the KKT conditions are useless unlesss the constraints are convex. In this case, there is a manageable notion of qualified constraint given by Slater's conditions. Theorem 50.6 gives necessary conditions for a function J to have a minimum on a subset U defined by convex inequality constraints in terms of the Karush–Kuhn–Tucker conditions. Furthermore, if J is also convex and if the KKT conditions hold, then J has a global minimum.

In Section 50.4, we apply Theorem 50.6 to the special case where the constraints are equality constraints, which can be expressed as $Ax = b$. In the special case where the convex objective function J is a convex quadratic functional of the form

$$J(x) = \frac{1}{2}x^\top Px + q^\top x + r,$$

where P is a $n \times n$ symmetric positive semidefinite matrix, the necessary and sufficient conditions for having a minimum are expressed by a linear system involving a matrix called the KKT matrix. We discuss conditions that guarantee that the KKT matrix is invertible, and how to solve the KKT system. We also briefly discuss variants of Newton's method dealing with equality constraints.

We illustrate the KKT conditions on an interesting example, the so-called hard margin support vector machine; see Sections 50.5 and 50.6. The problem is a classification problem, or more accurately a separation problem. Suppose we have two nonempty disjoint finite sets of p *blue* points $\{u_i\}_{i=1}^p$ and q *red* points $\{v_j\}_{j=1}^q$ in $\mathbb{R}^n$. Our goal is to find a hyperplane H of equation $w^\top x - b = 0$ (where $w \in \mathbb{R}^n$ is a nonzero vector and $b \in \mathbb{R}$), such that all the blue points u_i are in one of the two open half-spaces determined by H, and all the red points v_j are in the other open half-space determined by H.

If the two sets are indeed separable, then in general there are infinitely many hyperplanes separating them. Vapnik had the idea to find a hyperplane that maximizes the smallest distance between the points and the hyperplane. Such a hyperplane is indeed unique and is called a maximal hard margin hyperplane, or hard margin support vector machine. The support vectors are those for which the constraints are active.

Section 50.7 contains the most important results of the chapter. The notion of Lagrangian duality is presented. Given a primal optimization problem (P) consisting in minimizing an objective function $J(v)$ with respect to some inequality constraints $\varphi_i(v) \leq 0$, $i = 1, \ldots, m$, we define the *dual function* $G(\mu)$ as the result of minimizing the Lagrangian

$$L(v, \mu) = J(v) + \sum_{i=1}^m \mu_i \varphi_i(v)$$

with respect to v, with $\mu \in \mathbb{R}_+^m$. The dual program (D) is then to maximize $G(\mu)$ with respect to $\mu \in \mathbb{R}_+^m$. It turns out that G is a concave function, and the dual program is an unconstrained maximization. This is actually a misleading statement because G is generally a partial function, so maximizing $G(\mu)$ is equivalent to a constrained maximization problem in which the constraints specify the domain of G, but in many cases, we obtain a dual