Выпуклая оптимизация.
Численные методы оптимизации.

Краткий конспект семинаров

2015-2016 гг.

Конспект создан для 374-ой и 375-ой групп Ильнурой Усмановой, Юрием Дорном и Юрием Максимовым на основе меториалов собранных А.Г. Бирюковым, А.В. Гасниковым, С.А. Довгалем, А.В. Черновым, материалов авторов и различных лекционных источников.

NB: просьба использовать с осторожностью, вероятны множественные опечатки. *Текст содержит кроме собственно материала семинара, еще и некоторый лекционный материал про который я не говорил на семинаре. На всякий случай. Раздел со* $*$ *— факультатив. Кроме того, формулировки утверждений могут слегка отличаться от того, что дается на лекциях – в этом случае нужно следовать лекциям.*

NB домашние задания *этого семестра предполагают как решение математических задач так и реализацию алгоритмов оптимизации, о которых мы будем говорить в курсе. Последнее по желанию, но настоятельно рекомендуется. Данные к задачам будут выкладываться сюда, свой код участники могут прислать Юрию Максимову на почту*

Ссылка на примерное расписание семинаров: https://goo.gl/smCKe0

Содержание

1 Список литературы к курсу

Коллеги, данный список является ранжированием литературы с точки зрения семинариста и может существенно отличаться от рекомендованного на лекциях. Тем не менее я считаю, что правильный список именно такой

1. Лекции А.Б. Юдицкого по оптимизации. Наиболее компактное введение в предмет, весь учебный семестр вмещается в конспект из 4 лекций. Легко найти на персональной странице автора

2. Простая и понятная книга Ю.Е. Нестерова. Тут в основном методы, 2-ой семестр.

3. Лекции А.Б. Немировского. Более глубокое введение в предмет. В целом лекции Немировского я бы читал после лекций Юдицкого. Конспект на странице автора

4. Замечательная книжка Б.Т. Поляка. Если перед экзаменом (или после) будет время – рекомендую почитать.

5. Краткий конспект S. Boyd. Больше про выпуклый анализ, чем про методы.

6. Рекомендованные Вам книги А.Г. Бирюкова, В. Г. Жадана и МГУшный курс (Сухарев/Тимихов/Федоров. Курс методов оптимизации). Наверное на них стоит ориентироваться, эти книги рекомендованы Вам в качестве основных.

7. Книга Галеев Э.М. и Тихомиров В.М. Оптимизация: теория, примеры, задачи. Можно скачать в http://gen.lib.rus.ec

2 Семинар 1: Субградиентный спуск и окрестности (ЮМ)

2.1 Определения, обозначения

Напомним основные понятия, используемые в этом семинаре

Определение. Субградиентом функции $f : \mathbb{R}^n \to \mathbb{R}^1$ в точке x называется такой вектор $g(x)$, что выполнено неравенство

$$f(\bar{x}) \geq f(x) + \langle g(x), \bar{x} - x \rangle, \quad \forall \bar{x} \in \mathbb{R}^n$$

или в более удобном виде:

$$f(\bar{x}) - (f(x) + \langle g(x), \bar{x} - x \rangle) \geq 0 \quad \forall \bar{x} \in \mathbb{R}^n.$$

Важно напомнить, что вектор x обитает в прямом пространстве $E = (\mathbb{R}^n, \|\cdot\|)$, в то время как субградиент $g(x)$ (как и любая линейная функция над элементами прямого пространства) в пространстве двойственном $E^* = (\mathbb{R}^n, \|\cdot\|_*)$. Заметим, что нормы в указанных пространствах мы волльны выбирать сами и подчас это оказывается исключительно полезным.

2.2 Субградиентный метод

2.2.1 Алгоритмы безусловной оптимизации

Рассматривается задача безусловной минимизации выпуклой функции

$$f(x) \to \min_{x \in \mathbb{R}^n} \tag{1}$$?01:convex-un

Нашей ближайшей задачей задачей будет построить какой-нибудь итерационный процесс сходящийся к точке минимума функции f и, заодно, выяснить условия на функцию при которых мы можем надеятся на то, что указанный процесс существует.

