



Лекция 7

Численные методы минимизации регуляризирующих функционалов

Насонов Андрей Владимирович

<http://imaging.cs.msu.ru/>

Laboratory of Mathematics Methods of Image Processing
Department of Computational Mathematics and Cybernetics
Lomonosov Moscow State University

Задача восстановления изображений

- Общая математическая модель:

$$u = Az + n$$

Наблюдаемое изображение

Оператор, искажающий изображение

Реальное изображение

Шум

- Требуется найти z при известном u
- Задача является некорректно поставленной



Задача восстановления изображений

■ Некорректно поставленная задача

$$Az = u$$

- Решение не существует
- Решение не является единственным
- Неустойчивость

■ Метод регуляризации Тихонова

$$z_{\alpha} = \arg \min_z (\|Az - u\| + \alpha \Omega[z])$$

- с одной стороны, ищется z такое, чтобы оно соответствовало наблюдаемому изображению u
- с другой стороны, минимизируется штрафная функция

Минимизация регуляризующего функционала

■ Общий случай

- $f(z): \mathbb{R}^n \rightarrow \mathbb{R}$

- f – строго выпуклый функционал

$$f(\lambda z_1 + (1 - \lambda)z_2) < \lambda f(z_1) + (1 - \lambda)f(z_2) \\ 0 < \lambda < 1$$

- $z_R: \min f(z) = f(z_R)$

- значение z_R существует и единственно

- Если $f(z)$ дифференцируема в z_R , тогда $\nabla f(z_R) = 0$

- $\nabla f(z) = \left(\frac{\partial f}{\partial z_1}, \frac{\partial f}{\partial z_2}, \dots, \frac{\partial f}{\partial z_n} \right)^T$ – градиент

Дифференцирование

■ Производная Гато ($f: Z \rightarrow U$)

$$df(z, h) = \lim_{t \rightarrow 0} \frac{f(z + th) - f(z)}{t}$$

$df(z, h)$ – слабый дифференциал f в z

Если $df(z, h)$ – линейный оператор $f'(z)$,

$$df(z, h) = f'(z)h,$$

тогда $f'(z)$ называется слабой производной (производной Гато) в z .

Дифференцирование

■ Производная Фреше ($f: Z \rightarrow U$)

$$\lim_{h \rightarrow 0} \frac{\|f(z+h) - f(z) - Ah\|_U}{\|h\|_Z} = 0$$

$$f(z+h) = f(z) + Ah + o(h)$$

- $A: Z \rightarrow U$ – ограниченный линейный оператор, называемый производной f в z : $Df(z) = A$
- Совпадает с производной Гато
- Для $U = \mathbb{R}$

$$Ah = (\nabla f, h)$$

$$f(z+h) = f(z) + (\nabla f, h) + o(h)$$

Дифференцирование

- В конечномерном пространстве производная определяется матрицей Якоби

$$f(z): \mathbb{R}^n \rightarrow \mathbb{R}^m$$

$$[Df(z)](h) = J_f(z)h = \sum_{i=1}^n h_i \frac{\partial f}{\partial z_i}(z)$$

$$J_f(z) = \begin{pmatrix} \frac{\partial f_1}{\partial z_1} & \dots & \frac{\partial f_1}{\partial z_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial z_1} & \dots & \frac{\partial f_m}{\partial z_n} \end{pmatrix}$$



Дифференцирование

■ Свойство линейности

- $D(f + g)(z) = Df(z) + Dg(z)$
- $D(\lambda f)(z) = \lambda Df(z)$

■ Производная композиции функционалов

- $f: Z \rightarrow W, g: W \rightarrow U$ – дифференцируемые функционалы
- $g \circ f: Z \rightarrow U$ – композиция
- $D(g \circ f)(z) = Dg(f(z)) \circ Df(z)$

Дифференцирование

■ Примеры

- $D(\|z\|_2^2) = 2z^T$
- $D(Az) = A$, если A – линейный оператор
- $D(\|Az\|_2^2) = 2(Az)^T \circ A = (2A^T Az)^T$
- $D(\|Az - u\|_2^2) = 2(Az - u)^T \circ A = (2A^T Az - A^T u)^T$

