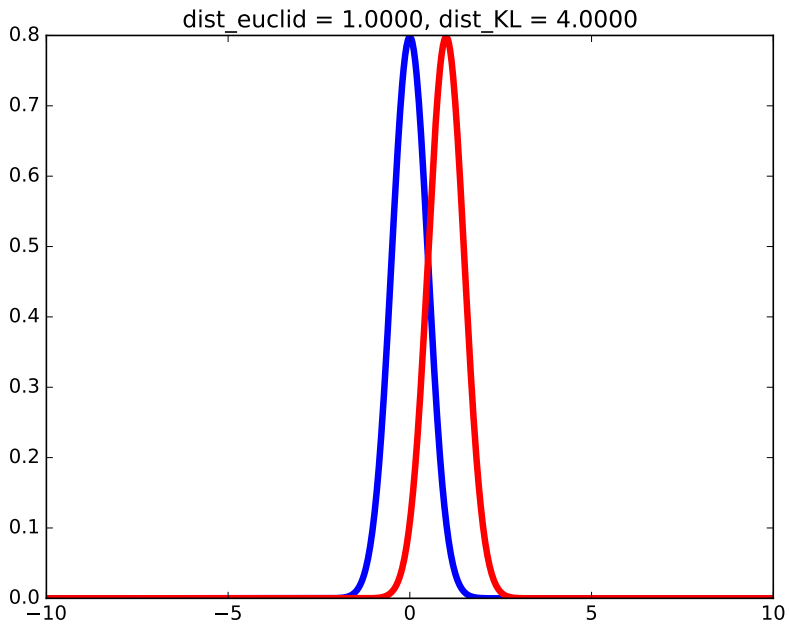
Optimizing Neural Networks using Kronecker-factored Approximate Curvature (K-FAC)

Dmitry Kropotov

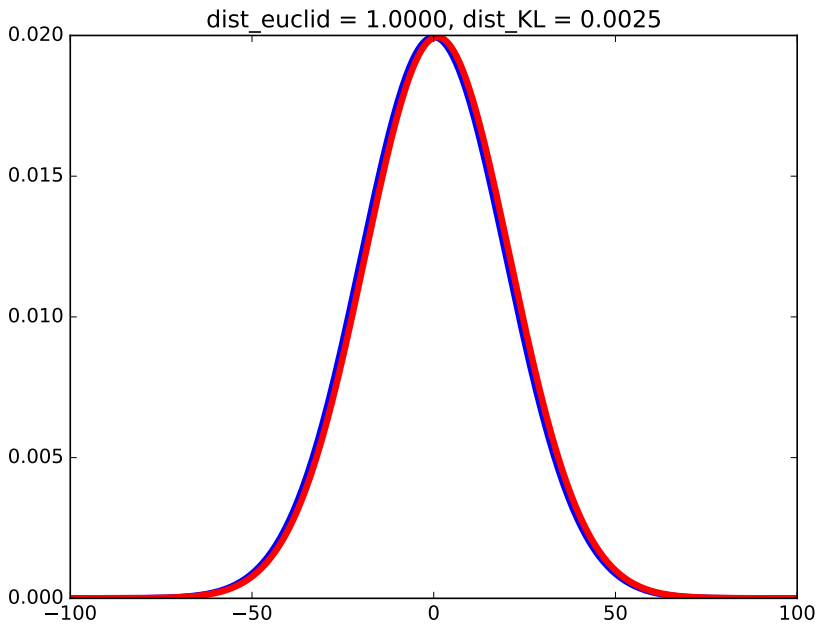
Bayesgroup seminar

7 октября 2016 г.