Первое, что мы бы хотели иметь — это ограниченность градиента (субградиента) функции. В самом деле, давайте посмотрим на задачу от одной переменной:

[**TODO: tikz picture, 3 параболы**] Чем меньше градиент функции тем, вообще говоря, мы ближе к значению в оптимальной точке. Большой градиент дает слишком мало информации о расположении минимума внутри отрезка. Так мы вынуждены перебирать слишком много точек.

С другой стороны, даже если суб-градиент маленький, нам необходимо иное условие: мы не должны находиться слишком далеко от оптимальной точки.

Итак, два основных условия, при выполнении которых мы можем рассчитывать на успех итерационной процедуры:

- Константа Липшица градиента ограницена числом L в каждой точки области:

$$f(\bar{x}) \geq f(x) + \langle g, \bar{x} - x \rangle, \quad \forall g \in \partial f(x) : \ \|g\|_* \leq L$$

- Расстояние от точки старта до решения x^* ограничено:

$$\|x^0 - x^*\| \leq R.$$

  Отметим, что некоторые методы требуют чуть больше, а именно ограниченности множества Лебега функции. К этому вопросу мы вернемся чуть позже, когда будем разбирать быстрый градиентный метод Нестерова.

Определение. Оракул первого порядка суть машина за одну итерацию способная выдать по заданной точке x значение функции $f(x)$ и (какой-нибудь) суб-градиент $g(x)$ функции f в точке x. [1]

В основном в рамках курса мы будем заниматься **аналитической (итерационной) сложностью** алгоритмов отпимизации, а именно под сложностью алгоритма мы будем понимать числом обращений к оракулу.

Временами, в конце семестра, мы будем вспоминать об **арифметической сложности** методов оптимизации, понимаемой как число арифметических операций в рамках стандартной RAM модели

2.2.2 Субградиентый метод

Базовым шагом субградиентного метода является реккурентность

$$x_{k+1} = x_k - \alpha_k g^k \ , \quad k \geq 1, \tag{2}$$
?eq:subgradie

где x_k — текущая точка, $g^k = g(x_k) \in \partial f(x_k)$ — субградиент функции в точке x_k.

Крайне важно отметить, что особенности субградиентный метод вообще говоря не является методом спуска. Во первых направление $-g^k$ не обязательно направление убывания функции (см. задачу 1). Во вторых, если даже оно является направлением убывания функции, то шаг α_k может быть слишком большим, так что $f(x_{k+1}) > f(x_k)$. Эта ситуация типична, когда мы находимся в окрестности точки минимимума.

Таким образом необходимо следить за лучшей точкой в последовательности

$$f^{(k)}_{\text{best}} = \min\{f^{(k-1)}_{\text{best}}, f(x^{(k)})\}.$$

Давайте внимательно посмотрим на равенство 2. У него есть одна существенная проблема, состоящая в том, что в правой части равенства стоят элементы разных пространств ($x_k \in E = (\mathbb{R}^n, \|\cdot\|)$ и $g^k \in (\mathbb{R}^n, \|\cdot\|_*)$) и складывать их не корректно (в каком пространстве лежит полученный объект?). Кроме одного важного случая, когда все объекты лежат в Гильбертовом пространстве[2]. В этом случае мы ограничены нормой ℓ_2 в обоих пространствах.

Посмотрим как меняется расстояние от теукщей точки до оптимума, при совершении итерации:

$$\begin{array}{rcl} \|x_{k+1} - x^*\|_2^2 & = & \|x_k - \alpha_k g^k - x^*\|_2^2 \\ & = & \|x_k - x^*\|_2^2 - 2\alpha_k {g^k}^\top (x_k - x^*) + \alpha_k^2 \|g^k\|_2^2 \\ & \leq & \|x_k - x^*\|_2^2 - 2\alpha_k (f(x_k) - f^*) + \alpha_k^2 \|g^k\|_2^2 \qquad \text{(определение субградиента)} \\ & \leq & \|x_1 - x_*\|_2^2 - 2\displaystyle\sum_{i=1}^k \alpha_i (f(x_i) - f^*) + \sum_{i=1}^k \alpha_i^2 \|g^i\|_2^2 \end{array}$$

Откуда:

$$\sum_{i=1}^k \alpha_i^2 \|g^i\|_2^2 + \|x_1 - x_*\|_2^2 \geq 2 \sum_{i=1}^k \alpha_i (f(x_i) - f^*) \geq \left(\sum_{i=1}^k \alpha_i\right) (f^k_{best} - f^*)$$