■ Если A – оператор свёртки $Az = z * H$, тогда

$$A^T z = z * H^*,$$

$$\text{где } H^*(x, y) = H(-x, -y)$$

Дифференцирование

■ Частный случай

$$f(z) = \arg \min_z (\|Az - u\|_2^2 + \alpha \|Bz\|_2^2)$$

– B – линейный оператор, например оператор Лапласа Δ

■ Тогда сводится к решению линейного уравнения

$$A^T(Az - u) + \alpha B^T Bz = 0$$

$$(A^T A + \alpha B^T B)z = A^T u$$

положительно
определённая матрица

- Матрица огромных размеров, обращение невозможно
- Решение только итерационное

Дифференцирование

■ Проблема

$$D(TV(z)) = ?$$

– функционал не является дифференцируемым

■ Требуется вычисление производной функции

$$f(x) = |x| \text{ в } x = 0$$

– Один из способов – приближение:

$$f_\varepsilon(x) = \sqrt{x^2 + \varepsilon^2}$$

– Недостаток: высокая вычислительная сложность

Дифференцирование

■ Субдифференциал

– $f(z): Z \rightarrow \mathbb{R}$,

– f – выпуклый функционал

$$f(\lambda z_1 + (1 - \lambda)z_2) \leq \lambda f(z_1) + (1 - \lambda)f(z_2)$$

$$0 < \lambda < 1$$

■ Множество элементов g , для которых

$$f(z) - f(z_0) \geq (g, z - z_0)$$

для всех $z \in Z$ называется субдифференциалом, а каждый из элементов g – субградиентом $f(z)$ в z_0

Дифференцирование

■ Свойства субдифференциала

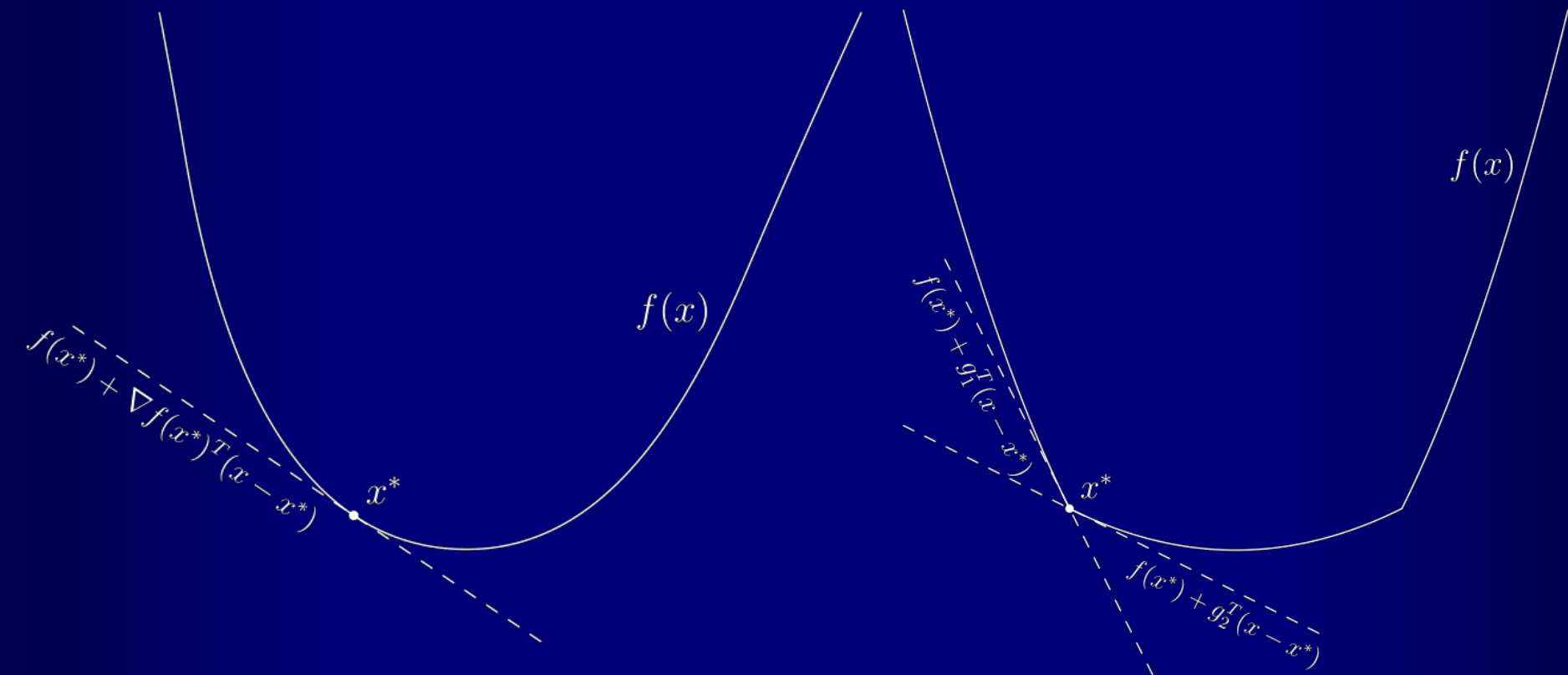
- $\partial f(z)$ – возможно пустое, выпуклое множество
- Линейность:
 - $\partial(f_1(z) + f_2(z)) = \partial f_1(z) + \partial f_2(z)$
 - $\partial(\lambda f(z)) = \lambda \partial f(z), \lambda > 0$
- $f(z): Z \rightarrow \mathbb{R}$ выпукла и непрерывна в z_0
 $\Rightarrow f(z)$ субдифференцируема в z_0 , $\partial f(z_0) \neq \emptyset$,
 $\partial f(z_0)$ – компактное и выпуклое множество

