This program is obviously equivalent to the original one. By Example 50.8(7), the conjugate of the square norm is given by

$$\frac{1}{2}\Big(\|y\|^D\Big)^2,$$

so the dual of the reformulated program is

$$\begin{aligned}
&\text{maximize} \quad -\frac{1}{2}\Big(\|\mu\|^D\Big)^2 + b^\top \mu \\
&\text{subject to} \\
&\qquad\qquad A^\top \mu = 0.
\end{aligned}$$

Note that this dual is different from the dual obtained in Example 50.13.

The objective function of the dual program in Example 50.13 is linear, but we have the nonlinear constraint $\|\mu\|^D \leq 1$. On the other hand, the objective function of the dual program of Example 50.14 is quadratic, whereas its constraints are affine. We have other examples of this trade-off with the Programs (SVM_{h2}) (quadratic objective function, affine constraints), and (SVM_{h1}) (linear objective function, one nonlinear constraint).

Sometimes, it is also helpful to replace a constraint by an increasing function of this constraint; for example, to use the constraint $\|w\|_2^2\,(= w^\top w) \leq 1$ instead of $\|w\|_2 \leq 1$.

In Chapter 55 we revisit the problem of solving an overdetermined or underdetermined linear system $Ax = b$ considered in Section 23.1, from a different point of view.

50.13 Uzawa's Method

Let us go back to our Minimization Problem

$$\begin{aligned}
&\text{minimize} \quad J(v) \\
&\text{subject to} \quad \varphi_i(v) \leq 0, \quad i = 1, \ldots, m,
\end{aligned}$$

where the functions J and φ_i are defined on some open subset Ω of a finite-dimensional Euclidean vector space V (more generally, a real Hilbert space V). As usual, let

$$U = \{v \in V \mid \varphi_i(v) \leq 0,\ 1 \leq i \leq m\}.$$

If the functional J satisfies the inequalities of Proposition 49.18 and if the functions φ_i are convex, in theory, the projected-gradient method converges to the unique minimizer of J over U. Unfortunately, it is usually impossible to compute the projection map $p_U \colon V \to U$.

On the other hand, the domain of the Lagrange dual function $G \colon \mathbb{R}_+^m \to \mathbb{R}$ given by

$$G(\mu) = \inf_{v \in \Omega} L(v, \mu) \quad \mu \in \mathbb{R}_+^m,$$