Beta binomial density:

$$\begin{aligned} BetaBin(k|n,\alpha,\beta) &= \binom{n}{k} \cdot \frac{Be(\alpha+k,\beta+n-k)}{Be(\alpha,\beta)} \\ &= \frac{\Gamma(n+1)}{\Gamma(k+1)\Gamma(n-k+1)} \cdot \frac{\Gamma(\alpha+k)\Gamma(\beta+n-k)}{\Gamma(\alpha+\beta+n)} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \end{aligned}$$

Definition of μ, α, β:

$$\mu = \frac{e^y}{1+e^y} \qquad \alpha = \mu\left(\frac{1-\rho}{\rho}\right) \qquad \beta = (\mu-1)\left(\frac{\rho-1}{\rho}\right) = (\mu-1) - \left(\frac{1-\rho}{\rho}\right)$$

Negative log-likelihood:

$$\begin{aligned} nll = &-\log(\Gamma(n+1)) + log(\Gamma(k+1)) + \log(\Gamma(n-k+1)) \\ &+ \log(\Gamma(\alpha)) + \log(\Gamma(\beta)) - \log(\Gamma(\alpha+k)) - \log(\Gamma(\beta+n-k)) \\ &+ \log(\Gamma(\alpha+\beta+n)) - \log(\Gamma(\alpha+\beta)) \end{aligned}$$

$$nll_{\textbf{truncated}} = \log(\Gamma(\alpha)) + \log(\Gamma(\beta)) - \log(\Gamma(\alpha+k)) - \log(\Gamma(\beta+n-k))$$

NLL with pseudocounts ($k+1$, $n+2$):

$$nll_{\textbf{truncated}} = \log(\Gamma(\alpha)) + \log(\Gamma(\beta)) - \log(\Gamma(\alpha+k+1)) - \log(\Gamma(\beta+n-k+1))$$

NLL for $y \to \infty$:

$$\begin{aligned} \lim_{y\to\infty} \mu &= 1 \\ \lim_{y\to\infty} \alpha &= \lim_{y\to\infty} \mu\left(\frac{1-\rho}{\rho}\right) = \frac{1-\rho}{\rho} \\ \lim_{y\to\infty} \alpha+k+1 &= \frac{1-\rho}{\rho} + k + 1 \\ \lim_{y\to\infty} \mu - 1 &= 0 \\ \lim_{y\to\infty} \beta &= \lim_{y\to\infty} (\mu-1)\left(\frac{\rho-1}{\rho}\right) = 0 \\ \lim_{y\to\infty} \beta+n-k+1 &= n-k+1 \\ \lim_{y\to\infty} \log(\Gamma(\alpha)) &= \log(\Gamma(\frac{1-\rho}{\rho})) \\ \lim_{y\to\infty} \log(\Gamma(\alpha+k+1)) &= \log(\Gamma(\frac{1-\rho}{\rho}+k+1)) \\ \lim_{y\to\infty} \log(\Gamma(\beta)) &= \infty(\log(\Gamma(0))) : \text{not defined}) \\ \lim_{y\to\infty} \log(\Gamma(\beta+n-k+1)) &= \log(\Gamma(n-k+1)) \end{aligned}$$

NLL for $y \to -\infty$:

$$\lim_{y \to -\infty} \mu = 0$$

$$\lim_{y \to -\infty} \alpha = \lim_{y \to -\infty} \mu \left(\frac{1-\rho}{\rho} \right) = 0$$

$$\lim_{y \to -\infty} \alpha + k + 1 = k + 1$$

$$\lim_{y \to -\infty} \mu - 1 = -1$$

$$\lim_{y \to -\infty} \beta = -1 \cdot \frac{\rho - 1}{\rho} = \frac{1-\rho}{\rho}$$