Дифференцирование

■ Свойства субдифференциала

- $f(z): Z \rightarrow \mathbb{R}$ выпукла и конечна
- Тогда $f(z)$ имеет производную Гато в $z_0 \in Z$ тогда и только тогда, когда её субдифференциал состоит из одной точки и равен градиенту $\partial f(z_0) = \{\nabla f(z_0)\}$
- $f(z)$ имеет локальный минимум в z_0 тогда и только тогда, когда 0 принадлежит субдифференциалу в z_0

Дифференцирование



Дифференцируемая функция
Субдифференциал состоит из одного
субградиента, совпадающего с
производной

Недифференцируемая функция
Субдифференциал – множество
субградиентов

Дифференцирование

■ Примеры:

$$f(x) = |x|$$
$$f'(x) = \begin{cases} 1, & x > 0, \\ [-1, 1], & x = 0, \\ -1, & x < 0. \end{cases}$$

– Многомерный случай

$$D(\|z\|_1) = (f'(z_1), f'(z_2), \dots, f'(z_n))$$

Дифференцирование

- Нахождение субдифференциала функционала полной вариации

$$D(TV[z]) = ?$$

$$TV[z] = |||\nabla z|||_1 = \sum_{i,j} |\nabla z_{i,j}|$$

- Проблема численного дифференцирования на дискретной двумерной сетке

$$\nabla z = (z_x, z_y) = ?$$

Дифференцирование

■ Простейшие разностные способы вычисления производной

- 1. Обычная производная

$$\nabla z_{i,j} = (z_{i+1,j} - z_{i,j}, z_{i,j+1} - z_{i,j})$$

- 2. Симметричная производная

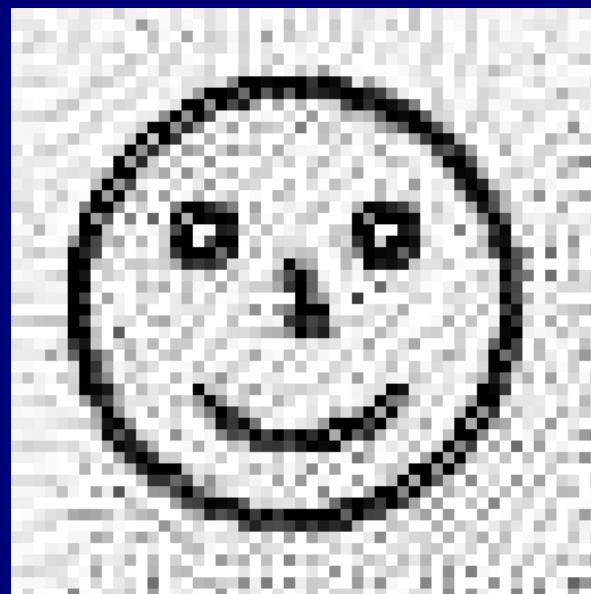
$$\nabla z_{i,j} = \frac{1}{2} (z_{i+1,j} - z_{i-1,j}, z_{i,j+1} - z_{i,j-1})$$

