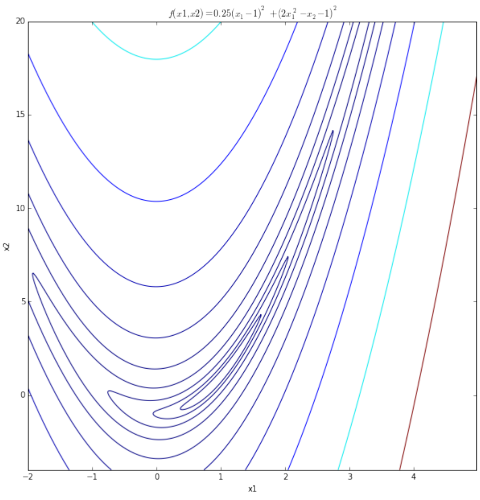
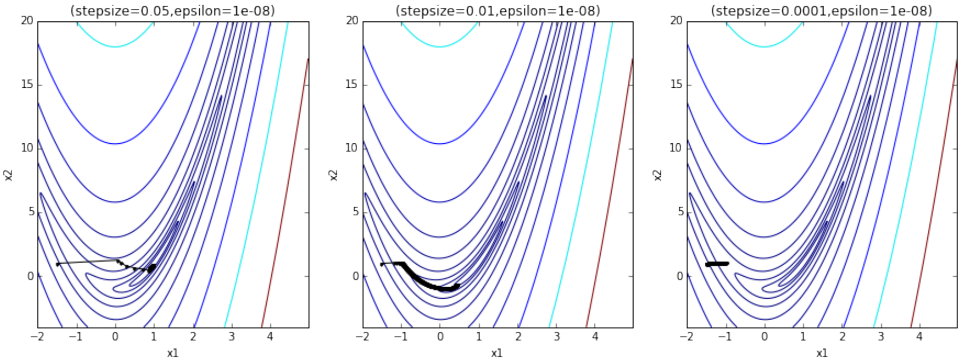
Ластовичек Дмитрий. АД
Методы оптимизации
Домашнее задание №2

Везде в качестве нормы использовалась норма L_∞
Вычислим первые производные:
$\frac{df}{dx_1} = 0.5(x_1 - 1) + 2(2x_1^2 - x_2 - 1)4x_1 = 16x_1^3 - (8x_2 + 7.5)x_1 - 0.5$
$\frac{df}{dx_i}|_{n>i>1} = -2(2x_{i-1}^2 - x_i - 1) + 2(2x_i^2 - x_{i+1} - 1)4x_i = 16x_i^3 - (8x_{i+1} + 6)x_i - 2(2x_{i-1}^2 - 1)$
$\frac{df}{dx_n} = -2(2x_{n-1}^2 - x_n - 1) = 2x_n - 4x_{n-1}^2 + 2$
Вычислим вторые производные:
$\frac{d^2f}{dx_1^2} = 48x_1^2 - 8x_2 - 7.5$
$\frac{d^2f}{dx_i^2}|_{n>i>1} = 48x_i^2 - 8x_{i+1} - 6$
$\frac{d^2f}{dx_n^2} = 2$
$\frac{d^2f}{dx_i dx_j}|_{|i-j|>1} = 0$
$\frac{d^2f}{dx_i dx_{i+1}} = -8x_i$
Линии уровня для n=2

