equivalently

$$\begin{aligned}
&\text{minimize} \quad \left\| \sum_{j=1}^{q} \mu_j v_j - \sum_{i=1}^{p} \lambda_i u_i \right\|_2 \\
&\text{subject to} \\
&\qquad \sum_{i=1}^{p} \lambda_i = 1, \ \ \lambda \geq 0 \\
&\qquad \sum_{j=1}^{q} \mu_j = 1, \ \ \mu \geq 0.
\end{aligned}$$

Geometrically, $\sum_{i=1}^p \lambda_i u_i$ with $\sum_{i=1}^p \lambda_i = 1$ and $\lambda \geq 0$ is a convex combination of the u_is, and $\sum_{j=1}^q \mu_j v_j$ with $\sum_{j=1}^q \mu_j = 1$ and $\mu \geq 0$ is a convex combination of the v_js, so the dual program is to minimize the distance between the polyhedron $\mathrm{conv}(u_1, \ldots, u_p)$ (the convex hull of the u_is) and the polyhedron $\mathrm{conv}(v_1, \ldots, v_q)$ (the convex hull of the v_js). Since both polyhedra are compact, the shortest distance between then is achieved. In fact, there is some vertex u_i such that if $P(u_i)$ is its projection onto $\mathrm{conv}(v_1, \ldots, v_q)$ (which exists by Hilbert space theory), then the length of the line segment $(u_i, P(u_i))$ is the shortest distance between the two polyhedra (and similarly there is some vertex v_j such that if $P(v_j)$ is its projection onto $\mathrm{conv}(u_1, \ldots, u_p)$ then the length of the line segment $(v_j, P(v_j))$ is the shortest distance between the two polyhedra).

If the two subsets are separable, in which case Problem (SVM_{h1}) has an optimal solution $\delta > 0$, because the objective function is convex and the convex constraint $\|w\|_2 \leq 1$ is qualified since δ may be negative, by Theorem 50.17(2) the duality gap is zero, so δ is half of the minimum distance between the two convex polyhedra $\mathrm{conv}(u_1, \ldots, u_p)$ and $\mathrm{conv}(v_1, \ldots, v_q)$; see Figure 50.19.

It should be noted that the constraint $\|w\| \leq 1$ yields a formulation of the dual problem which has the advantage of having a nice geometric interpretation: finding the minimal distance between the convex polyhedra $\mathrm{conv}(u_1, \ldots, u_p)$ and $\mathrm{conv}(v_1, \ldots, v_q)$. Unfortunately this formulation is not useful for actually solving the problem. However, if the equivalent constraint $\|w\|^2 \,(= w^\top w) \leq 1$ is used, then the dual problem is much more useful as a solving tool.

In Chapter 54 we consider the case where the sets of points $\{u_1, \ldots, u_p\}$ and $\{v_1, \ldots, v_q\}$ are not linearly separable.

50.12 Some Techniques to Obtain a More Useful Dual Program

In some cases, it is advantageous to reformulate a primal optimization problem to obtain a more useful dual problem. Three different reformulations are proposed in Boyd and Van-