 1)$ такие, что

$\mathcal{L}'_x = \sum_{k=0}^{m} \lambda_k f'_k(\hat{x}) = 0$ (стационарность)

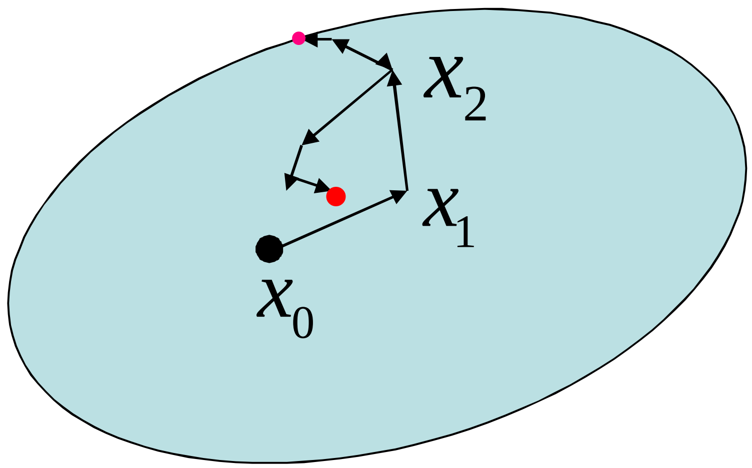
$\lambda_k \ge 0, \quad k = 1, \ldots, m$ (неотрицательность)

$\lambda_k f_k(\hat{x}) = 0, \quad k = 1,\ldots,m$ (дополняющая нежёсткость)

2^m случаев

Другой подход

Градиентный спуск (градиентная релаксация)

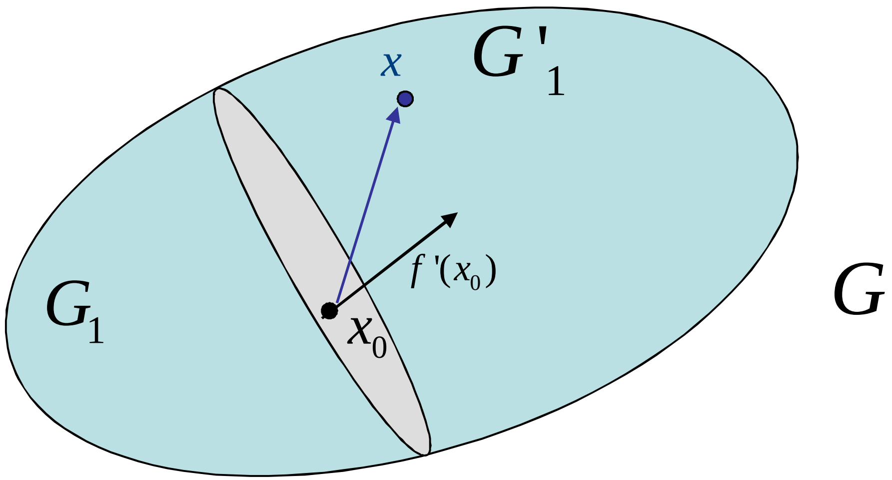


$$x_1 = x_0 - \alpha_0 \; f'(x_0)$$

$$x_2 = x_1 - \alpha_1 \; f'(x_1)$$

$\cdots\cdots\cdots$

Центрированные сечения: основная идея



$G'_1 = \left\{ x \in G, \ \left\langle x - x_0, f'(x_0) \right\rangle \geq 0 \right\}$ $\qquad$ $G = G_1 \cup G'_1$

Для любого $x \in G'_1$ $\quad f(x) \geq f(x_0) \Leftarrow$ следует из выпуклости

Удаляем часть G'_1 и ограничиваем на G_1 вместо G

Итерация: берем точку $x_1 \in G_1$ вычисляем $f'(x_1)$ строим сечение

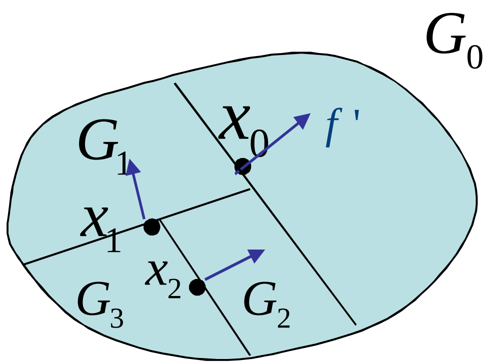
$\Rightarrow$ получаем тело G_2 и т.д. ...