Дифференцирование

- Проблемы при использовании простейших способов дифференцирования в обработке изображений



Использование несимметричной производной приводит к неравномерной обработке и смазу



Использование симметричной производной приводит к шуму в виде сетки из-за того, что центральный пиксель не участвует в вычислениях

Дифференцирование

■ Идея

- Вместо вычисления модуля градиента в каждом пикселе считать суммарную вариацию
- В одномерном случае:

$$TV[f] = \int_{-\infty}^{+\infty} |f'(x)| dx = \sum_{k=1}^K |f(x_k) - f(x_{k-1})|,$$

$\{x_k\}$ – множество всех локальных экстремумов

- Поэтому

$$TV[z] = \sum_i |z_{i+1} - z_i|$$

Дифференцирование

- В двумерном случае:

$$TV[z] = \sum_{i,j} |z_{i+1,j} - z_{i,j}| + \sum_{i,j} |z_{i,j+1} - z_{i,j}|$$

– такой способ приводит к небольшой анизотропии

- Более аккуратный вариант – сумма разностных производных по различным направлениям

$$BTV[z] = \sum_{(x,y) \in Q} \frac{1}{\sqrt{x^2 + y^2}} \sum_{i,j} |z_{i+x,j+y} - z_{i,j}|$$

– Множество Q – множество смещений

– Например, $Q = \{(1,0), (0,1), (1,1), (1,-1)\}$

Дифференцирование

■ Запись в операторном виде

$$BTV[z] = \sum_{(x,y) \in Q} \frac{1}{\sqrt{x^2 + y^2}} \|S_X^x S_Y^y z - z\|_1,$$

– $S_X z$ – оператор сдвига на 1 пиксель по горизонтали:

$$(S_X z)_{i,j} = z_{i+1,j}$$

– $S_Y z$ – оператор сдвига на 1 пиксель по вертикали:

$$(S_Y z)_{i,j} = z_{i,j+1}$$

Дифференцирование

■ Вычисление субградиента

$$D(S_X) = (S_X)^T = S_X^{-1}$$

$$D(S_Y) = (S_Y)^T = S_Y^{-1}$$

$$D(BTV[z]) = \sum_{(x,y) \in Q} \frac{1}{\sqrt{x^2+y^2}} (S_X^{-x} S_Y^{-y} - I) \operatorname{sgn}(S_X^x S_Y^y z - z)$$

- I – единичный оператор
- $\operatorname{sgn}(z)$ – применение функции sgn поэлементно:

$$\operatorname{sgn}(z)_{i,j} = \begin{cases} 1, & z_{i,j} > 0, \\ 0, & z_{i,j} = 0, \\ -1, & z_{i,j} < 0. \end{cases}$$

Полная обобщённая вариация

$$\text{TGV}[z] = \min_v (\alpha_1 \|\nabla z - v\|_1 + \alpha_2 \|\nabla v\|_1)$$

$$z_\alpha = \arg \min_{z,v} (\|Az - u\| + \alpha_1 \|\nabla z - v\|_1 + \alpha_2 \|\nabla v\|_1)$$

- Проблема: минимизация по двум переменным
 - 1. Минимизация в пространстве $Z \times V$
 - 2. Поочерёдная минимизация сначала по z , затем по v .

Полная обобщённая вариация

- Представление в следующем виде:

$$\text{TGV}[z] = \alpha_1 \text{BTV}[z] + \alpha_2 \text{BTV}_2[z]$$

$$\text{BTV}[z] = \sum_{(x,y) \in Q} \frac{1}{\sqrt{x^2 + y^2}} \|S_X^x S_Y^y z - z\|_1$$

$$\text{BTV}_2[z] = \sum_{(x,y) \in Q} \frac{1}{\sqrt{x^2 + y^2}} \|S_X^x S_Y^y z + S_X^{-x} S_Y^{-y} z - 2z\|_1$$