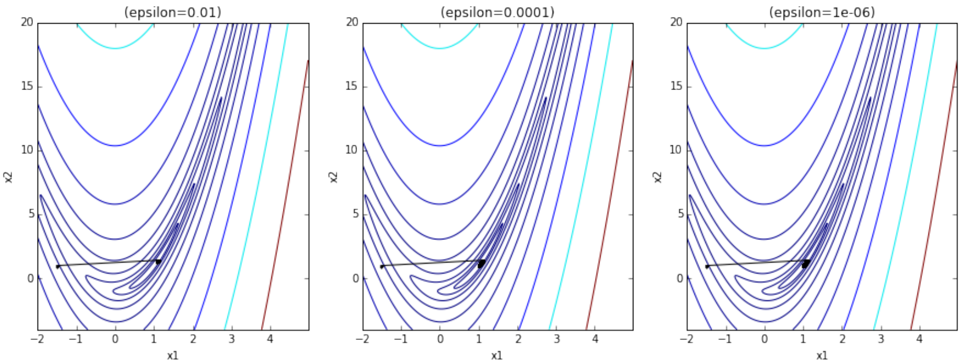


Метод градиента



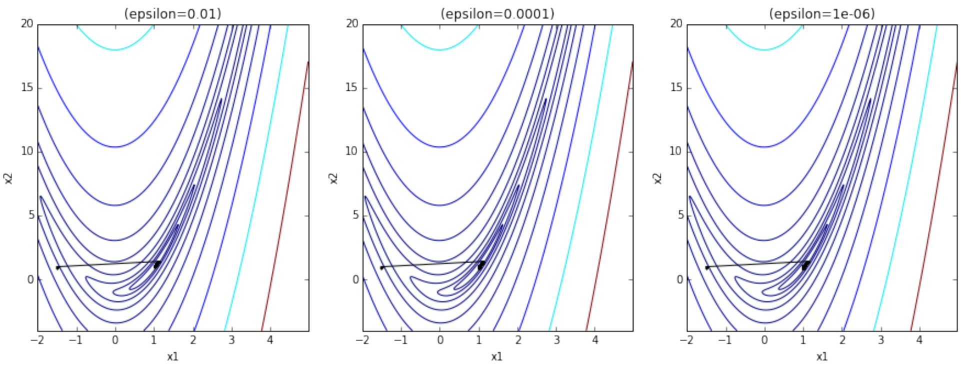
На изображениях видно, что скорость сходимости градиентного метода сильно зависит от выбора шага. При этом в случае неудачно выбранного шага метод может расходится, или сходится чересчур медленно(см. 3 изображение)

Метод наискорейшего спуска



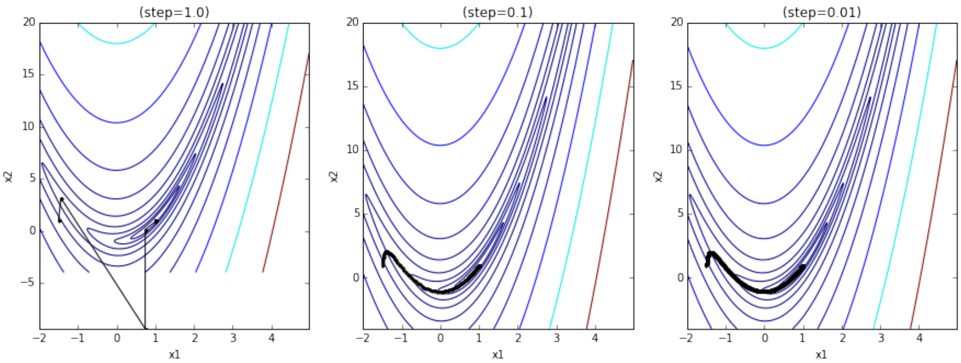
В сравнении с градиентным методом, метод наиск. спуска практически сходится за пару итераций. На данных изображениях видно, что скорость сходимости зависит от критерия остановки(ϵ)

Метод сопряженного градиента



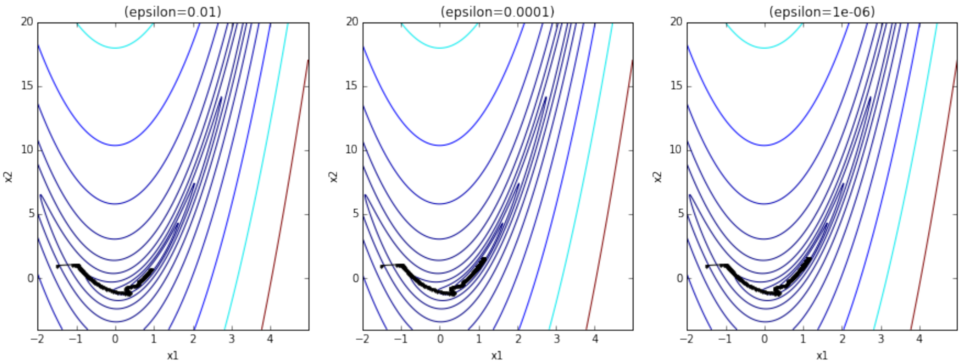
Реализация метода сопр. градиента для произвольной функции близка к методу наиск. спуска. В данном случае лишь добавляется некая "инерционность"от предыдущих шагов. На изображениях видим что сходится данный метод быстрее метода Наиск. спуска. Однако в разных случаях, как мы увидим ниже, возможно обратное

