B Theorem 2.2: Near-optimality of optimal circular input for full convolution

Theorem B.1 *Let $p_v(x)$ denote the activation of a single pooling unit in a valid convolution, square-pooling architecture in response to an input x, and let x_v^{opt} and x_c^{opt} denote the optimal norm-one inputs for valid and circular convolution, respectively. Then if x_c^{opt} is composed of a single sinusoid,*

$$\lim_{n\to\infty} \left| p_v\left(x_v^{opt}\right) - p_v\left(x_c^{opt}\right)\right| = 0.$$

Proof We proceed by first establishing that the maximal eigenvalues of $V_p^* V_p$ limit to those of $C_p^* C_p$. Then we show that the optimal input for circular convolution asymptotically attains the same value when applied to valid convolution. We begin with some definitions.

The strong norm of a square matrix A is $||A|| = \max_{x\neq 0} \frac{||Ax||}{||x||} = \sqrt{\max_k \lambda_k}$, where λ_k are the eigenvalues of the Hermitian positive semidefinite matrix $A^* A$.

The weak norm of a matrix $A \in \mathbb{R}^{p\times p}$ is $|A| = \left(\frac{1}{p}\sum_{i=1}^{p}\sum_{j=1}^{p}|A_{ij}|^2\right)^{\frac{1}{2}}$.

Two sequences of $n \times n$ matrices $\{A_p\}$ and $\{B_p\}$ are *asymptotically equivalent* if

1. A_p and B_p are uniformly bounded in the strong norm

$$||A_p||, ||B_p|| \leq M < \infty, p = 1, 2, \ldots$$

2. and $A_p - B_p = D_p$ goes to zero in weak norm as $p \to \infty$,

$$\lim_{p\to\infty} |D_p| = 0.$$

Lemma B.2 *Let V_p and C_p denote matrices performing valid and circular convolution of a filter $f \in \mathbb{R}^{k\times k}$ with an input of size p, respectively. The sequences of matrices $\left\{V_p^* V_p\right\}$ and $\left\{C_p^* C_p\right\}$ are asymptotically equivalent.*

Proof Let $D_p = V_p^* V_p - C_p^* C_p$. First we will show that $\lim_{p\to\infty} |D_p| = 0$. We do this by showing that the number of nonzero elements in D_p is proportional only to n, not n^2. Note that both circular and valid convolution compute the same $n-k+1 \times n-k+1$ filter responses in the interior of the input. Hence nonzero entries in D_p can come only from the $n^2 - (n-k+1)^2 = 2(k-1)n - (k-1)^2$ filter responses that circular convolution computes but valid convolution does not. Each of these filter responses, when squared, will contribute at most $Q = 2\binom{k^2}{2} + 2k^2$ terms to D_p, where the factor of 2 is due to the symmetry of the quadratic form. This is a significant overestimate, but importantly is only a function of k and not p. Further, we note that for $n > 2k$, the maximum element of D_p is independent of p, that is, $\max_{i,j} |d_{ij}| = M$. Therefore

$$\begin{aligned} |D_p| &= \sqrt{\frac{1}{n^2}\sum_{i=1}^{n^2}\sum_{j=1}^{n^2}|d_{ij}|^2} && (1)\\ &\leq \left(\frac{1}{n^2}\left(2(k-1)n - (k-1)^2\right)QM^2\right)^{\frac{1}{2}} && (2)\\ &\leq Kn^{-\frac{1}{2}} && (3)\end{aligned}$$

where K is not a function of n. Hence $\lim_{p\to\infty} |D_p| = 0$. □

Next we show that the matrices are uniformly bounded in the strong norm. For Hermitian matrices, $||A||^2 = \max_k |\alpha_k|$, the maximum magnitude eigenvalue of A. For the circular convolution case this is simply the square of the magnitude of the maximal Fourier coefficient of f, and hence is bounded for all p. For valid convolution, we note that $||V_p x||^2 = \sum_i^{n-k+1}\left(v_i^T x\right)^2$, where v_i^T is the i^{th} row of V_p. The vector v_i^T contains the filter coefficients f and is otherwise zero; hence it has only k^2 nonzero entries. We can therefore form the vector $\tilde{x}_i \in \mathbb{R}^{k^2}$ from just those elements of x which will be involved in computing the dot product, such that $v_i^T x = f^T \tilde{x}_i$. Then we have