Методы минимизации

- Общий метод субградиентного спуска
 - $f(z): Z \rightarrow \mathbb{R}$, f – выпуклый функционал

$$z^{(k+1)} = z^{(k)} - \beta_k g^{(k)}$$

$g^{(k)}$ – любой субградиент из $\partial f(z^{(k)})$

$\beta_k > 0$ – шаг на k -й итерации

Методы минимизации

■ Общий метод субградиентного спуска

- Не гарантируется монотонное уменьшение $f(z^{(k)})$
- Рассматривается последовательность

$$f_{best}^{(k)} = \min \left\{ f_{best}^{(k-1)}, f(z^{(k)}) \right\}$$

Она монотонно убывает и ограничена $\Rightarrow$ имеет предел

Методы минимизации

Проблема выбора шага

■ Простейшие способы

- 1. Постоянный коэффициент: $\beta_k = \beta$
- 2. Постоянная норма шага: $\beta_k = \frac{\gamma}{\|g^{(k)}\|_2}$
- 3. Последовательность β_k , суммируемая в квадратах

$$\beta_k > 0, \quad \sum_{k=1}^{\infty} \beta_k = \infty, \quad \sum_{k=1}^{\infty} \beta_k^2 < \infty$$

Пример:

$$\beta_k = \frac{a}{b+k}, \quad a > 0, b \geq 0$$

Методы минимизации

Проблема выбора шага

■ Простейшие способы

- 4. Несуммируемая бесконечно убывающая последовательность коэффициентов

$$\beta_k > 0, \quad \sum_{k=1}^{\infty} \beta_k = \infty, \quad \lim_{k \rightarrow \infty} \beta_k = 0$$

Пример:

$$\beta_k = \frac{a}{\sqrt{k}}, \quad a > 0.$$

- 5. Несуммируемая бесконечно убывающая последовательность шагов

$$\gamma_k > 0, \quad \sum_{k=1}^{\infty} \gamma_k = \infty, \quad \lim_{k \rightarrow \infty} \gamma_k = 0$$



Методы минимизации

Проблема выбора шага

■ УСЛОВИЯ СХОДИМОСТИ

- $f(z)$ – липшицева функция

$$f(z_1 - z_2) \leq G \|z_1 - z_2\|_2 \Rightarrow \|g^{(k)}\|_2 \leq G \quad \forall k$$

- начальное приближение задано с точностью R :

$$\|z^{(1)} - z_T\| \leq R$$

$$f_{best}^{(k)} - f(z_T) \leq \frac{R^2 + G^2 \sum_{i=1}^k \beta_i^2}{2 \sum_{i=1}^k \beta_i}$$



Методы минимизации

Проблема выбора шага

■ Условия сходимости

- 1. Постоянный коэффициент – сходимость к субоптимальному решению:

$$f_{best}^{(k)} - f(z_T) \rightarrow G^2 \beta / 2$$

$f_{best}^{(k)} - f(z_T) \leq G^2 \beta$ как максимум
через $R^2 / (G^2 \beta^2)$ итераций

- 2. Постоянный шаг

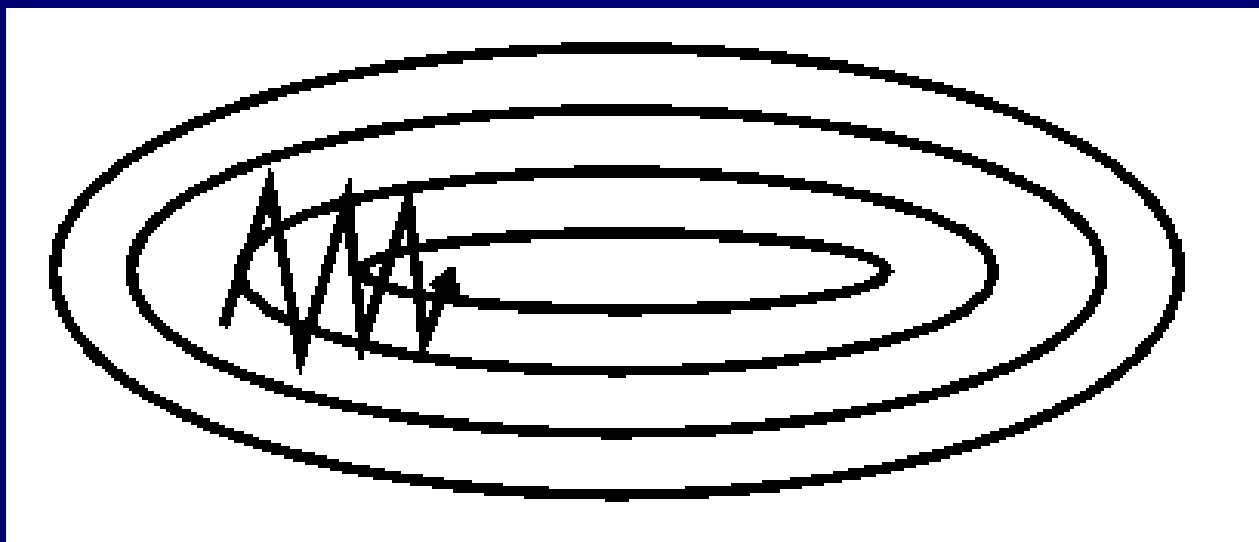
$$f_{best}^{(k)} - f(z_T) \rightarrow G \gamma / 2$$