$$\lim_{y \to -\infty} \beta + n - k + 1 = \frac{1-\rho}{\rho} + n - k + 1$$

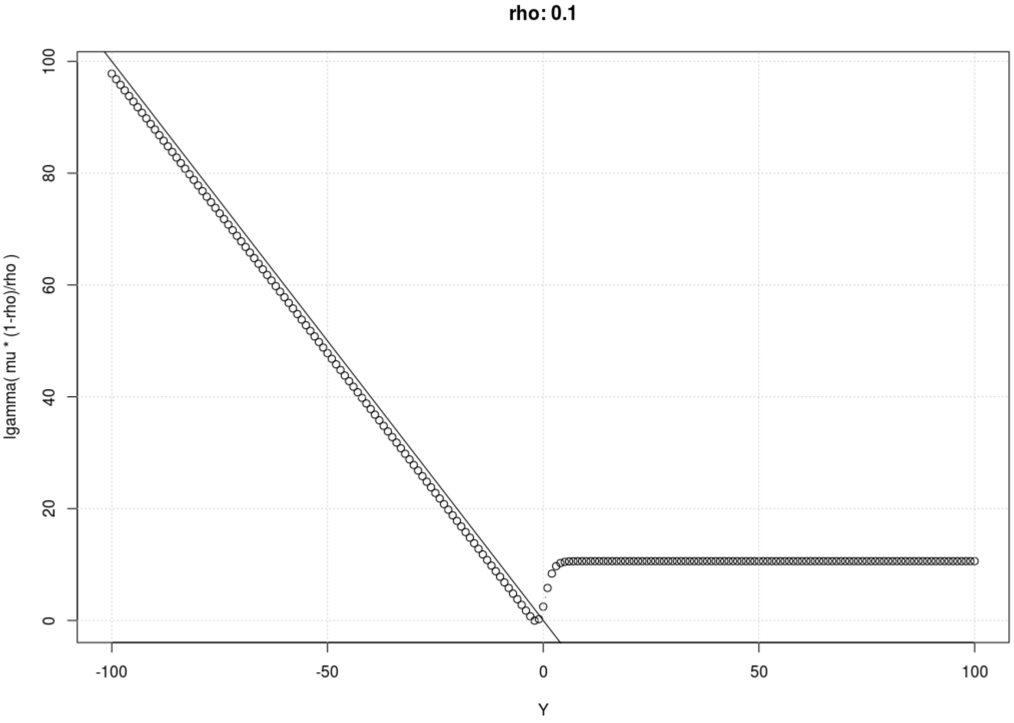
$$\lim_{y \to -\infty} \log(\Gamma(\alpha)) = \infty(\log(\Gamma(0))) : \text{not defined})$$

$$\lim_{y \to -\infty} \log(\Gamma(\alpha + k + 1)) = \log(\Gamma(k+1))$$

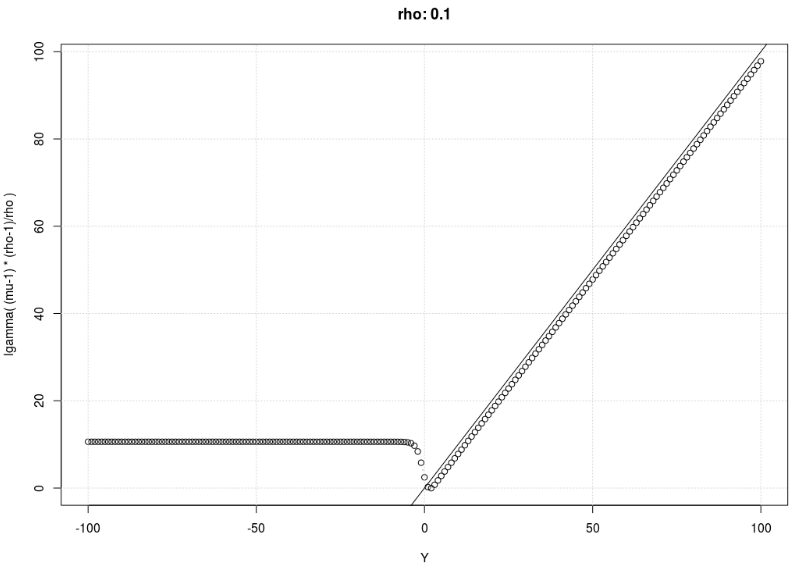
$$\lim_{y \to -\infty} \log(\Gamma(\beta)) = \log(\Gamma(\frac{1-\rho}{\rho}))$$

$$\lim_{y \to -\infty} \log(\Gamma(\beta + n - k + 1)) = \log(\Gamma(\frac{1-\rho}{\rho} + n - k + 1))$$

