convex functions then followed immediately from the previous reasoning. The main results are dualization, meaning the transformation of optimization problems via the Lagrangian mechanism into possibly simpler problems, and that optimality properties can be estimated via the KKT gap (Theorem 6.27).

Interior point algorithms are practical applications of the duality reasoning; these seek to find a solution to optimization problems by satisfying the KKT optimality conditions. Here we were able to employ some of the concepts introduced at an earlier stage, such as predictor corrector methods and numerical ways of finding roots of equations. These algorithms are robust tools to find solutions on moderately sized problems ($10^3 - 10^4$ examples). Larger problems require decomposition methods, to be discussed in Section 10.4, or randomized methods. The chapter concluded with an overview of randomized methods for maximizing functions or finding the best subset of elements. These techniques are useful once datasets are so large that we cannot reasonably hope to find exact solutions to optimization problems.

6.7 Problems

6.1 (Level Sets •) *Given the function $f : \mathbb{R}^2 \to \mathbb{R}$ with $f(x) := |x_1|^p + |x_2|^p$, for which p do we obtain a convex function?*

Now consider the sets $\{x | f(x) \leq c\}$ for some $c > 0$. Can you give an explicit parametrization of the boundary of the set? Is it easier to deal with this parametrization? Can you find other examples (see also [489] and Chapter 8 for details)?

6.2 (Convex Hulls •) *Show that for any set X, its convex hull $\operatorname{co} X$ is convex. Furthermore, show that $\operatorname{co} X = X$ if X is convex.*

6.3 (Method of False Position [334] •••) *Given a unimodal (possessing one minimum) differentiable function $f : \mathbb{R} \to \mathbb{R}$, develop a quadratic method for minimizing f.*

Hint: Recall the Newton method. There we used $f''(x)$ to make a quadratic approximation of f. Two values of $f'(x)$ are also sufficient to obtain this information, however.

What happens if we may only use f? What does the iteration scheme look like? See Figure 6.8 for a hint.

6.4 (Convex Minimization in one Variable ••) *Denote by f a convex function on $[a, b]$. Show that the algorithm below finds the minimum of f. What is the* rate *of convergence in x to $\operatorname{argmin}_x f(x)$? Can you obtain a bound in $f(x)$ wrt. $\min_x f(x)$?*

input a, b, f *and threshold* ε
  $x_1 = a, x_2 = \frac{a+b}{2}, x_3 = b$ *and compute* $f(x_1), f(x_2), f(x_3)$
  repeat
    if $x_3 - x_2 > x_2 - x_1$ ***then***

$x_4 = \frac{x_2+x_3}{2}$ *and compute* $f(x_4)$
else
$x_4 = \frac{x_1+x_2}{2}$ *and compute* $f(x_4)$
end if
Keep the two points closest to the point with the minimum value of $f(x_i)$ *and rename them such that* $x_1 < x_2 < x_3$.
until $x_3 - x_1 \geq \varepsilon$

6.5 (Newton Method in $\mathbb{R}^d$ ••) *Extend the Newton method to functions on* $\mathbb{R}^d$. *What does the iteration rule look like? Under which conditions does the algorithm converge? Do you have to extend Theorem 6.13 to prove convergence?*

6.6 (Rewriting Quadratic Functionals •) *Given a function*

$$f(x) = x^\top Q x + c^\top x + d, \tag{6.97}$$

rewrite it into the form of (6.18). Give explicit expressions for $x^* = \operatorname{argmin}_x f(x)$ *and the difference in the additive constants.*

6.7 (Kantorovich Inequality [278] •••) *Prove Theorem 6.16. Hint: note that without loss of generality we may require* $\|x\|^2 = 1$. *Second, perform a transformation of coordinates into the eigensystem of K. Finally, note that in the new coordinate system we are dealing with convex combinations of eigenvalues* λ_i *and* $\frac{1}{\lambda_i}$. *First show (6.24) for only two eigenvalues. Then argue that only the largest and smallest eigenvalues matter.*

6.8 (Random Subsets •) *Generate m random numbers drawn uniformly from the interval* $[0,1]$. *Plot their distribution function. Plot the distribution of maxima of subsets of random numbers. What can you say about the distribution of the maxima? What happens if you draw randomly from the Laplace distribution, with density* $p(\xi) = e^{-\xi}$ *(for* $\xi \geq 0$*)?*

6.9 (Matching Pursuit [342] ••) *Denote by* $f_1, \ldots, f_M$ *a set of functions* $\mathcal{X} \to \mathbb{R}$, *by* $\{x_1, \ldots, x_m\} \subset \mathcal{X}$ *a set of locations and by* $\{y_1, \ldots, y_m\} \subset \mathcal{Y}$ *a set of corresponding observations.*

Design a sparse greedy algorithm that finds a linear combination of functions $f := \sum_i \alpha_i f_i$ *minimizing the squared loss between* $f(x_i)$ *and* y_i.

