Image Processing 1, Exercise 9

1 Neural network formalism

[basic] This exercise will investigate the mathematical formalism of convolutional neural networks.

Let $f \in \ell_2(\mathbb{Z})$ and consider the 3-channel convolutional layer with input f:

$$\begin{bmatrix} \sigma\big((h_1 * f)[k]\big) \\ \sigma\big((h_2 * f)[k]\big) \\ \sigma\big((h_3 * f)[k]\big) \end{bmatrix},$$

where h_1, h_2, h_3 are filters of the form

$$h_1 = \begin{bmatrix} 0 & \boxed{1} & 0 \end{bmatrix} \qquad h_2 = \begin{bmatrix} 1 & \boxed{-4} & 1 \end{bmatrix} \qquad h_3 = \begin{bmatrix} 0 & \boxed{1} & 0 \end{bmatrix}$$

and $\sigma : \mathbb{R} \to \mathbb{R}$ is a fixed pointwise nonlinearity. The 3-channel convolutional layer can be implemented as the neural network architecture

$$\boldsymbol{\sigma}(\mathbf{W}\boldsymbol{f}_W[k]) = \begin{bmatrix} \sigma\big((h_1 * f)[k]\big) \\ \sigma\big((h_2 * f)[k]\big) \\ \sigma\big((h_3 * f)[k]\big) \end{bmatrix},$$

where we recall that $\boldsymbol{\sigma}$ applies σ component-wise and

$$f_W[k] = (f[k - k_0])_{k_0 \in W}$$

denotes the patch extraction operator. Let $W = \{-2, -1, 0, 1, 2\}$, so that

$$f_W[k] = \begin{bmatrix} f[k+2] \\ f[k+1] \\ f[k] \\ f[k-1] \\ f[k-2] \end{bmatrix} \in \mathbb{R}^5.$$

Specify the size and the entries of the matrix $\mathbf{W}$, with justification.

Solution: The matrix will be 3×5 and will take the form

$$\mathbf{W} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & -4 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

2 Stability of convolution layers

[basic] To investigate the stability of convolution layers, you need to have a solid understanding of discrete convolution operators.

The convolution masks used in CNNs are usually of finite length, which implies that they are absolutely summable. Here, we show that the latter property is sufficient for the convolution layers to be stable; that is, Lipschitz continuous.

(a) Prove that the convolution operator $\mathrm{T}_{\mathrm{LSI}} : f \mapsto h * f$ with $h \in \ell_1(\mathbb{Z}^d)$ is Lipschitz continuous. Hint: Derive a simple bound for the frequency response of h.

Solution: First, we recall that the condition $h \in \ell_1(\mathbb{Z}^d)$ is equivalent to

$$\|h\|_{\ell_1} = \sum_{\boldsymbol{k}\in\mathbb{Z}^d} \big|h[\boldsymbol{k}]\big| < +\infty.$$

Consequently,

$$\forall \boldsymbol{\omega} \in \mathbb{R}^d: \quad \big|H(\mathrm{e}^{\mathrm{j}\boldsymbol{\omega}})\big| = \Big|\sum_{\boldsymbol{k}\in\mathbb{Z}^d} h[\boldsymbol{k}]\mathrm{e}^{-\mathrm{j}\langle\boldsymbol{\omega},\boldsymbol{k}\rangle}\Big| \le \sum_{\boldsymbol{k}\in\mathbb{Z}^d} |h[\boldsymbol{k}]| \underbrace{\big|\mathrm{e}^{-\mathrm{j}\langle\boldsymbol{\omega},\boldsymbol{k}\rangle}\big|}_{=1} = \|h\|_{\ell_1},$$

which implies that $\mathrm{Lip}(\mathrm{L}_{\mathrm{LSI}}) \le \|h\|_{\ell_1} < +\infty$.

(b) Lipschitz constant of spatial averagers: Consider the linear averaging operator $\mathrm{T}_{\mathrm{ave}}: f \mapsto g * f$ with an impulse response g such that $g[\boldsymbol{k}] \ge 0$ for all $\boldsymbol{k}$ and $\sum_{\boldsymbol{k}\in\mathbb{Z}^d} g[\boldsymbol{k}] = 1$ (lowpass condition). Show that $\mathrm{Lip}(\mathrm{T}_{\mathrm{ave}}) = 1$, which tells us that the insertion of such an operator cannot degrade the overall stability of a neural network.