Метод Ньютона



Из исследований данного метода выяснилось, что рекомендуется добавлять в метод параметр шага, иначе метод может расходится. на изображении 1 видим реализацию чистого метода Ньютона без шага. Видим, что метод ведет себя довольно непредсказуемо. При выборе удачного шага метод близок к методу градиент. спуска. Явный недостаток метода заключается в том, что на больших размероностях гессиан (который при этом нужно вычислять) может быть плохо обусловлен и не обращается, таким образом метод перестает работать.

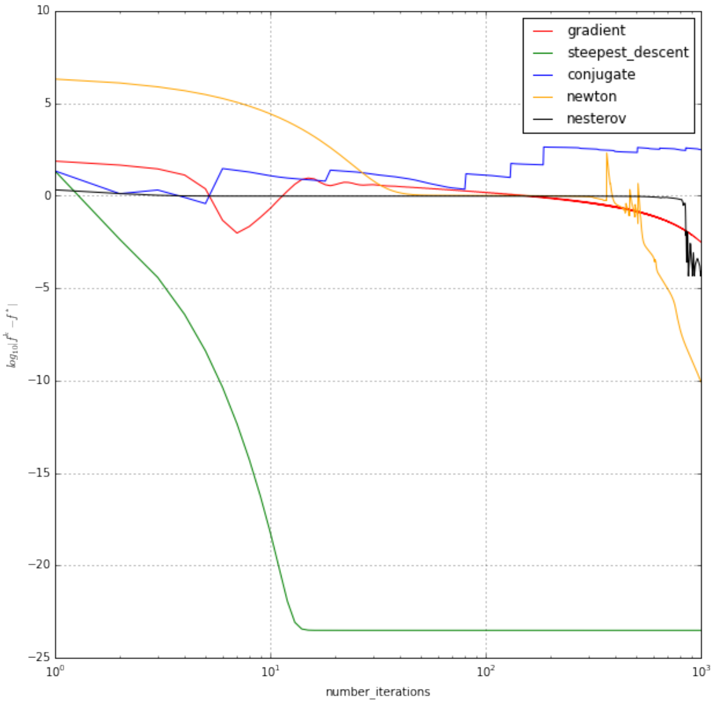
Метод Нестерова



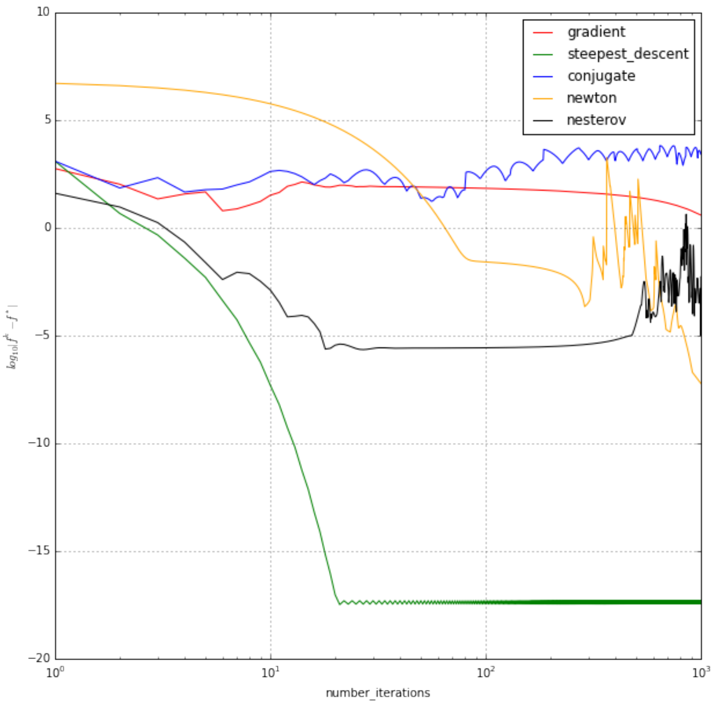
Сходимость

	Метод	n	N	||grad f(x')||	||x* - x'||	|f* - f(x')|	Сошелся?
0	gradient	2	9674	9.985821e-09	3.545143e-07	1.822988e-15	True
1	gradient	5	6446516	2.462305e-01	2.188763e+00	1.584599e-03	False
2	gradient	13	4277649	6.626776e+00	3.698908e+00	1.511535e+00	False
3	gradient	32	2221397	1.027219e+01	3.804144e+00	1.442518e+01	False
4	gradient	76	1053594	1.028015e+01	4.071310e+00	4.079723e+01	False
5	gradient	180	474121	1.133839e+01	4.567963e+00	1.305351e+02	False
6	gradient	425	206209	1.125048e+01	5.394310e+00	2.701364e+02	False
7	gradient	1000	87107	1.129870e+01	7.278420e+00	6.618817e+02	False
8	steepest descent	2	4416	9.994942e-09	3.212542e-07	1.497200e-15	True
9	steepest descent	5	348464	2.394603e-08	3.647346e-03	4.674375e-11	True
10	steepest descent	13	26	5.331049e-09	6.366238e-08	1.552348e-18	True
11	steepest descent	32	26	5.335938e-09	3.704032e-10	1.610500e-18	True
12	steepest descent	76	26	5.335938e-09	3.704032e-10	1.610500e-18	True