После достаточного числа итераций мы локализуем точку минимума

Как выбирать точки $x_0, x_1, x_2, \ldots$?

Ответ: x_k -- центр тяжести G_k

$$x_k = gr\,(G_k) = \frac{1}{Vol\;G_k} \int\limits_{G_k} x\,d\,v$$



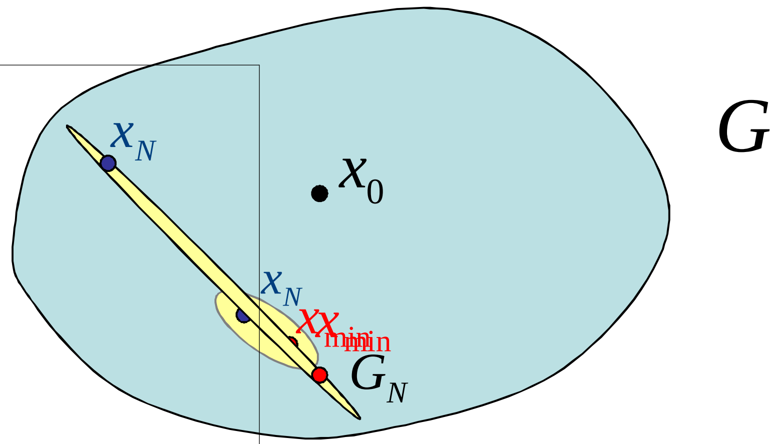
концепция ``оракула''

$$x_0 = gr(G_0)$$
$$x_1 = gr(G_1)$$
$$x_2 = gr(G_2)$$

В результате получаем тело G_N малого объёма, содержащее точку минимума $x_{\min}$ функции $f(x)$

$$x_0 = gr(G)$$

$$Vol\, G_N \le \varepsilon^d\, Vol\, G$$



$\left| f(x_N) - f_{\min} \right|$ может быть велико, когда объём мал ?

ТЕОРЕМА 1 (Немировсий, Юдин) Если $f : G \to \mathbf{R}$ выпукла и непрерывна, и если $f(x) \ge f(x_*)$ для любого $x \in G \setminus G_N$, где $Vol\, G_N \le \varepsilon^d\, Vol\, G$, то

$$\left| f(x_*) - f_{\min} \right| \le \varepsilon \left| f_{\max} - f_{\min} \right|$$

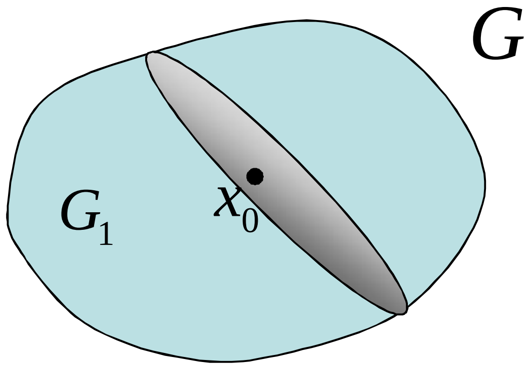
Берем $\quad x_* = \min \{ x_0, x_1, \ldots, x_N \}$

Насколько велико должно быть N для неравенства $\quad Vol\, G_N \le \varepsilon^d\, Vol\, G$?

ТЕОРЕМА 2 (Grunbaum, Minkowski) Если $x_0 \equiv gr(G)$, то для любой гиперплоскости, проходящей через x_0, имеем $Vol\, G_1 \le (1 - 1/e) Vol\, G$.

$$\frac{Vol\, G_1}{Vol\, G} \le 1 - 1/e = 0.632...$$

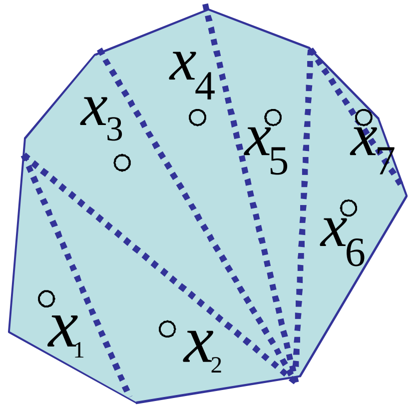
$$O\,(f\,', d\,, \varepsilon) \le C\,d\,\ln\frac{1}{\varepsilon}$$