Solution: Since $g[\cdot]$ is non-negative, we have that $\|g\|_{\ell_1} = \sum_{\boldsymbol{k}\in\mathbb{Z}^d} |g[\boldsymbol{k}]| = \sum_{\boldsymbol{k}\in\mathbb{Z}^d} g[\boldsymbol{k}] = 1$, which, in view, of the first question, implies that $\mathrm{Lip}(\mathrm{T}_{\mathrm{ave}}) = G_{\max} \le 1$. To show that this estimate is sharp, we now simply observe that $\big|G(\mathrm{e}^{\mathrm{j}\boldsymbol{\omega}})\big|$ achieves its maximum $G_{\max} = \mathrm{Lip}(\mathrm{T}_{\mathrm{ave}}) = 1$ at the origin. Indeed,

$$G(\mathrm{e}^{\mathrm{j}\boldsymbol{0}}) = \sum_{\boldsymbol{k}\in\mathbb{Z}^d} g[\boldsymbol{k}]\mathrm{e}^{-\mathrm{j}\langle\boldsymbol{\omega},\boldsymbol{k}\rangle}\Big|_{\boldsymbol{\omega}=\boldsymbol{0}} = \sum_{\boldsymbol{k}\in\mathbb{Z}^d} g[\boldsymbol{k}] = 1. \tag{1}$$

(c) Composition of filters: Prove that $h_1, h_2 \in \ell_1(\mathbb{Z}^d) \Rightarrow h_2 * h_1 \in \ell_1(\mathbb{Z}^d)$. What happens when h_1 and h_2 are both averagers as in Question (b)?

Solution: To prove the claim, we estimate the ℓ_1-norm of $h_2 * h_1$ as

$$\begin{aligned}
\|h_2 * h_1\|_{\ell_1} &= \sum_{\boldsymbol{k}\in\mathbb{Z}^d} \Big|\sum_{\boldsymbol{n}\in\mathbb{Z}^d} h_2[\boldsymbol{n}]h_1[\boldsymbol{k}-\boldsymbol{n}]\Big| \le \sum_{\boldsymbol{k}\in\mathbb{Z}^d}\sum_{\boldsymbol{n}\in\mathbb{Z}^d} \big|h_2[\boldsymbol{n}]\big|\,\big|h_1[\boldsymbol{k}-\boldsymbol{n}]\big| \\
&\le \sum_{\boldsymbol{m}\in\mathbb{Z}^d}\sum_{\boldsymbol{n}\in\mathbb{Z}^d} \big|h_2[\boldsymbol{n}]\big|\,\big|h_1[\boldsymbol{m}]\big| \qquad \text{(by change of variable)} \\
&\le \Big(\sum_{\boldsymbol{n}\in\mathbb{Z}^d} \big|h_2[\boldsymbol{n}]\big|\Big)\Big(\sum_{\boldsymbol{m}\in\mathbb{Z}^d} \big|h_1[\boldsymbol{m}]\big|\Big) = \|h_2\|_{\ell_1}\|h_1\|_{\ell_1},
\end{aligned}$$

where the interchange of sums is legitimate since the arguments are all positive.

When h_1 and h_2 are both averagers—that is, their impulse responses are non-negative and sum up to one—then, the same holds true for $h_2 * h_1$, with the above bound being sharp. Specifically, we have that $\sup_{\boldsymbol{\omega}\in\mathbb{R}^d} \big|H_2(\mathrm{e}^{\mathrm{j}\boldsymbol{\omega}})H_1(\mathrm{e}^{\mathrm{j}\boldsymbol{\omega}})\big| = H_2(\mathrm{e}^{\mathrm{j}\boldsymbol{0}})H_1(\mathrm{e}^{\mathrm{j}\boldsymbol{0}}) = 1$.

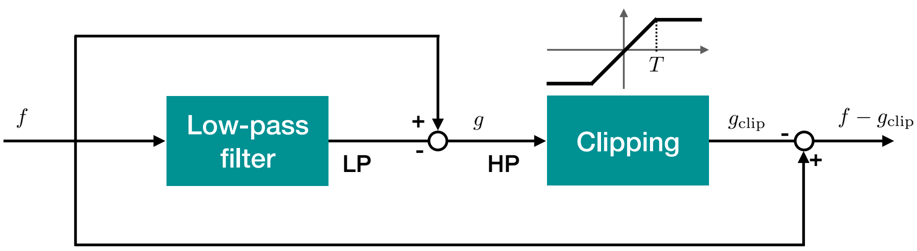

Figure 1: Block diagram of a simple residual denoiser.

3 Analysis of residual denoiser

[intermediate] This exercise will train you to derive stability bounds for simple neuronal architectures that combine convolution operators and pointwise nonlinearities.

The task here is to perform a stability analysis of the residual denoiser in Fig. 1 with $\mathrm{T_{HP}} : f \mapsto f - h * f$ where h is Gaussian-like smoother. Specifically, h is a lowpass filter whose frequency response is such that $0 \le H(\mathrm{e}^{\mathrm{j}\boldsymbol{\omega}}) \le H(\mathrm{e}^{\mathrm{j}\mathbf{0}}) = 1$. The pointwise clipping operator is defined as

$$\sigma_T(x) = \begin{cases} T, & x \ge T \\ x, & x \in [-T, T] \\ -T, & x \le -T \end{cases}$$

(a) Prove that $\mathrm{Lip}(\mathrm{T_{HP}}) \le 1$.