Значит

$$f^k_{best} - f^* \leq \left(\sum_{i=1}^k \alpha_i^2 \|g^i\|_2^2 + \|x_1 - x_*\|_2^2\right) \Bigg/ \left(\sum_{i=1}^k \alpha_i\right) \leq \left(L^2 \sum_{i=1}^k \alpha_i^2 + R^2\right) \Bigg/ \left(\sum_{i=1}^k \alpha_i\right) \tag{3}$$
?eq:subgrad-e

[1] Концепция оптимизационно оракула весьма близка к концепции Тьюринговского оракула (курс Алгоритмов и моделей вычислений), который за единицу времени умеет решать задачу из оракульного класса

[2] Этот факт влечет так называемая Теорема Рица–Фреше, которую вы проходите в курсе функционального анализа

Выбор шага. Остался важный вопрос: выбор шага α_k обеспечивающий сходимость 3

- Суммируемый квадрат, но не суммируемый ряд α_k;

$$\|\alpha\|_2^2 = \sum_{k=1}^{\infty} \alpha_k^2 < \infty \qquad \sum_{k=1}^{\infty} \alpha_k = \infty,$$

  обеспечивает сходимость $f_{best}^k \to f^*$. Отметим, что правая часть 3 симметрична по своим переменным и выпукла, значит достигает минимума $(R^2 + L^2 k\alpha^2)/(2k\alpha) = LR/\sqrt{k}$ при $\alpha_1 = \cdots = \alpha_k = R/(L\sqrt{k})$.

Проблема такого выбора в его универсальности. Действительно, в худшем случае выбор оптимален и при необходимой точности ε по функции мы сойдемся за L^2R^2/ε^2 итераций.

Но в реальности норма суб-градиента L зависит от точки и может быть меньше чем используемая нами оценка. Оценка расстояния до решения – дело также весьма нетривиальное. Как правило грубая оценка это размер области на которой происходит оптимизация или размер множества подуровня функции[3]

Критерий остановки. При решении практических задач нам бы хотелось иметь критерий остановки при проверке которого на данном шаге мы можем остановить итерационный процесс.

Простой критерий можно вытащить из неравенства 3. А именно

$$f^* \geq l_k = \frac{2\sum_{i=1}^k \alpha_i f(x_i) - R^2 - \sum\limits_{i=1}^{k} \alpha_i^2 \|g^i\|_2^2}{2\sum_{i=1}^k \alpha_i} \qquad (4)$$?eq:subgrad-t

Критерий не зависит от L, но все еще зависит от расстояния до решения и не очень эффективен на практике.

К Поляку: как бороться?

Шаг по Поляку. Важным и достаточно распространным частным случаем является ситуация когда оптимальное значение f^* известно. Тогда естественное правило выбора шага происходит из аппроксимации функции рядом Тейлора:

$$f(x_k - \alpha_k g^k) \approx f(x_k) + g^{k\top}(x_k - \alpha_k g^k - x_k) = f(x_k) - \alpha_k g^{k\top} g^k.$$

Откуда в окрестности оптимальной точки $\alpha_k = \frac{f(x_k) - f^*}{\|g^k\|_2^2}$. Альтернативная мотивация следует из базового неравенства для субградиентного спуска

$$\|x_{k+1} - x^*\|_2^2 \leq \|x_k - x^*\|_2^2 - 2(f(x_k) - f^*) + \alpha_k^2 \|g^k\|_2^2.$$

Выбранный шаг (по Поляку) просто минимизирует правую часть последнего неравенства.

Algorithm 1: Субградиентный метод для решения задачи 1

Input: Выпуклая оптимизационная задача в форме 1, константа Липшица градиента L, оракул первого порядка

Output: Точка x^n и значении функции $f(x^n)$ в этой точке, что выполено $f(x^n) - \min f(x) \leq \varepsilon$

$x_1 \leftarrow$ выбираем стартовую точку (любую)

while *не выполнится критерий 4 (либо оценка на число шагов, близость к оптимуму и тп.)* **do**

  $x_{k+1} = x_k - \alpha_k g^k$

return $x_k,\ f(x_k)$

---

[3] $L_\alpha = \{x : f(x) \leq \alpha\}$ – множество Лебега, оно же множество подуровня функции f.

