

# GRADIENT CLIPPING

Cliffs in the loss function can cause the learning process to overshoot a desirable minimum. This happens because the gradient at the cliff is so high. One solution is to prevent the gradient from taking extreme values:

$$\text{if } \|g\| > v:$$
$$g \leftarrow \frac{g v}{\|g\|}$$

Annotations:

- $g$ : gradient
- $v$ : Threshold
- $\|g\|$ : norm of gradient