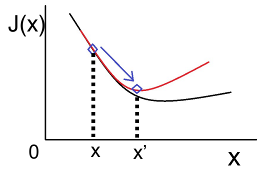
Convergence of Numerical Newton's Method: High-Dimensional Perturbed Cost Functions

Scott Reid

Department of Applied Physics, Stanford University

Numerical Newton's Method

In Newton's method, the cost function is locally modeled as a parabola and in each iteration the algorithm steps to the minimum of that parabola. This requires knowledge of the Hessian matrix and gradient. If we do not have access to the analytic derivatives, we must take our derivatives numerically.



$$\nabla[J(\vec{x})]_i = \frac{J(\vec{x}+\delta\hat{e}_i) - J(\vec{x})}{\delta}$$

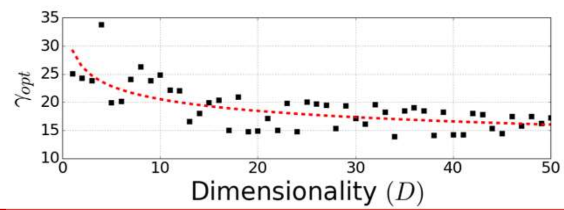
$$H[J(\vec{x})]_{ij} = \frac{J(\vec{x}+\delta\hat{e}_i+\delta\hat{e}_j) - J(\vec{x}+\delta\hat{e}_i) - J(\vec{x}+\delta\hat{e}_j) + J(\vec{x})}{\delta^2}$$

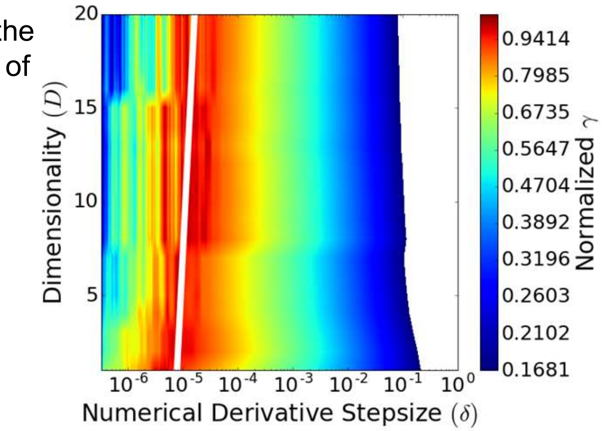
Optimizing Gradient Stepsize δ in D Dimensions

I find that the optimal stepsize δ which maximizes the convergence rate γ depends on the dimensionality of the cost function.

$$J(\vec{x}) \equiv ||\vec{x}||^2 \qquad \gamma \equiv -\log\frac{J(\vec{x}^{(1)})}{J(\vec{x}^{(0)})}$$

$$\delta_{\text{opt}} \approx 7.5\cdot 10^{-6} \times e^{D/59.4} \qquad \gamma_{\text{opt}} \approx 29.3 \cdot D^{-0.155}$$





Guiding Questions

(1) What numerical gradient stepsize δ will optimize the convergence rate?

(2) How is the convergence rate affected as our cost function becomes more complex?

Results

(1) Optimal stepsize depends on D.

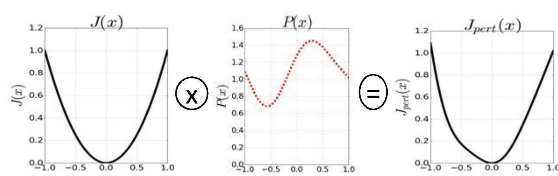
(2) Convergence rate decreases with cost function complexity.

(3) Convergence is less affected by complexity in higher dimensions.

Perturbing the Cost Function

I consider more complicated cost functions by multiplying the unperturbed cost function by a randomly generating perturbing mask function.

$$J_{\text{pert}}(\vec{x}) = P(\vec{x}) \cdot ||\vec{x}||^2$$



I find that the convergence rate is affected by the perturbation density N roughly as:

$$\gamma(D,N) \approx \gamma_o(D)e^{-\Gamma(D)\cdot N} \qquad \Gamma(D) \approx 0.869 \cdot e^{-D/3.46}$$