2.2.3 Метод проекции субградиента

Естественным обобщением задачи 1 является задача минимизации выпуклой функции на выпуклом множестве

$$
\begin{aligned}
& f(x) \to \min \\
& \texttt{s.t.:} x \in Q
\end{aligned}
\tag{5}
$$

?01:convex-co

Шаг метода (Евклидовой) проекции субградиента определяется равенством

$$
x_{k+1} = \pi_Q(x_k - \alpha_k g^k),
$$

где g^k произвольный субградиент функции f в точке x_k.

Вопрос сходимости метода решается очень просто. Пусть z_{k+1} – точка в которую нас привел шаг по анти-градиенту, $z_{k+1} = x_k - \alpha_k g^k$ (и $z_{k+1} = \pi_Q(x_k - \alpha_k g^k)$). Заметим, что

$$
\|x_{k+1} - x^*\|_2 \le \|\Pi(z_{k+1}) - x^*\|_2 \le \|z_{k+1} - x^*\|_2
$$

В то же время, для разности $\|z_{k+1} - x^*\|$ справедлива базовая оценка субградиентного метода

$$
\|z_{k+1} - x^*\|_2 \le \|x_k - x^*\|_2^2 - 2\alpha_k(f(x_k) - f^*) + \alpha_k^2 \|g^k\|_2^2
$$

Сложив телескопическую сумму получаем ту же оценку, что и для субградиентного метода.

2.2.4 В чем (не)-проблема?

2.3 Градиентный спуск

2.3.1 Методы одномерной минимизации

Будет дописано ЮМ

2.3.2 Правила выбора шага и сходимость

2.4 *Зеркальный спуск

2.4.1 *Сопряженные функции и их свойства

2.4.2 *Градиентное отображение и обратное к нему

2.4.3 *Алгоритм и его анализ

2.4.4 *Стохастическая версия

2.5 Домашнее задание

2.5.1 Задачи на повторение

2.5.2 Задачи на бумажке

- Как изменяться оценки (худшего случая) сходимости (числа итераций) метода проекции субградиента, если проектирование на множество Q осуществляется не на каждом шаге, а на k-ом шаге, $k > 1$?
- Верно-ли что анти-субградиент выпуклой функции в некоторой точке может не быть направлением ее убывания. Может ли быть эта точка отличной от точки минимума функции?

- Верно ли, что для выпуклой гладкой функции $f(x_1, x_2) = g(x_1) + h(x_2)$ точка минимума (x_1^*, x_2^*) может быть вычислена минимизацией (выпуклых гладких) функций $g(x_1)$ и $h(x_2)$ в отдельности? Предполагая, что вычисление $g, h, \nabla g, \nabla h$ могут быть выполнены за единицу времени каждое. Если верно, то оцените выигрыш в числе операций от этой схемы (считая, что нужно найти (x_1, x_2) такое что $f(x_1, x_2) - \min\limits_{x_1, x_2} f(x_1, x_2) \leq \varepsilon$)

2.5.3 Задачи за компьютером

3 Семинар 2: Ньютоновские и квази-ньютоновские методы (ИУ)

3.1 Методы спуска. Скорости сходимости

Метод спуска - строит последовательность $\{x_k\} : f(x_{k+1} < f(x_k)$.

Опр. *Направление убывания* $f(x)$ в точке $x \in R^n$ - ненулевой вектор $s \in R^n : f(x + \alpha s) < f(x)$ для достаточно малых $\alpha > 0$ Множество всех направлений убывания - конус K_d

$$\forall s \in K_d \;\; \langle \nabla f(x), s \rangle \leq 0$$
$$\forall s : \langle \nabla f(x), s \rangle < 0 \Rightarrow s \in K_d$$

(Док-во - упр. на паре)

Опр. *Итерационные методы спуска* - задаются:

$$x_{k+1} = xk + \alpha_k s_k, \quad s_k \in K_d(x_k), \alpha_k > 0$$

4 Правила выбора длины шага α_k

4.0.1 Одномерная минимизация

$$\alpha_k = arg \min_{\alpha \geq 0} f(x_k + \alpha s_k) = arg \min_{\alpha \geq 0} \varphi(\alpha)$$

4.0.2 Точное решение

$$\phi'(\alpha_k) = \langle \nabla f(x_{k+1}), s_k \rangle = 0$$

Если $f(x)$ - квадратичная сильно выпуклая:

$$f(x) = \frac{1}{2} \langle x, Ax \rangle + \langle b, x \rangle,$$

$$\nabla f(x) = Ax + b$$

A - симметричная, положительно определенная, то

$$\alpha_k = -\frac{\langle \nabla f(x_k), s_k \rangle}{\langle s_k, As_k \rangle}$$

4.0.3 Правило Армихо.