$\log(\Gamma(\alpha))$:



$\log(\Gamma(\beta))$:



$\Rightarrow$ $\log(\Gamma(\alpha))$ and $\log(\Gamma(\beta))$ approximated as a straight line with slope -1 and 1, for $y \to -\infty$ and $y \to \infty$, respectively.

Gradient of d:

$$\begin{aligned}\frac{d}{dW_d} nll_{\textbf{truncated}} &= \left[\psi\left(\mu\left(\frac{1-\rho}{\rho}\right)\right)\cdot\frac{1-\rho}{\rho}\cdot\frac{e^Y}{(1+e^Y)^2}\right]^T\cdot XW_e \\ &+ \left[\psi\left((\mu-1)\left(\frac{\rho-1}{\rho}\right)\right)\cdot\frac{\rho-1}{\rho}\cdot\frac{e^Y}{(1+e^Y)^2}\right]^T\cdot XW_e \\ &- \left[\psi\left(\mu\left(\frac{1-\rho}{\rho}\right)+k+1\right)\cdot\frac{1-\rho}{\rho}\cdot\frac{e^Y}{(1+e^Y)^2}\right]^T\cdot XW_e \\ &- \left[\psi\left((\mu-1)\left(\frac{\rho-1}{\rho}\right)+n-k+1\right)\cdot\frac{\rho-1}{\rho}\cdot\frac{e^Y}{(1+e^Y)^2}\right]^T\cdot XW_e\end{aligned}$$

Approximation of the gradient:
For $y \to \infty$:

$$\begin{aligned}\frac{e^Y}{(1+e^Y)^2} &\to 0 \\ \psi\left((\mu-1)\left(\frac{\rho-1}{\rho}\right)\right) &\to -\infty \\ \psi\left((\mu-1)\left(\frac{\rho-1}{\rho}\right)\right)\cdot\frac{\rho-1}{\rho}\cdot\frac{e^Y}{(1+e^Y)^2} &\to 1 \\ \psi\left(\mu\left(\frac{1-\rho}{\rho}\right)\right)\cdot\frac{1-\rho}{\rho}\cdot\frac{e^Y}{(1+e^Y)^2} &\to 0 \\ \psi\left(\mu\left(\frac{1-\rho}{\rho}\right)+k+1\right)\cdot\frac{1-\rho}{\rho}\cdot\frac{e^Y}{(1+e^Y)^2} &\to 0 \\ \psi\left((\mu-1)\left(\frac{\rho-1}{\rho}\right)+n-k+1\right)\cdot\frac{\rho-1}{\rho}\cdot\frac{e^Y}{(1+e^Y)^2} &\to 0\end{aligned}$$

For $y \to -\infty$:

$$\frac{e^Y}{\left(1+e^Y\right)^2} \to 0$$

$$\psi\left(\mu\left(\frac{1-\rho}{\rho}\right)\right) \to -\infty$$

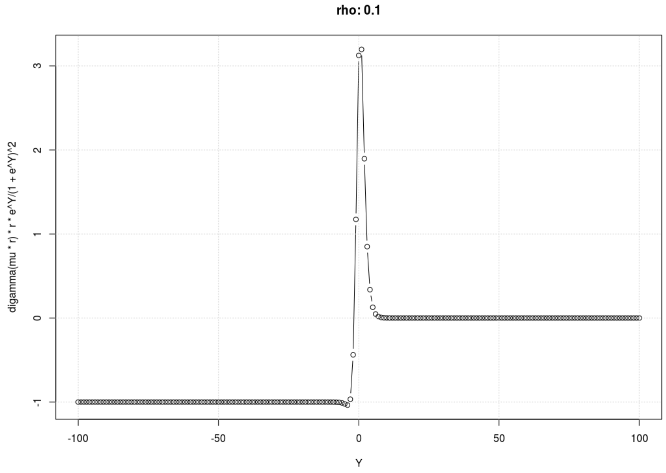
$$\psi\left(\mu\left(\frac{1-\rho}{\rho}\right)\right) \cdot \frac{1-\rho}{\rho} \cdot \frac{e^Y}{\left(1+e^Y\right)^2} \to -1$$