13	steepest descent	180	26	5.335938e-09	3.704032e-10	1.610500e-18	True
14	steepest descent	425	26	5.335938e-09	3.704032e-10	1.610500e-18	True
15	steepest descent	1000	26	5.335938e-09	3.704032e-10	1.610500e-18	True
16	conjugate	2	60	9.572306e-09	2.613158e-07	9.917234e-16	True
17	conjugate	5	301199	NaN	NaN	NaN	False
18	conjugate	13	177360	NaN	NaN	NaN	False
19	conjugate	32	76719	2.700434e-07	2.812501e+00	6.325268e-01	False
20	conjugate	76	40742	2.867093e-07	2.812501e+00	6.325268e-01	False
21	conjugate	180	16751	3.061752e-07	2.812501e+00	6.325268e-01	False
22	conjugate	425	6937	1.701678e-07	2.812501e+00	6.325268e-01	False
23	conjugate	1000	3058	1.419353e-07	2.812501e+00	6.325268e-01	False
24	newton	2	114	8.346094e-09	5.899464e-08	5.166099e-17	True
25	newton	5	1215	9.152944e-09	2.273622e-06	1.954188e-17	True
26	newton	13	1233119	1.544644e+07	3.987430e+11	1.968192e+01	False
27	nesterov	2	10687	3.322111e-09	1.185320e-07	2.037922e-16	True
28	nesterov	5	51637	9.757223e-09	1.008134e-03	3.576222e-12	True
29	nesterov	13	1685293	2.761584e-06	1.965465e+00	9.830312e-01	False
30	nesterov	32	828069	3.243469e-06	1.965465e+00	9.830312e-01	False
31	nesterov	76	381661	2.533069e-06	1.965465e+00	9.830312e-01	False
32	nesterov	180	179946	2.256360e-06	1.965465e+00	9.830312e-01	False
33	nesterov	425	76598	2.114588e-06	1.965465e+00	9.830312e-01	False
34	nesterov	1000	34272	2.056526e-06	1.965465e+00	9.830312e-01	False

Из данной таблицы видим что на различных размерностях лучше всего себя повел метод наиск. спуска. Возможно не последее место здесь сыграла его простота



