(17) Observe that for all $w \in \mathbb{R}^d$ and for all data points (x_i, y_i), by the definition of the margin $\gamma(w)$,

$$y_i \frac{w^T x_i}{\|w\| \cdot \gamma(w)} \geq 1$$

but $v = \frac{w}{\|w\| \gamma(w)}$. Maximizing $\gamma(w)$ is equivalent to minimizing $\|v\| = \frac{1}{\gamma(w)}$. The optimization problem becomes

$$\min \|v\| \text{ subject to } y_i \cdot v^T x_i \geq 1 \quad \forall i = 1, \dots, n.$$

This is a convex optimization problem with linear constraints. Equivalent (and more practical) formulation:

$$\min \|v\|^2 \text{ subject to } y_i v^T x_i \geq 1 \quad \forall i = 1, \dots, n.$$

To show that the optimal v lies in the subspace spanned by those x_i for which $y_i v^T x_i = 1$ (these are the support vectors), suppose it is not true. Then (just like in the proof of the representer theorem) v has a component in the orthogonal complement of the support vectors. By projecting orthogonally to the subspace spanned by the support vectors, the projection $\tilde{v}$ satisfies $\tilde{v}^T x_i = v^T x_i$ for all support vectors — and therefore has the same margin — but $\|\tilde{v}\| < \|v\|$, contradicting the optimality of v.