- 3, 4, 5. Остальные случаи

$$f_{best}^{(k)} - f(z_T) \rightarrow 0$$

Проблема медленной сходимости

- Субградиент не всегда указывает в сторону настоящего минимума



- Возможны сильные осцилляции при медленном движении в сторону минимума

Методы моментов

■ Классический метод моментов (I)

$$\begin{aligned}g^{(k)} &= \nabla f(z^{(k)}) \\v^{(k+1)} &= \mu v^{(k)} - g^{(k)} \\z^{(k+1)} &= z^{(k)} + \beta_k v^{(k+1)}\end{aligned}$$

- Идея: пытаться продолжать двигаться в том же направлении
- Сложность выбора параметра μ ($0 \leq \mu < 1$)

Методы моментов

■ Классический метод моментов (II)

$$\begin{aligned}g^{(k)} &= \nabla f(z^{(k)}) \\v^{(k+1)} &= \mu v^{(k)} - \beta_k g^{(k)} \\z^{(k+1)} &= z^{(k)} + v^{(k+1)}\end{aligned}$$

- Идея: пытаться продолжать двигаться в том же направлении
- Сложность выбора параметра μ ($0 \leq \mu < 1$)

Методы моментов

■ Метод ускоренного градиента Нестерова (I)

$$g^{(k)} = \nabla f(z^{(k)} + \beta_{k-1} \mu v^k)$$

$$v^{(k+1)} = \mu v^{(k)} - g^{(k)}$$

$$z^{(k+1)} = z^{(k)} + \beta_k v^{(k+1)}$$

- Идея: коррекция ошибки и применение градиента к предыдущей точке

Методы моментов

■ Метод ускоренного градиента Нестерова (II)

$$g^{(k)} = \nabla f(z^{(k)} + \mu v^k)$$

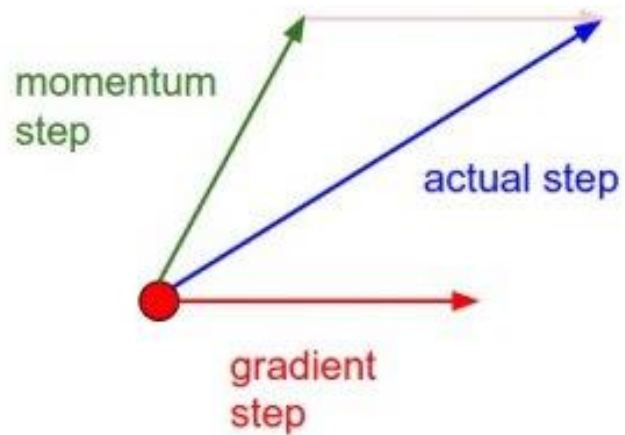
$$v^{(k+1)} = \mu v^{(k)} - \beta_k g^{(k)}$$

