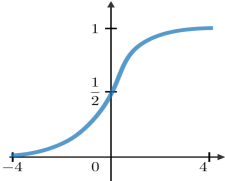
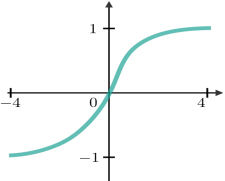
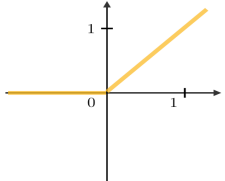


□ **Backpropagation temporelle** – L'étape de backpropagation est appliquée dans la dimension temporelle. À l'instant  $T$ , la dérivée du loss  $\mathcal{L}$  par rapport à la matrice de coefficients  $W$  est donnée par :

$$\frac{\partial \mathcal{L}^{(T)}}{\partial W} = \sum_{t=1}^T \left. \frac{\partial \mathcal{L}^{(T)}}{\partial W} \right|_{(t)}$$

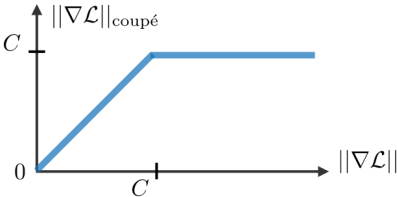
2.2 Dépendances à long terme

□ **Fonctions d'activation communément utilisées** – Les fonctions d'activation les plus utilisées dans les RNNs sont décrits ci-dessous :

Sigmoïde	Tanh	RELU
$g(z) = \frac{1}{1 + e^{-z}}$	$g(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$	$g(z) = \max(0, z)$
		

□ **Gradient qui disparaît/explose** – Les phénomènes de gradient qui disparaît et qui explose (en anglais *vanishing gradient* et *exploding gradient*) sont souvent rencontrés dans le contexte des RNNs. Ceci est dû au fait qu'il est difficile de capturer des dépendances à long terme à cause du gradient multiplicatif qui peut décroître/croître de manière exponentielle en fonction du nombre de couches.

□ **Coupage de gradient** – Cette technique est utilisée pour atténuer le phénomène de gradient qui explose qui peut être rencontré lors de l'étape de backpropagation. En plafonnant la valeur qui peut être prise par le gradient, ce phénomène est maîtrisé en pratique.



□ **Types de porte** – Pour remédier au problème du gradient qui disparaît, certains types de porte sont spécifiquement utilisés dans des variantes de RNNs et ont un but bien défini. Les portes sont souvent notées  $\Gamma$  et sont telles que :

$$\Gamma = \sigma(Wx^{<t>} + Ua^{<t-1>} + b)$$