Приближенное решение. (картинка стр.32) Задаем:

$$0 < \varepsilon < 1, \quad 0 < \theta < 1$$

Шаг 1 Проверяем, достаточно ли быстро уменьшается значение функции:

$$f(x_k + \alpha s_k) \leq f(x_k) + \varepsilon\alpha\langle\nabla f(x_k), s_k\rangle$$

Шаг 2 Если нет - уменьшаем $\alpha = \theta\alpha$
. Если да, заканчиваем, $\alpha_k = \alpha$

(Доказать, что процедура завершится)

4.0.4 Правило постоянного шага. Априорного выбора.

4.0.5 Правило Голдстейна.

Выбираем $0 < \varepsilon_1 < \varepsilon_2 < 1$ (картинка стр. 34) $\quad \alpha_k$ должно удовлетворять

$$\varepsilon_1 < \frac{f(x_k + \alpha s_k) - f(x_k)}{\alpha\langle\nabla f(x_k), s_k\rangle} < \varepsilon_2$$

4.1 Метод градиентного спуска

метод 1 порядка Выбираем направление наискорейшего локального убывания функции! Направление - антиградиент

$$s_k = -\nabla f(x_k)$$

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k)$$

Метод наискорейшего спуска - правило одномерной минимизации
(напишите метод наискорейшего спуска для квадратичной функции)

По факту градиентный шаг является спуском в минимум параболы:

$$x_{k+1} = arg \min_{x \in Q} \langle\nabla f(x_k), x - x_k\rangle + \frac{1}{2\alpha_k}\|x - x_k\|^2$$

Скорость сходимости - линейная. Обобщение - субградиентный.

5 Метод Ньютона

метод 2го порядка

5.1 Классический

На каждом шаге аппроксимируем функцию $f(x)$ квадратичной

$$\phi(x) \approx f(x_k) + \langle\nabla f(x_k), x - x_k\rangle + \frac{1}{2}f''(x_k)\|x - x_k\|^2$$

и минимизируем ее.

$$\nabla\phi(x_{k+1}) = 0$$

$$\nabla f(x_k) + f''(x_k)(x_{k+1} - x_k) = 0$$
$$x_{k+1} = x_k - f''(x_k)^{-1}\nabla f(x_k)$$

Метод локальный. (Упр. - привести пример когда зацикливается)

5.1.1 Теорема

Покажем, что скорость сходимости квадратична в случае, когда мы находимся в достаточно маленькой окрестности $x^* : f''(x) > 0$ и $\|f''^{-1}\| < C$ - ограничена.

$$0 = f'(x^*) = f'(x_k - \varepsilon_k) = f'(x_k) + f''(x_k)(x_k - x^*) + O(\|\varepsilon\|^2)$$
$$0 = f'(x_k) - f''(x_k)\varepsilon_k + O(\|\varepsilon\|^2)$$

Из-за ограниченности:

$$0 = f''^{-1}(x_k)f'(x_k) - \varepsilon_k + O(\|\varepsilon_k\|^2) =$$

Но

$$= x_k - x_{k+1} - \varepsilon_k = x^* - x_{k+1} = -\varepsilon_{k+1}$$

Отсюда:

$$\varepsilon_{k+1} = O(\|\varepsilon_k\|^2)$$
$$\|\varepsilon_{k+1}\| \leq c\|\varepsilon_k\|^2$$

Если окрестность настолько маленькая, что $\|\varepsilon_k\| \leq \frac{\alpha}{c}, 0 < \alpha < 1$, то видно что метод сходится, и скорость квадратичная.

Однако, несмотря на свою предельнуют естественную интерпретацию, Метод ньютона имеет несколько скрытых недостатков. Прежде всего, может случиться, что в текущей точке этого процесса гессиан целевой функции является вырожденным. Тогда в методе происходит аварийная остановка. Во вторых, может случиться, что метод расходится, или сходится к седловой точке, или к точке лок максимума.

