minimizing over α and μ.

Solving the dual program (for example, using ADMM, see Section 56.3) does not determine b, and for this we use the KKT conditions. The KKT conditions (for the primal program) are

$$\begin{aligned}
\lambda_i(w^\top x_i + b - y_i - \epsilon - \xi_i) &= 0, \quad i = 1, \ldots, m\\
\mu_i(-w^\top x_i - b + y_i - \epsilon - \xi_i') &= 0, \quad i = 1, \ldots, m\\
\gamma\epsilon &= 0\\
\alpha_i\xi_i &= 0, \quad i = 1, \ldots, m\\
\beta_i\xi_i' &= 0, \quad i = 1, \ldots, m.
\end{aligned}$$

If $\epsilon > 0$, since the equations

$$\begin{aligned}
w^\top x_i + b - y_i &= \epsilon + \xi_i\\
-w^\top x_i - b + y_i &= \epsilon + \xi_i'
\end{aligned}$$

cannot hold simultaneously, we must have

$$\lambda_i\mu_i = 0, \quad i = 1, \ldots, m. \tag{$\lambda\mu$}$$

From the equations

$$\lambda_i + \alpha_i = \frac{C}{m}, \quad \mu_i + \beta_i = \frac{C}{m}, \quad \alpha_i\xi_i = 0, \quad \beta_i\xi_i' = 0,$$

we get the equations

$$\left(\frac{C}{m} - \lambda_i\right)\xi_i = 0, \quad \left(\frac{C}{m} - \mu_i\right)\xi_i' = 0, \quad i = 1, \ldots, m. \tag{$*$}$$

Suppose we have optimal solution with $\epsilon > 0$. Using the above equations and the fact that $\lambda_i\mu_i = 0$ we obtain the following classification of the points x_i in terms of λ and μ.

(1) $0 < \lambda_i < C/m$. By $(*)$, $\xi_i = 0$, so the equation $w^\top x_i + b - y_i = \epsilon$ holds and x_i is on the blue margin hyperplane $H_{w,b-\epsilon}$. See Figure 56.5.

(2) $0 < \mu_i < C/m$. By $(*)$, $\xi_i' = 0$, so the equation $-w^\top x_i - b + y_i = \epsilon$ holds and x_i is on the red margin hyperplane $H_{w,b+\epsilon}$. See Figure 56.5.

(3) $\lambda_i = C/m$. By $(\lambda\mu)$, $\mu_i = 0$, and by $(*)$, $\xi_i' = 0$. Thus we have

$$\begin{aligned}
w^\top x_i + b - y_i &= \epsilon + \xi_i\\
-w^\top x_i - b + y_i &\leq \epsilon.
\end{aligned}$$

The second inequality is equivalent to $-\epsilon \leq w^\top x_i + b - y_i$, and since $\epsilon > 0$ and $\xi_i \geq 0$ it is trivially satisfied. If $\xi_i = 0$, then x_i is on the blue margin $H_{w,b-\epsilon}$, else x_i is an error and it lies in the open half-space bounded by the blue margin $H_{w,b-\epsilon}$ and not containing the best fit hyperplane $H_{w,b}$ (it is outside of the ϵ-slab). See Figure 56.5.