Solution: $\mathrm{T_{HP}}$ is a highpass filter whose frequency response is $W(\mathrm{e}^{\mathrm{j}\boldsymbol{\omega}}) = 1 - H(\mathrm{e}^{\mathrm{j}\boldsymbol{\omega}})$. The condition $0 \le H(\mathrm{e}^{\mathrm{j}\boldsymbol{\omega}}) \le 1$ implies that $0 \le W(\mathrm{e}^{\mathrm{j}\boldsymbol{\omega}}) \le 1$, so that $W_{\max} \le 1$, which is the desired result. We also note that this estimate is sharp if and only if $W(\mathrm{e}^{\mathrm{j}\boldsymbol{\omega}})$ has some frequency nulls.

(b) Compute $\mathrm{Lip}(\sigma_T)$ with $\sigma_T : \mathbb{R} \to \mathbb{R}$.

Solution: The derivative of the clipping pointwise nonlinearity is simply

$$\sigma_T'(x) = \begin{cases} 0, & x \ge T \\ 1, & x \in [-T, T] \\ 0, & x \le -T, \end{cases}$$

which yields $\mathrm{Lip}(\sigma_T) = \sup_{x \in \mathbb{R}} \left|\sigma_T'(x)\right| = 1$.

(c) Global stability of pointwise operators. Let $\sigma : \mathbb{R} \to \mathbb{R}$ with $\mathrm{Lip}(\sigma) = L$. We then define the image-wide version of the operator $\boldsymbol{\sigma}$ with $\boldsymbol{\sigma}\{f\}[\boldsymbol{k}] = \sigma(f[\boldsymbol{k}])$, for all $\boldsymbol{k} \in \mathbb{Z}^d$. Prove that $\boldsymbol{\sigma} : \ell_2(\mathbb{Z}^d) \to \ell_2(\mathbb{Z}^d)$ with $\mathrm{Lip}(\boldsymbol{\sigma}) = L$. The latter is a basic result that is listed in the course and that can be invoked in exercises for deriving basic stability estimates.

Solution: To characterize the stability of $\boldsymbol{\sigma}$, we construct the estimate

$$\|\boldsymbol{\sigma}\{f\} - \boldsymbol{\sigma}\{g\}\|_{\ell_2}^2 = \sum_{\boldsymbol{k} \in \mathbb{Z}^d} \left|\sigma(f[\boldsymbol{k}]) - \sigma(g[\boldsymbol{k}])\right|^2 \le \sum_{\boldsymbol{k} \in \mathbb{Z}^d} \left|L|f[\boldsymbol{k}]) - g[\boldsymbol{k}]|\right|^2 \le L^2 \|f - g\|_{\ell_2}^2. \quad (2)$$

In particular, if we take $f \in \ell_2(\mathbb{Z}^d)$ with $g = 0$, then we get that $\|\boldsymbol{\sigma}\{f\}\|_{\ell_2} \le L\|f\|_{\ell_2} < \infty$, which implies that $\boldsymbol{\sigma}\{f\} \in \ell_2(\mathbb{Z}^d)$. More generally, for any $f, g \in \ell_2(\mathbb{Z}^d)$, we have that $\|\boldsymbol{\sigma}\{f\} - \boldsymbol{\sigma}\{g\}\|_{\ell_2} \le L\|f - g\|_{\ell_2}$, which proves that $\mathrm{Lip}(\boldsymbol{\sigma}) \le L$.
To prove that the bound is sharp, we recall that the hypothesis $\mathrm{Lip}(\sigma) = L$ implies the existence of $x_1, x_2 \in \mathbb{R}$ such that $\left|\sigma(x_2) - \sigma(x_1)\right| = L|x_2 - x_1|$. This suggests to consider the pair of impulsive signals $x_1\delta[\cdot], x_2\delta[\cdot] \in \ell_2(\mathbb{Z}^d)$. Indeed, $\|\boldsymbol{\sigma}\{x_1\delta[\cdot]\} - \boldsymbol{\sigma}\{x_2\delta[\cdot]\}\|_{\ell_2} = \left|\sigma(x_2) - \sigma(x_1)\right| = L|x_2 - x_1| = L\|x_1\delta[\cdot] - x_2\delta[\cdot]\|_{\ell_2}$, which shows that the bound is achievable.

(d) Use the previous results to obtain a simple bound of the Lipschitz constant of the whole system.

Solution: The composition of the highpass filter and the clipping operation results in a transformation with $L_1 = \mathrm{Lip}(\boldsymbol{\sigma}_T \circ \mathrm{T_{HP}}) \le 1 \times 1 = 1$ (upper branch of the block diagram). The output is then formed by taking a linear combination with the identity, which yields the upper Lipschitz bound: $L_1 + \mathrm{Lip}(-\mathrm{Identity}) = 2$.