$$\gamma = \frac{\|\nabla M_t - \nabla M_{t-1}\|^2}{\|\nabla M_{t-1}\|^2}$$

$$CG_t = \nabla M_t + \gamma CG_{t-1}$$

$$p_{\text{new}} = p_{\text{old}} + \text{learning rate } CG$$

regularization happens here ...