$$\psi\left((\mu-1)\left(\frac{\rho-1}{\rho}\right)\right) \cdot \frac{\rho-1}{\rho} \cdot \frac{e^Y}{\left(1+e^Y\right)^2} \to 0$$

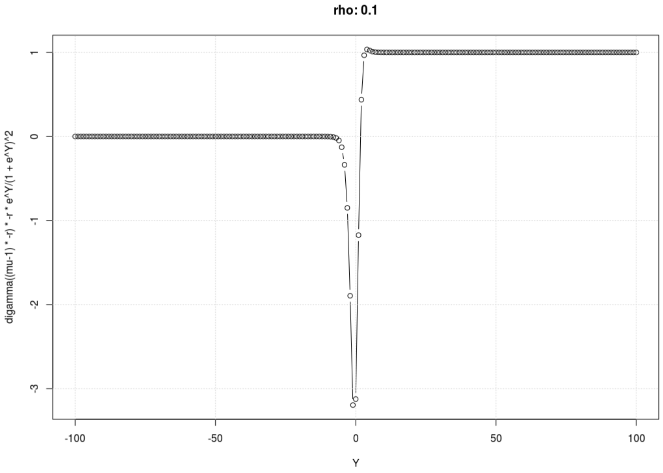
$$\psi\left(\mu\left(\frac{1-\rho}{\rho}\right)+k+1\right) \cdot \frac{1-\rho}{\rho} \cdot \frac{e^Y}{\left(1+e^Y\right)^2} \to 0$$

$$\psi\left((\mu-1)\left(\frac{\rho-1}{\rho}\right)+n-k+1\right) \cdot \frac{\rho-1}{\rho} \cdot \frac{e^Y}{\left(1+e^Y\right)^2} \to 0$$

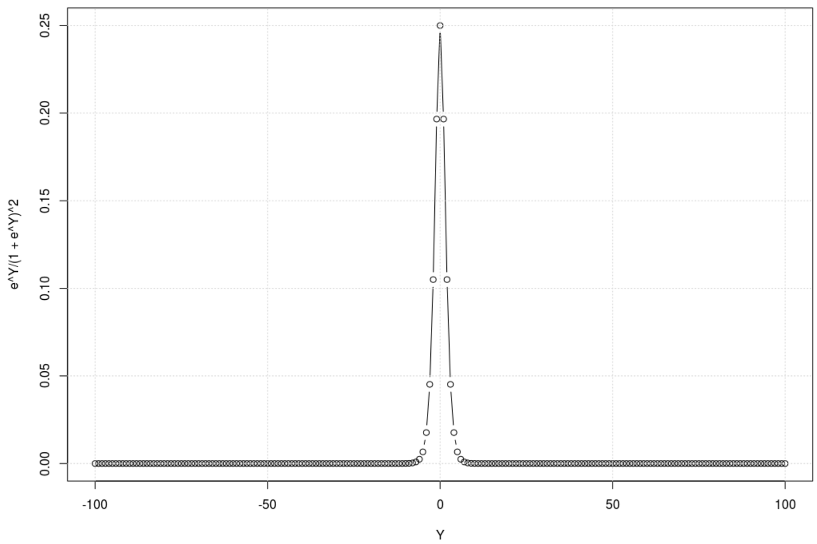
$\psi(\alpha) \cdot \frac{1-\rho}{\rho} \cdot \frac{e^Y}{(1+e^Y)^2}$:



$\psi(\beta) \cdot \frac{\rho-1}{\rho} \cdot \frac{e^Y}{(1+e^Y)^2}$:

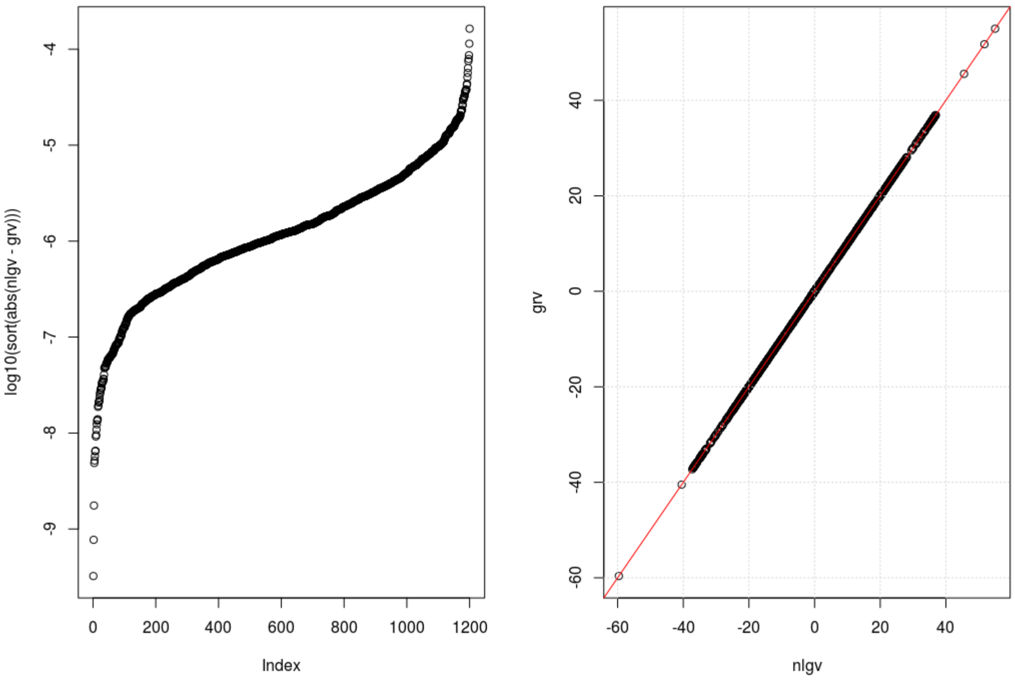


$\frac{e^Y}{(1+e^Y)^2}$:

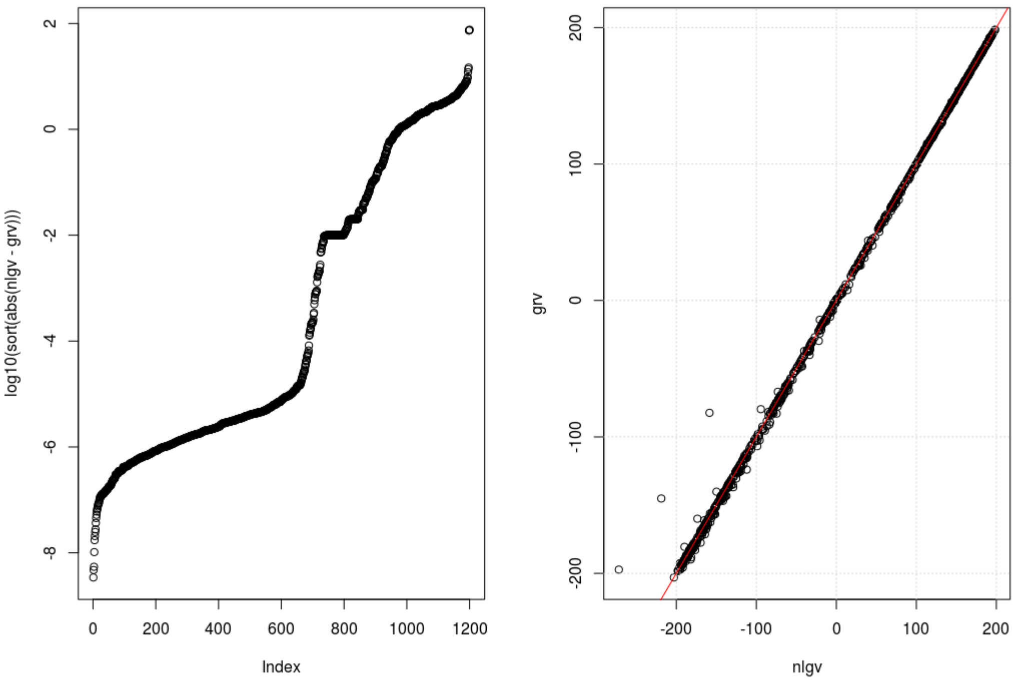


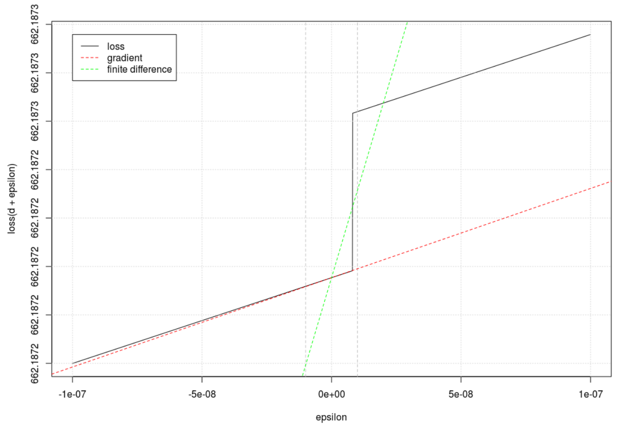
Gradient with approximation compared to finite difference approximation:

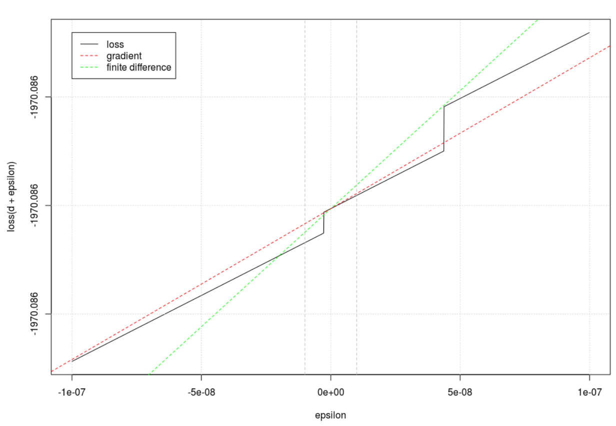
values in D in [-1, 1]



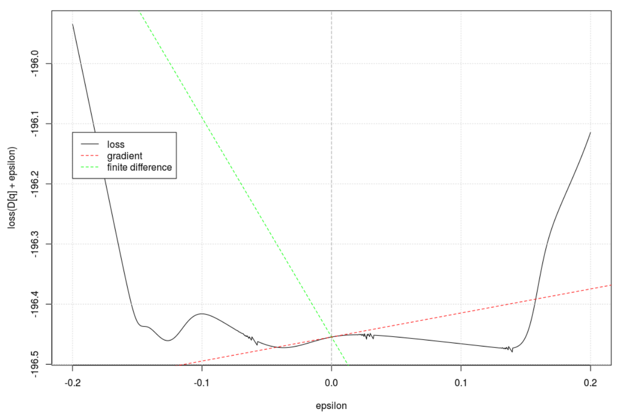
values in D in [-5, 5]: some inaccuracies compared to finite difference approximation because of small jumps in the loss function (probably because of the approximation), but gradient is fine (see plots below, direction of gradient is computed correctly), and the Fraser autoencoder fitting works