Distance between distributions

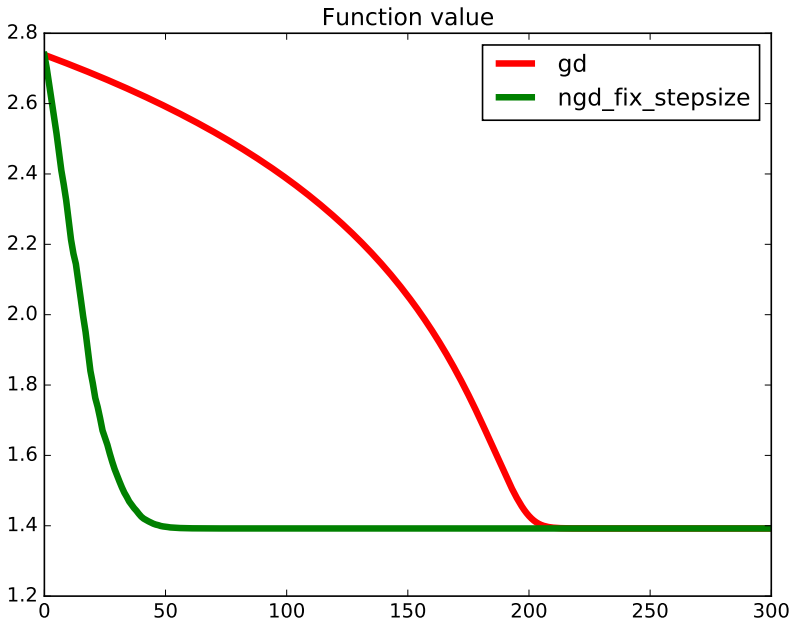


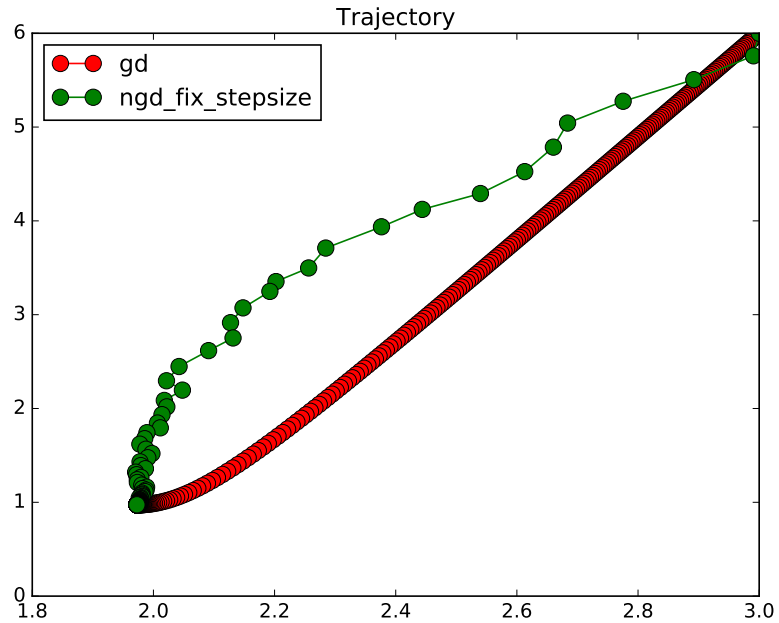
$$\mu_1 = 0, \mu_2 = 1$$
$$\sigma_1 = \sigma_2 = 0.5$$



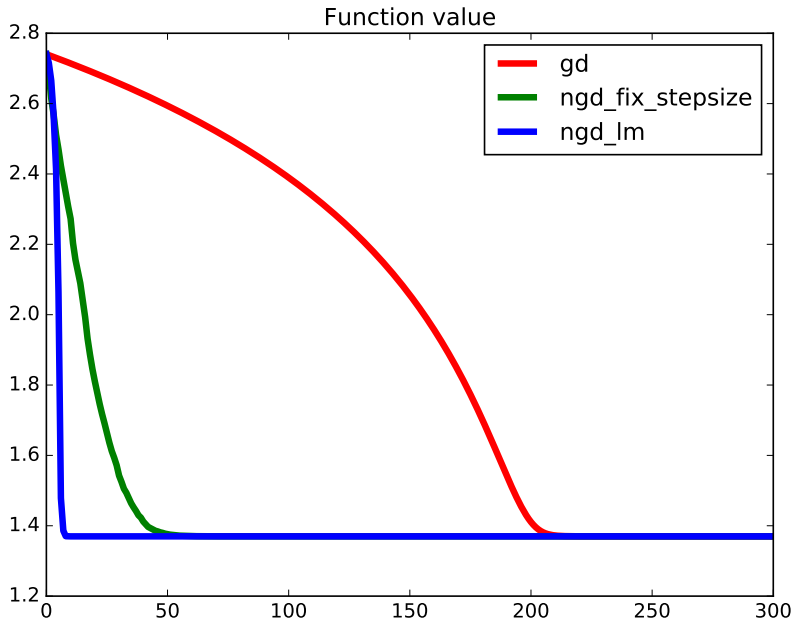
$$\mu_1 = 0, \mu_2 = 1$$
$$\sigma_1 = \sigma_2 = 20$$

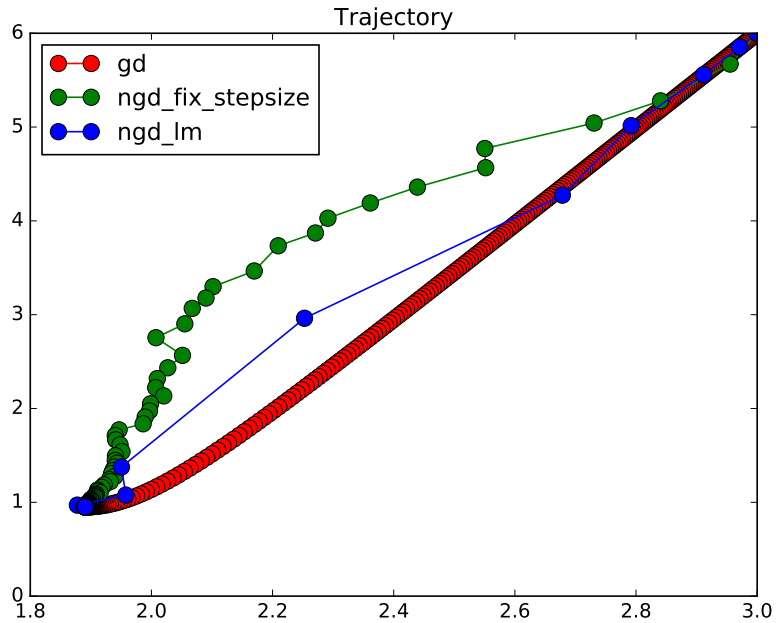
GD vs. NGD


Function value
gd
ngd_fix_stepsize
2.8
2.6
2.4
2.2
2.0
1.8
1.6
1.4
1.2
0
50
100
150
200
250
300