Проблема: Как найти центр тяжести ?

$$x_0 = gr\,(G) = \frac{1}{Vol\;G}\int\limits_G x\,d\,v$$

$$gr\,(G) = \frac{1}{Vol\;G}\sum_{k=1}^{N} x_k\;Vol\,\Delta_k$$



N растёт экспоненциально от размерности d и от числа вершин

``проклятие размерности”

Вычисление gr(G) является NP-сложной задачей

А.Ю.Левин не публиковал результат 4 года (1961-1965), пытаясь найти практически применимый алгоритм.

Метод (описанных) эллипсоидов
А.С.Немировский (1977), Н.З.Шор (1977)

$$O_{e\,ell}(f', d, \varepsilon) \le C\, d^2 \ln\frac{1}{\varepsilon}$$

Метод описанных симплексов
L.Levin, D.Yamitsky (1980)

$$O_{e\,sim}(f', d, \varepsilon) \le C\, d^3 \ln\frac{1}{\varepsilon}$$

Метод вписанных эллипсоидов
Л.Хачиян, С.Тарасов, А.Эрлих (1988)

$$O_{e\,ins}(f', d, \varepsilon) \le C\, d \ln\frac{1}{\varepsilon}$$

В последнем методе вместо объёма $Vol\,(G)$ используется другой минимизатор.

ОПРЕДЕЛЕНИЕ. Неотрицательная функция μ, заданная на множестве выпуклых тел, называется минимизатором, если $\mu(t\,G) = t^d\,\mu(G)$ и $G_1 \subset G_2 \;\Rightarrow\; \mu(G_1) \le \mu(G_2)$.

ТЕОРЕМА 3 (фольклор) Если $f : G \to \mathbf{R}$ выпукла и непрерывна, и если $f(x) \ge f(x_*)$ для любого $x \in G \setminus G_N$, где $\mu(G_N) \le \varepsilon^d \mu(G)$, то

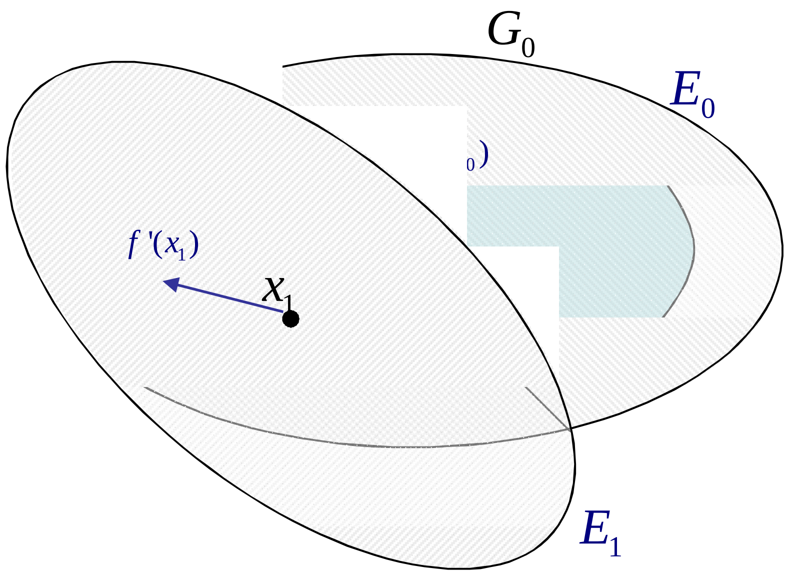
$$\left| f(x_*) - f_{\min} \right| \le \varepsilon \left| f_{\max} - f_{\min} \right|$$

В методе центров тяжести в качестве минимизатора используется объём.
В методе вписанных эллипсоидов – объём эллипсоида Джона. Вместо центра тяжести – центр эллипсоида Джона.
При других минимизаторах вместо центра тяжести возникают *аналитический центр* , *волюметрический центр*, и т.д.