$$z^{(k+1)} = z^{(k)} + v^{(k+1)}$$

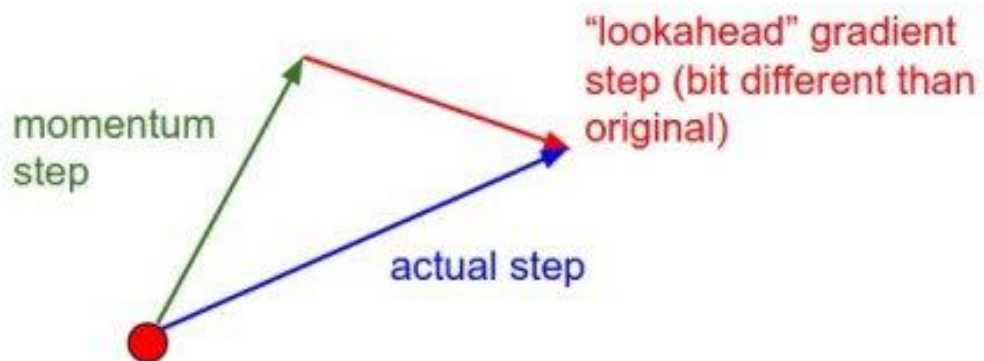
- Идея: коррекция ошибки и применение градиента к предыдущей точке

Методы моментов

Momentum update



Nesterov momentum update



Результаты



Размытое изображение
Ядро размытия – фильтр Гаусса с $\sigma = 3$

Результаты

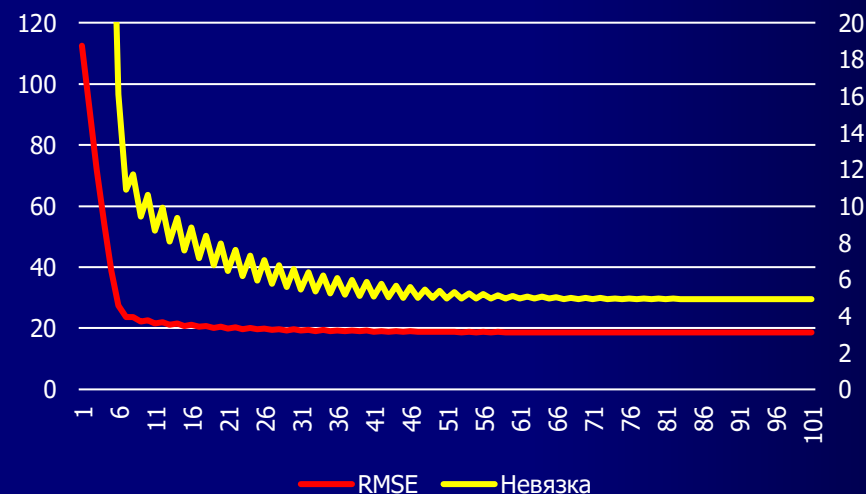
- Регуляризирующий функционал

$$z_{\alpha} = \arg \min_z (\|Az - u\|_1 + \alpha_1 BTV[z] + \alpha_2 BTV_2[z])$$

$$\alpha_1 = 0,01, \quad \alpha_2 = 0,01$$

Результаты

■ Простой субградиентный метод

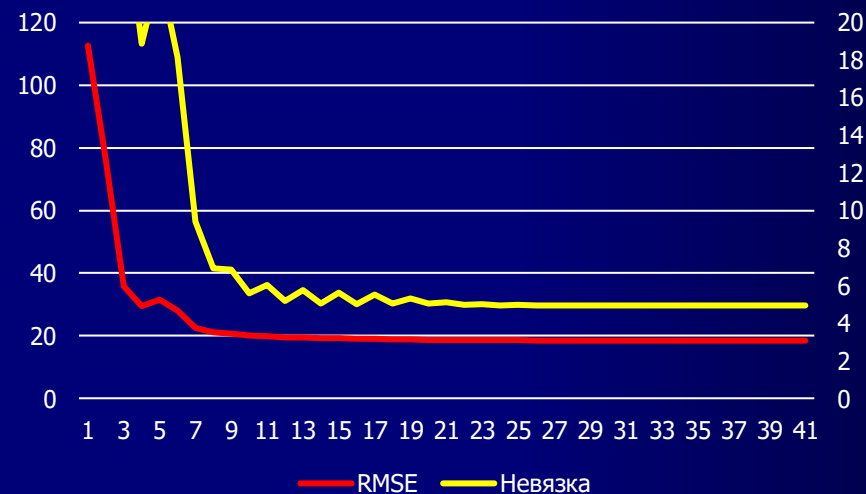


100 итераций

$$\gamma_k = 25 \cdot 0,01^{k/100}$$

Результаты

■ Метод моментов Нестерова



40 итераций, $\mu = 0,85$

$$\gamma_k = 25 \cdot 0,01^{k/40}$$

Результаты