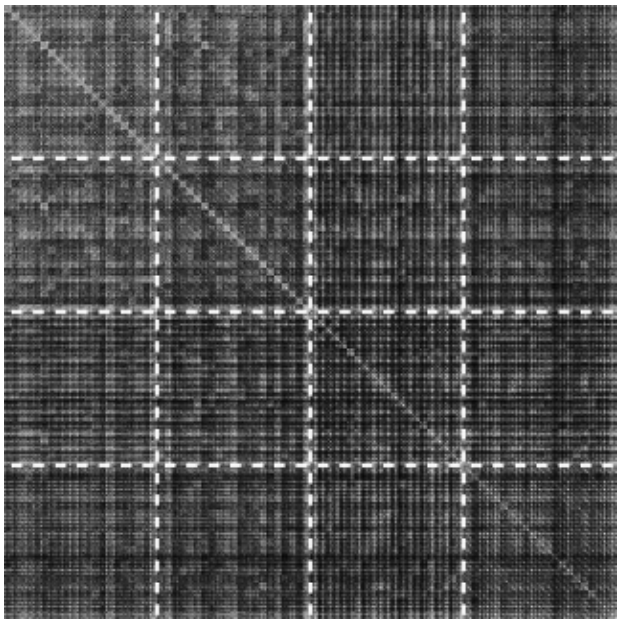

Trajectory
gd
ngd_fix_stepsize
6
5
4
3
2
1
0
1.8
2.0
2.2
2.4
2.6
2.8
3.0

Adding Levenberg-Marquardt for adjusting step size

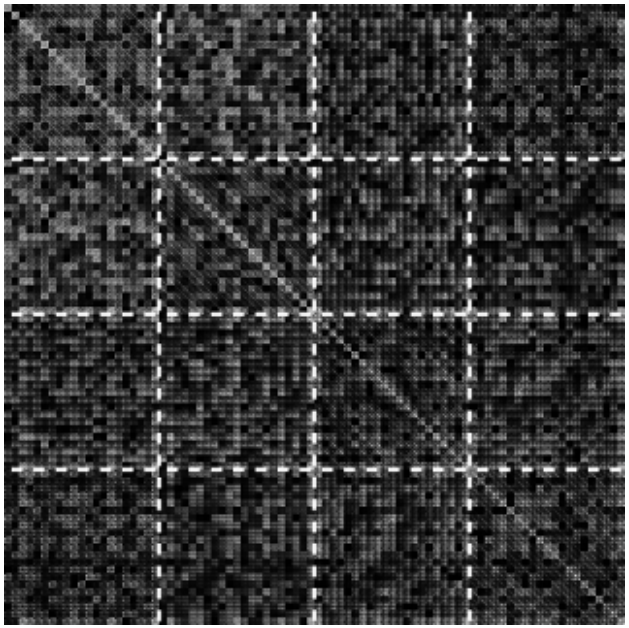




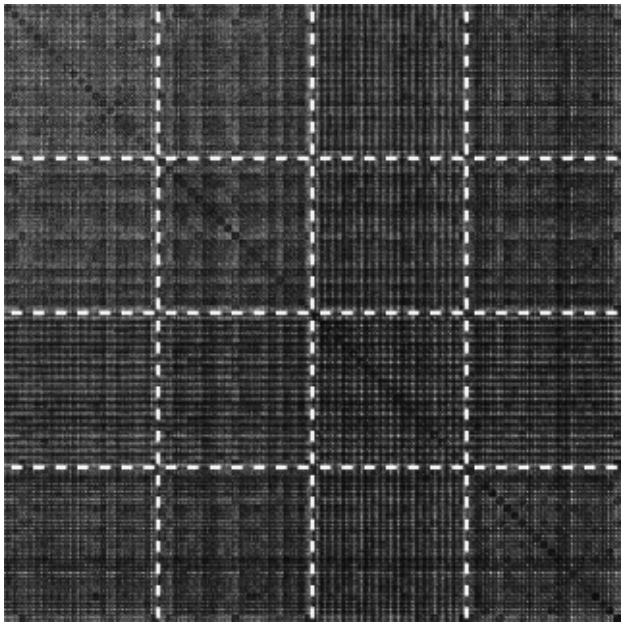
Quality of K-FAC Fisher matrix approximation



True Fisher



K-FAC



difference between the two