Основная идея метода эллипсоидов Немировского-Шора:

Лемма. Полуэллипсоид можно поместить в эллипсоид меньшего объёма .

Алгоритм:

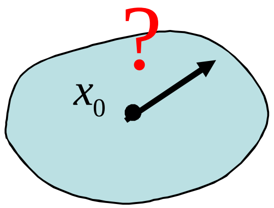


Метод эллипсоидов сделал метод центрированных сечений практически применимым

В 1979 Л.Хачиян применил метод эллипсоидов для доказательства полиномиальной разрешимости задач линейного программирования.

Негладкие задачи

Как проводить плоскости сечения, если функция $f(x)$ не дифференцируема ?



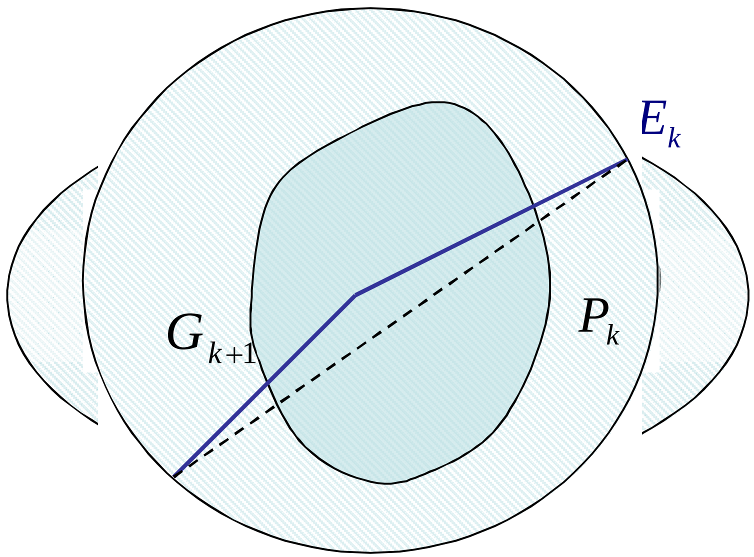
Оракул $O_e(f, d, \varepsilon)$ даёт нам значения f вместо градиента

Кузовкин, Тихомиров (1967) $O(f, d, \varepsilon) \leq Ce^{\alpha d} \ln \frac{1}{\varepsilon}, \quad \alpha > 0$

Немировский, Юдин (1979) $O_e(f, d, \varepsilon) \leq Cd^7 \ln^2 d \ln \frac{1}{\varepsilon}$

Протасов (1996) $O_e(f, d, \varepsilon) \leq Cd^2 \ln d \ln \frac{1}{\varepsilon}$

k-тая итерация метода эллипсоидов.



Мы не можем построить отделяющую плоскость.
Мы лишь можем построить достаточно широкий конус.

1) Линейным преобразованием переводим эллипсоид в евклидов шар.

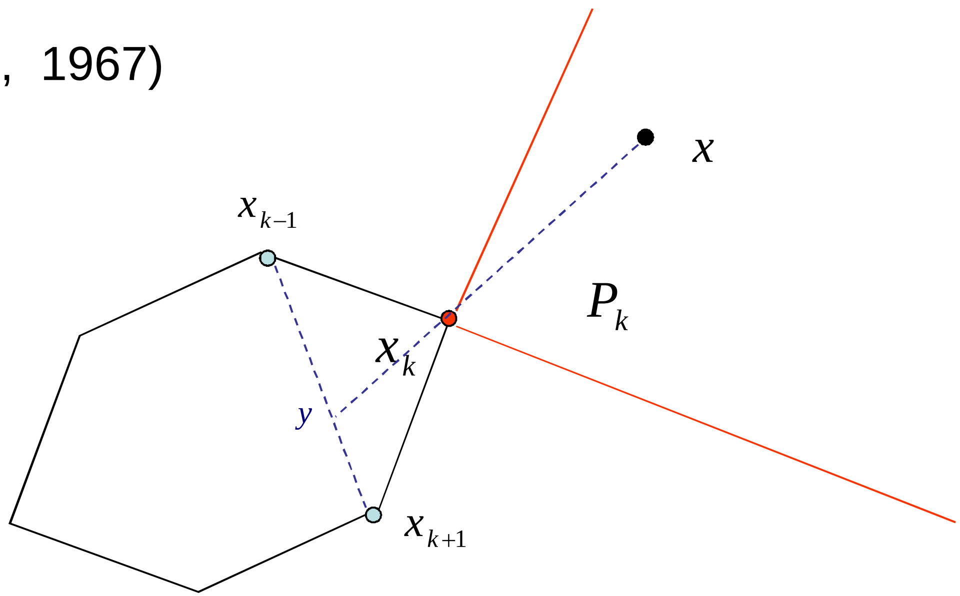
2) Строим круговой конус с вершиной $x_k : \ \forall \ x \in P_k \quad f(x) \geq f(x_k)$