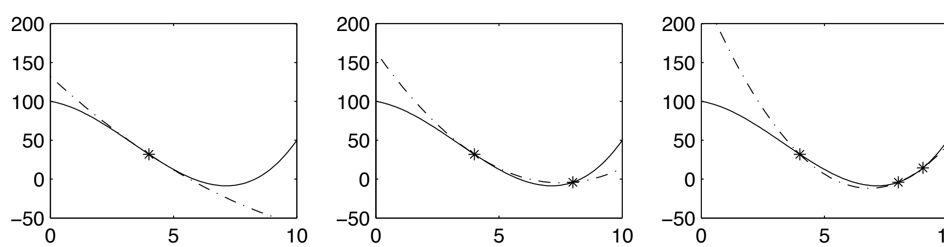

Figure 6.8 From left to right: Newton method, method of false position, quadratic interpolation through 3 points. Solid line: $f(x)$, dash-dotted line: interpolation.

6.10 (Reduced Set Approximation [474] ●●) *Let $f(x) = \sum_{i=1}^{m} \alpha_i k(x_i, x)$ be a kernel expansion in a Reproducing Kernel Hilbert Space $\mathcal{H}_k$ (see Section 2.2.3). Give a sparse greedy algorithm that finds an approximation to f in $\mathcal{H}_k$ by using fewer terms. See also Chapter 18 for more detail.*

6.11 (Equality Constraints in LP and QP ●●) *Find the dual optimization problem and the necessary KKT conditions for the following optimization problem:*

$$\begin{aligned} &\underset{x}{\text{minimize}} && c^\top x, \\ &\textit{subject to} && Ax + b \leq 0, \\ & && Cx + d = 0, \end{aligned} \tag{6.98}$$

where $c, x \in \mathbb{R}^m$, $b \in \mathbb{R}^n$, $d \in \mathbb{R}^{n'}$, $A \in R^{n \times m}$ and $C \in \mathbb{R}^{n'}$. Hint: split up the equality constraints into two inequality constraints. Note that you may combine the two Lagrange multipliers again to obtain a free variable. Derive the corresponding conditions for

$$\begin{aligned} &\underset{x}{\text{minimize}} && \tfrac{1}{2} x^\top K x + c^\top x, \\ &\textit{subject to} && Ax + b \leq 0, \\ & && Cx + d = 0, \end{aligned} \tag{6.99}$$

where K is a strictly positive definite matrix.

6.12 (Not Strictly Definite Quadratic Parts ●●●) *How do you have to change the dual of (6.99) if K does not have full rank? Is it better not to dualize in this case? Do the KKT conditions still hold?*

6.13 (Dual Problems of Quadratic Programs ●●) *Denote by P a quadratic optimization problem of type (6.72) and by $(\cdot)^D$ the dualization operation. Prove that the following is true,*

$$((P^D)^D)^D = P^D \textit{ and } (((P^D)^D)^D)^D = (P^D)^D, \tag{6.100}$$

where in general $(P^D)^D \neq P$. Hint: use (6.80). Caution: you have to check whether $KA^\top$ has full rank.

6.14 (Interior Point Equations for Linear Programs [336] ●●●) *Derive the interior point equations for linear programs. Hint: use the expansions for the quadratic programs and note that the reduced KKT system has only a diagonal term where we had K before.*

How does the complexity of the problem scale with the size of A?

6.15 (Update Step in Interior Point Codes ●) *Show that the maximum value of λ satisfying (6.84) can be found by*

$$\frac{1}{\lambda} = \max\left(1, (\epsilon - 1)^{-1} \min_{i \in [n]} \frac{\Delta\alpha_i}{\alpha_i}, (\epsilon - 1)^{-1} \min_{i \in [n]} \frac{\Delta\xi_i}{\xi_i}\right). \tag{6.101}$$