



A contour plot illustrating the optimization of a loss function $L_{\mathcal{B}}$. The plot features concentric elliptical contours, with the center representing the minimum loss (white) and the outer edges representing higher loss values (red, yellow, green, blue). A black dot marks the current parameter value $w(t)$, and a small grey circle indicates the next step $w(t+1)$ after a gradient descent update. The update is shown as a small step towards the center, following the negative gradient direction.

$$w(t+1) = w(t) - 0.03 \nabla L_{\mathcal{B}}(w(t))$$