On the computation of a discrete signal power only using Half-spectra STFT magnitude bins

Angel M. Gómez
Universidad de Granada (Spain)

September 21, 2018

1 The basics

We start from the Parseval's theorem definition in the DFT domain:

$$E_x = \sum_{l=0}^{L-1} |x[l]|^2 = \frac{1}{L} \sum_{k=0}^{L-1} |X[k]|^2 \tag{1}$$

where E_x is the signal energy, L is the number of samples, and $X(k)$ is the DFT with L points of $x(l)$. Then, the signal power, P_x, can be obtained as,

$$P_x = \frac{1}{L} \sum_{l=0}^{L-1} |x[l]|^2 = \frac{1}{L^2} \sum_{k=0}^{L-1} |X[k]|^2. \tag{2}$$

2 Segmentation and STFT equivalence

Lets assume that $x(l)$ is segmented in M frames of length N (even number). Whether frames are overlapped or not (but the complete signal is covered), we can rewrite (2) as:

$$P_x = \frac{1}{M} \sum_{m=0}^{M-1} P_x(m) = \frac{1}{M} \sum_{m=0}^{M-1} \left(\frac{1}{N} \sum_{n=0}^{N-1} |x_m[n]|^2 \right) = \frac{1}{M} \sum_{m=0}^{M-1} \left(\frac{1}{N^2} \sum_{k=0}^{N-1} |X_m[k]|^2 \right) \tag{3}$$

where $P_x(m)$ and $X_m[k]$ are the power and the DFT of frame m, respectively.

Now, we will consider each frame is windowed so that we can understand the previous segmentation as a STFT. Windowing causes a distortion over the signal power which can be approximated as,

$$P_{x_w}(m) = \frac{1}{N} \sum_{n=0}^{N-1} |x_m[n]w[n]|^2 \approx \frac{1}{N} \sum_{n=0}^{N-1} |x[n]|^2 \cdot \frac{1}{N} \sum_{n=0}^{N-1} |w[n]|^2 = P_x(m) \cdot P_w \tag{4}$$

which can alternatively be expressed in terms of the RMS for infinite signals as,

$$RMS(x_w) = RMS(x) \cdot RMS(w) \tag{5}$$

as long as $x(n)$ and $w(n)$ are independent.

When windowing is assumed during spectra computation, $\hat{X}_m[k]$, we can approximate the signal power as,

$$P_x \approx \frac{1}{M} \sum_{m=0}^{M-1} \left(\frac{1}{N^2} \sum_{k=0}^{N-1} |\hat{X}_m[k]|^2 \right) \cdot \frac{1}{P_w} \tag{6}$$

where it must be taken into account that,

$$P_x(m) \approx \frac{1}{P_w} \cdot \frac{1}{N} \sum_{n=0}^{N-1} |x_m[n]w[n]|^2 \tag{7}$$

3 Average operator and Half-spectra equivalence

Lets now define the operator $\overline{\sum_{K}}(\cdot) = \frac{1}{K}\sum_{k=0}^{K-1}(\cdot)$, which can be understood as the average of the elements in $(\cdot)$, so that we can compactly rewrite (6) as,

$$P_x \approx \overline{\sum_{M}}\left(\overline{\sum_{N}}|\hat{X}_m[n]|^2\right)\cdot\frac{1}{NP_w} \tag{8}$$

Then, we will assume that $x(n)$ is a real signal and only half-spectra are available, i.e. $H_m[i]$ with $i = 0, \ldots, N/2$. If $H_m[0] = H_m[N/2] = 0$ (i.e. $x(n)$ is a band pass-signal) then,

$$\begin{aligned}
\overline{\sum_{N}}|\hat{X}_m[n]|^2 &= \frac{1}{N}\left(\sum_{n=1}^{N/2-1}|\hat{X}_m[n]|^2 + \sum_{n=N/2+1}^{N-1}|\hat{X}_m[n]|^2\right) && (9)\\
&= \frac{2}{N}\sum_{i=0}^{N/2}|\hat{X}_m[i]|^2 = \frac{2}{N}\frac{\frac{N}{2}+1}{\frac{N}{2}+1}\sum_{i=0}^{N/2}|H_m[i]|^2 && (10)\\
&= \frac{2(\frac{N}{2}+1)}{N}\overline{\sum_{I}}|H_m[i]|^2 = \frac{N+2}{N}\cdot\overline{\sum_{I}}|H_m[i]|^2 && (11)
\end{aligned}$$

Therefore, we can finally approximate the time-domain signal's power from its half-spectra STFT representation as,

$$P_x \approx \overline{\sum_{M}\sum_{I}}|H_m[i]|^2\cdot\frac{N+2}{N^2}\cdot\frac{1}{P_w} \tag{12}$$

where it is must be noted that N is the number of DFT frequency bins and $\overline{\sum_{K}}(\cdot)$ corresponds to the TensorFlow operator `tf.reduce_mean()`.