На изображении выше мы построили график для n=5. Видим, что метод Нестерова практически сразу сходится. Лучше всего себя ведет метод Наиск. спуска. Метод сопр. градиентов, напротив, ведет себя хуже всех и единственный не сходится



изображение выше отображает поведеие уже модуля производных для того же случая. Здесь можно попытаться понять, почему тот или иной метод себя ведет по разному. Нам же интересен метод сопр. градиента. Видим что производная не сходится к нулю. Возможно это связано с тем, что "инерционность"аккумулирует сумму всех предыдущих производных

Безусловная минимизация выпуклой функции

$f(x) = \sum_{i=1}^{n} e^{-x_i} + e^{\sum_{i=1}^{n} x_i - 1}$

Вычислим первые производные:

$\frac{df}{dx_i} = -e^{-x_i} + e^{\sum_{i=1}^{n} x_i - 1}$

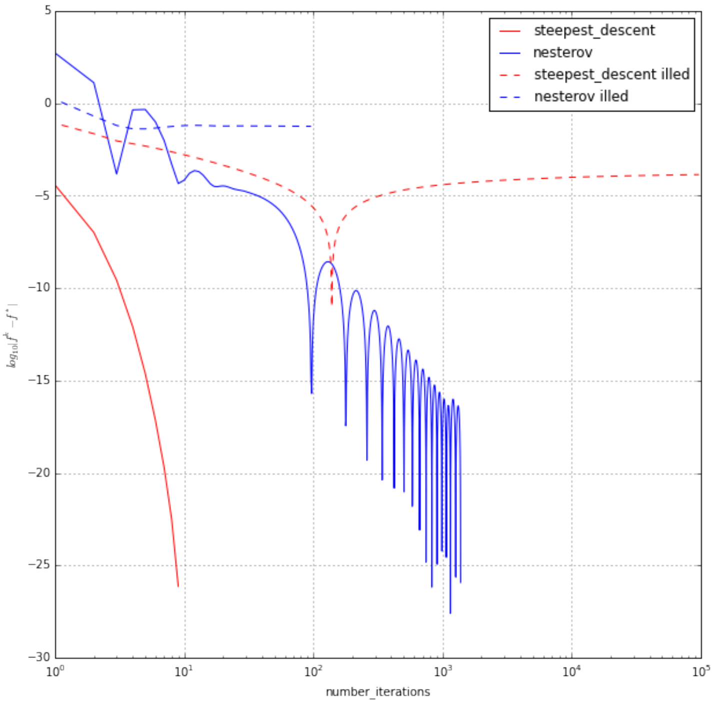
Вычислим вторые производные:

$\frac{d^2 f}{dx_i^2} = e^{-x_i} + e^{\sum_{i=1}^{n} x_i - 1}$

$\frac{d^2 f}{dx_i dx_j}|_{i \neq j} = e^{\sum_{i=1}^{n} x_i - 1}$

	Метод	N	||grad f(x')||	||x* - x'||	|f* - f(x')|	Скорость схо
0	steepest descent(gradient stop)	10	8.213606e-07	0.000006	4.504841e-12	None
1	steepest descent(func stop)	6	1.787821e-04	0.000951	4.478236e-07	None
2	nesterov(gradient stop)	1379	4.829861e-07	0.000004	5.528022e-12	None
3	nesterov(func stop)	97	1.702497e-04	0.000768	2.919644e-07	None
4	steepest descent(gradient stop)	96368	4.007038e-05	19.362832	2.122190e-02	None
5	steepest descent(func stop)	96107	1.271733e-04	19.349824	2.121939e-02	None
6	nesterov(gradient stop)	37170	9.999801e-07	226.772340	2.493577e-02	None
7	nesterov(func stop)	517306	4.036475e-06	1084.723037	2.593116e-02	None

В данной таблице строки 0-3 - для чистой функции, а 4-7 для испорченной. Стоит обратить внимание, что для плохо обусловленных задач (испорченная фунция) сходимость методов в разы уменьшается, что мы собственно здесь и увидели.



На изображении выше представлено сравнение скорости сходимости для функции и ее испорченной версии