5.2 Регуляризация Левенберга - Маркварта.

Если матрица $\nabla^2 f$ не является положительно определенной, ее можно регуляризовать с помощью единичной матрицы. То есть в шаге метода вместо гессиана использовать:

$$\nabla^2 f(x) + \gamma I > 0$$

$$x_{k+1} = x_k - [\nabla^2 f(x_k) + \gamma I]^{-1}\nabla f(x_k)$$

Иногда эта стратегия интерпретируется как комбинация градиентного метода с методом Ньютона.

5.3 С переменным шагом(Демпфированный метод Ньютона)

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

α_k - выбирается тем или иным способом. Например, правилом Армихо.

Теорема. Демпфированный метод Ньютона с правилом выбора шага Армихо(или одномерной минимизации, наискорейший спуск) для сильно выпуклой функции сходится из любой точки, со сверхлинейной скоростью, в случае липшицева второй производной - с квадратичной.

5.4 Доверительные области

В соответствии с этим подходом вокруг точки x_k надо зафиксировать окрестность, в которой аппроксимация второго порядка обязана быть достаточно хорошей. Эта окрестность $\Delta(x_k)$ называется доверительной областью. Можно, например, взять $\Delta(x_k) = \{x : \|x - x_k\| \leq \varepsilon, \quad \varepsilon > 0\}$. Тогда следующая точка x_{k+1} будет выбираться как решение задачи

$$\min_{x \in \Delta(x_k)} \left[\langle \nabla f(x_k), x - x_k \rangle + \frac{1}{2} \langle x - x_k, f''(x_k)(x - x_k) \rangle \right]$$

Если $\Delta(x) = R^n$, то это в точности стандартный ньютоновский шаг.

5.5 Квазиньютоновские методы

1 порядка

Идея в том, чтобы не вычислять каждый раз обратную матрицу вторых производных а постепенно приближаться к ней, итеративно накапливая информацию.

$$H_{k+1} = H_k + \Delta H$$

Тейлор:

$$\Delta y_k \approx \nabla f(x_k) - \nabla f(x_{k+1}) = f''(x_k) \|x_k - x_{k+1}\|$$

$$\Delta x_k \approx f''(x_k)^{-1} \Delta y_k$$

$$\Delta x_k = H_{k+1} \Delta y_k = H_k \Delta y_k + \Delta H_k \Delta y_k$$

$$\Delta H_k \Delta y_k = \Delta x_k - H_k \Delta y_k = p_k$$

- это - *квазиньютоновское условие*

Существует много способов удовлетворить этому равенству.

5.5.1 Метод Бройдена

Одноранговый апдейт

$$\Delta H_k = \mu_k q_k q_k^T$$

Это значит, что

$$\mu_k q_k q_k^T \Delta y_k = p_k$$

Оно выполняется, если:

- $q_k^T y \neq 0$
- $q_k = p_k$

Тогда:

$$\mu_k = (q_k^T \Delta y_k)^{-1}$$

В итоге получаем матрицу приращения:

$$\Delta H_k = \frac{(\Delta x_k - H_k \Delta y_k) (\Delta x_k - H_k \Delta y_k)^T}{\langle \Delta x_k - H_k \Delta y_k, \Delta y_k \rangle}$$

5.5.2 Метод Дэвидон-Флетчер-Пауэл

Двуранговый апдейт

$$\Delta H_k = \mu_1 \Delta x_k \Delta x_k^T + \mu_2 H_k \Delta y_k (H_k \Delta y_k)^T$$

Если

$$\mu_1 = \langle \Delta x_k, \Delta y_k \rangle^{-1}$$
$$\mu_2 = -\langle H_k \Delta y_k, \Delta y_k \rangle^{-1}$$

то квазиньютоновское условие выполняется.

5.5.3 Метод BFGS (Бройден - Флетчер - Гольдфарб - Шанно)

Двуранговый апдейт

$$G_k = H_k^{-1} \approx f''(x_k)$$
$$\Delta G_k = \frac{\Delta y_k \Delta y_k^T}{\Delta y^T \Delta x} + \frac{G_k \Delta x_k \Delta x_k^T G_k}{\Delta x_k^T G_k \Delta x_k}$$

Геометрическая интерпретация в R^1 - метод секущих.