3) Проводим гиперплоскость и отделяем G_{k+1}

Как построить достаточно широкий конус $P_k: \quad \forall \ x \in P_k \quad f(x) \geq f(x_k)$?

Идея (Кузовкин, Тихомиров, 1967)

Начнем со случая d=2



Мы можем вычислить f(x) в вершинах некоторого многоугольника.

Найдём вершину, значение функции в которой не меньше, чем в обеих соседних:

$$f(x_k) \geq \max\{ f(x_{k-1}), f(x_{k+1}) \}$$

$$\forall \ x \in P_k \quad f(x) \geq f(x_k)$$

$$f(x_k) \geq f(y) \quad \Rightarrow \quad f(x) \geq f(x_k)$$

Для больших размерностей d сложность очень высока.

Кузовкин, Тихомиров (1967) $O\,(f, d, \varepsilon) \quad \le \quad C e^{\alpha d} \ln \frac{1}{\varepsilon}, \quad \alpha > 1$

Немировский, Юдин (1979) $O_e(f, d, \varepsilon) \quad \le \quad C d^{\,7} \ln^2 d \ln \frac{1}{\varepsilon}$

Протасов (1996) $O_e(f, d, \varepsilon) \quad \le \quad C d^{\,2} \ln d \, \ln \frac{1}{\varepsilon}$

Идея последнего алгоритма

Пусть дана точка $x_0 \in$ int G.

1) Проведём произвольную двумерную плоскость L_1 через x_0.

Построим маленький правильный N-угольник $T_1 \subset L_1$ с вершиной x_0;

$x_0 = z_0, z_1, \ldots, z_{N-1}$ -- вершины T_1; $N \approx C d^{\,3}$.

Найдём его вершину z_j, для которой $f(z_j) \ \ge \ \max\left\{ f(z_{j-1}) \ , \ f(z_{j+1}) \ , \ f(x_0) \right\}$.

Для этого нужно перебрать $\approx \dfrac{\ln N}{\ln \rho}$ вершин T_1, где $\rho = \dfrac{\sqrt{5}+1}{2}$.

2) Положим $x_1 = z_j$.

3) Пусть $M_1 = \left\{ y \in \mathbf{R}^d \mid \; |y - z_{j-1}| = |y - z_{j+1}| \right\}$. $\dim M_1 = d - 1$.

Проводим произвольную двумерную плоскость $L_2 \subset M_1$ через x_1.

Строим маленький правильный N-угольник $T_2 \subset L_2$ с вершиной x_1. Найдём вершину x_2 .

Проведем соответствующую плоскость M_2 , $\dim M_2 = d - 2$, и т.д.

4) После $(d$ -1) итераций мы получаем x_{d-1} и прямую M_{d-1}.

Они являются вершиной и осью конуса Р с раствором $\varphi \approx \arccos\left(\frac{1}{d^2}\right)$

Для построения одного конуса мы должны вычислить функцию f в $\approx C\, d \ln d$ точках.

Для приближения минимума с относительной погрешностью ε нужно построить

$d \ln \frac{1}{\varepsilon}$ конусов (в методе с центрами тяжести)

$d^2 \ln \frac{1}{\varepsilon}$ конусов (в методе эллипсоидов).

$$O_e(f, d, \varepsilon) \quad \leq \quad C d^2 \ln d \ln \frac{1}{\varepsilon}$$

Метод Левина-Шора-Немировского-Юдина-Кузовкина-Тихомирова-Протасова

``Russian method”

Thank you!