$$\sum_i^{n-k+1} \left(v_i^T x\right)^2 = \sum_i^{n-k+1} \left(f^T \tilde{x}_i\right)^2, \tag{4}$$

$$\leq ||f||^2 \sum_i^{n-k+1} ||\tilde{x}_i||^2, \tag{5}$$

$$\leq k^2 ||f||^2 ||x||^2, \tag{6}$$

where the last inequality comes from the fact that each x_i can appear at most k^2 times in the sum. The strong norm is therefore bounded, since $||V_p^* V_p|| = \max_{x \neq 0} \frac{x^* V_p^* V_p x}{x^* x} \leq k^2 ||f||^2$. $\square$

Next we appeal to the following theorem, which is a variation on that stated by [1].

Theorem B.3 *Let $\alpha_{p,k}$ and $\beta_{p,k}$ denote the eigenvalues of V_p and C_p respectively. Let $\hat{f}(\omega_1, \omega_2)$ denote the 2D discrete time Fourier transform of the filter f,*

$$\hat{f}(\omega_1, \omega_2) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} f[j,k] e^{ij\omega_1} e^{ik\omega_2}.$$

Then

$$\lim_{p \to \infty} \max_k \alpha_{p,k} = \lim_{p \to \infty} \max_k \beta_{p,k} = M_{|\hat{f}|^2}$$

where $M_{\hat{f}}$ is the essential supremum of $|\hat{f}|^2$, that is, the smallest number for which $|\hat{f}(x,y)|^2 \leq M_{|\hat{f}|^2}$ except on a set of total length or measure 0.

Proof The proof is a straightforward generalization of that given in Theorem 4.2, Corollary 4.1, and Corollary 4.2 of [1]. $\square$

Hence we have established that the optimal pooling unit activity for valid and circular convolution converges as p grows. Next we show that the optimal norm-one solution for circular convolution, x_c, is near-optimal for valid convolution provided that x_c consists of a single sinusoid. The difference between objective values is

$$\left| \frac{x_c^* V_p^* V_p x_c}{x_c^* x_c} - \frac{x_c^* C_p^* C_p x_c}{x_c^* x_c} \right| = \left| \frac{x_c^* D_p x_c}{x_c^* x_c} \right|$$

Recall that the number of nonzero elements in D_p can be written as Kn where K is not a function of n. Now we establish a bound on each individual element of x_c; because x_c is a sinusoid that spans the entire input, and the total norm is constrained, the individual elements diminish in size as p grows. In particular,

$$|x_c[j,l]| = \left| \frac{1}{n} \sum_{m=0}^{n-1} \sum_{q=0}^{n-1} z[m,q] e^{i2\pi(\frac{jm}{n} + \frac{lq}{n})} \right| \tag{7}$$

$$\leq \frac{1}{n} \sum_{m=0}^{n-1} \sum_{q=0}^{n-1} |z[m,q]| \tag{8}$$

$$\leq \frac{\sqrt{2}}{n} \tag{9}$$

provided there is only one maximum frequency and hence only one (if the zero, DC frequency is maximal) or two (if a single nonzero frequency is maximal) nonzero entries in z. Let M be the maximum magnitude entry in D_p, and let $T = \max_{i,j,k,l} |x_c[i,j] x_c[k,l]|$ be the maximum magnitude of any pair of terms in x_c. We note that $T \leq |x_c[i,j]||x_c[k,l]| \leq \frac{2}{n^2}$. Hence

$$|x_c^* D_p x_c| \leq nKMT \tag{10}$$

$$\leq \frac{2KM}{n}, \tag{11}$$

and so $\lim_{p\to\infty} |x_c^* D_p x_c| = 0$. Therefore, since from Theorem B.3 we know circular and valid convolution limit to the same value, and from the preceding analysis we know x_c^{opt} applied to V_p attains the same objective when applied to C_p, we know that as $p \to \infty$, x_c^{opt} attains the optimal value for V_p. $\square$

References

[1] R. M. Gray. Toeplitz and Circulant Matrices: A review. *Foundations and Trends in Communications and Information Theory*, 2(3), 2005.