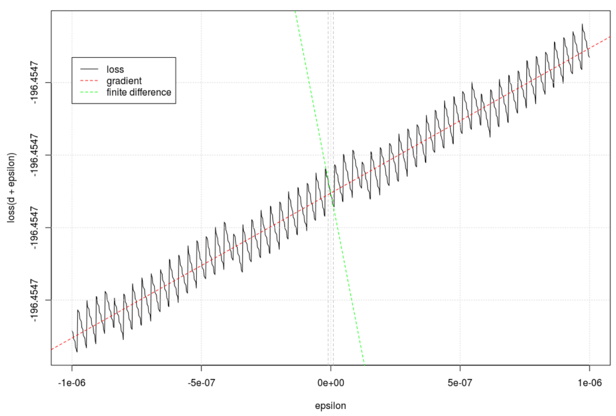




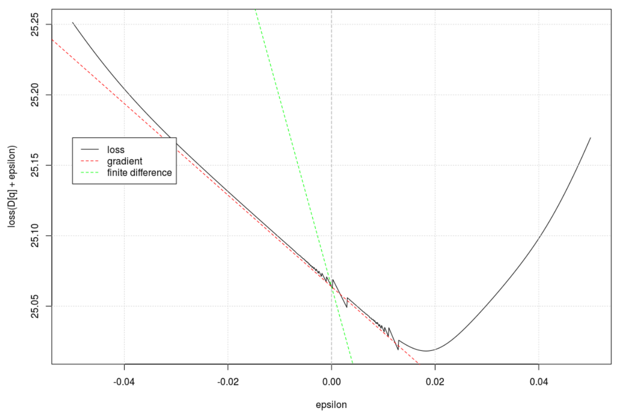
idx: 148 q: 4



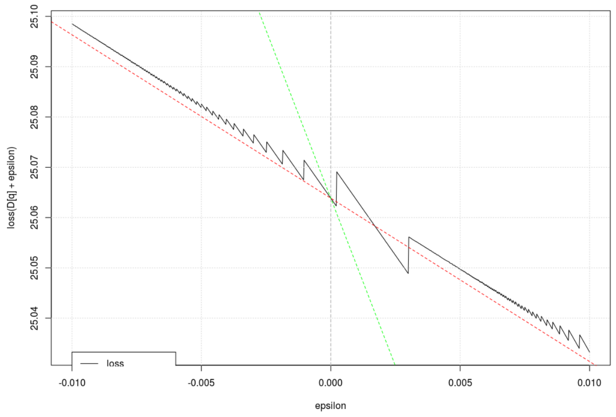
idx: 148 q: 4



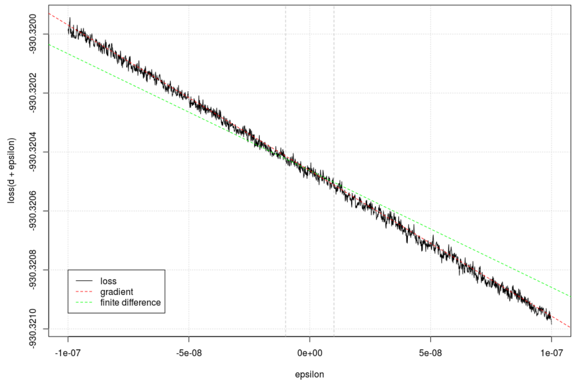
idx: 271 q: 4

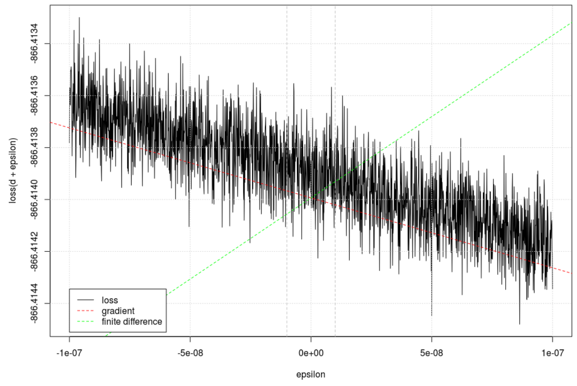


idx: 271 q: 4



For small rho ($\approx \rho < 10^{-8}$), the loss function is sometimes instable (also for small values in D where no approximation happens), but the gradient is correct (problems with finite difference in these cases):


loss
gradient
finite difference
loss(d + epsilon)
-1e-07
-5e-08
0e+00
5e-08
1e-07
epsilon


loss
gradient
finite difference
loss(d + epsilon)
-1e-07
-5e-08
0e+00
5e-08
1e-07
epsilon