Theorem 12.3 (Bousquet and Elisseeff [67]) *Assume that we have a β-stable algorithm with the additional requirement that $|f_Z(x)| \leq M$ for all $x \in \mathcal{X}$ and for all training samples $Z \subset \mathcal{X} \times \mathcal{Y}$. Then, for $m \geq \frac{8M^2}{\varepsilon^2}$, we have,*

$$\mathrm{P}\left\{\left|R_{\mathrm{emp}}[f_Z] - R[f_Z]\right| > \varepsilon\right\} \leq \frac{64Mm\beta + 8M^2}{m\varepsilon^2} \tag{12.5}$$

and for any $m \geq 1$

$$\mathrm{P}\left\{\left|R_{\mathrm{emp}}[f_Z] - R[f_Z]\right| > \varepsilon + \beta\right\} \leq 2\exp\left(-\frac{m\varepsilon^2}{2(m\beta + M)^2}\right). \tag{12.6}$$

This means that if β decreases with increasing m, or, in particular, if $\beta = O(m^{-1})$, then we obtain bounds that are optimal in their rate of convergence, specifically, bounds which have the same convergence rate as Hoeffding's bound (5.7).

To keep matters simple, we only prove (12.6). The details for the proof of (12.5), which is rather technical, can be found in [162].

Proof We first give a bound on the expected difference between $R_{\mathrm{emp}}[f_Z]$ and $R[f_Z]$ (hence the bias term) and subsequently will bound the variance. This leads to

$$\left|\mathbf{E}_Z\left[R_{\mathrm{emp}}[f_Z] - R[f_Z]\right]\right| = \left|\mathbf{E}_{Z,z}\left[\frac{1}{m}\sum_{i=1}^{m} c(x_i, y_i, f_Z(x_i)) - c(x, y, f_Z(x))\right]\right| \tag{12.7}$$

$$= \left|\mathbf{E}_Z\left[\frac{1}{m}\sum_{i=1}^{m} c(x, y, f_{Z^i}(x)) - c(x, y, f_Z(x))\right]\right| \leq \beta \tag{12.8}$$

The last equality (12.8) followed from the fact that, since we are taking the expectation over Z, z, we may as well replace z_i by z in the terms stemming from the empirical error. The bound then follows from the assumption that we have a uniformly β-stable algorithm.

Now that we have a bound on the expectation, we deal with the variance. Since we want to apply Theorem 12.1, we have to analyze the deviations of $(R_{\mathrm{emp}}[f_Z] - R[f_Z])$ from $(R_{\mathrm{emp}}[f_{Z^i}] - R[f_{Z^i}])$.

$$\left|\left(R_{\mathrm{emp}}\left[f_Z\right] - R\left[f_Z\right]\right) - \left(R_{\mathrm{emp}}\left[f_{Z^i}\right] - R\left[f_{Z^i}\right]\right)\right| \leq \tag{12.9}$$

$$\left|R\left[f_Z\right] - R\left[f_{Z^i}\right]\right| + \left|R_{\mathrm{emp}}\left[f_Z\right] - R_{\mathrm{emp}}\left[f_{Z^i}\right]\right| \leq \tag{12.10}$$

$$\beta + \frac{1}{m}\left|c\left(x_i, y_i, f_Z(x_i)\right) - c\left(x, y, f_{Z^i}(x)\right)\right|$$

$$+\frac{1}{m}\sum_{j \neq i}^{m}\left|c\left(x_j, y_j, f_Z(x_j)\right) - c\left(x_j, y_j, f_{Z^i}(x_j)\right)\right| \leq \beta + \tfrac{2M}{m} + \beta \tag{12.11}$$

Here (12.10) follows from the triangle inequality and the fact that the learning algorithm is β-stable. Finally, we split the empirical risks into their common parts depending on Z^i and the remainder. From (12.11) it follows that $c_i = 2\frac{\beta m + M}{m}$, as required by Theorem 12.1. This, in combination with (12.8), completes the proof.

■