1 Question 1

The problem to solve is:

$$\text{minimize } \frac{1}{2}\|Xw - y\|_2^2 + \lambda\|w\|_1 \tag{LASSO}$$

where $w \in \mathbb{R}^d$, $X = \begin{pmatrix} x_1^T \\ \vdots \\ x_n^T \end{pmatrix} \in \mathbb{R}^{n\times d}$, $y = (y_1, \ldots, y_n)^T \in \mathbb{R}^n$, and $\lambda > 0$ is the regularization parameter.

1.1 Deriving the dual problem of LASSO and reformulating it as a quadratic problem

The dual problem can be expressed as:

$$\text{minimize } v^T Q v + p^T v \quad \text{subject to } Av \preceq b \tag{QP}$$

where $v \in \mathbb{R}^n$ and $Q \succeq 0$.
To derive the dual problem, we start by introducing an equality constraint to the original formulation:

$$\begin{gathered}\min_{z,w} \frac{1}{2}\|z\|_2^2 + \lambda\|w\|_1 \\ \text{subject to } z = Xw - y\end{gathered} \tag{1}$$

The Lagrangian and the dual function are given as follows:

$$L(w, z, \mu) = \frac{1}{2}\|z\|_2^2 + \lambda\|w\|_1 + \mu^T(z - Xw + y) \tag{2}$$

$$g(\mu) = \inf_{w,z} L(w, z, \mu) = \inf_z \left(\frac{1}{2}\|z\|_2^2 + \mu^T z\right) + \inf_w \left(\lambda\|w\|_1 - \mu^T Xw\right) + \mu^T y \tag{3}$$

The infimum with respect to z is calculated using the first-order condition $\nabla f(z) = 0$, as the function is convex in z:

$$\nabla f(z) = z + \mu = 0 \quad \implies z = -\mu \tag{4}$$

Next, to compute the infimum with respect to w, we use the result of Exercise 2.1 from a previous homework. The conjugate function of $f(x) = \|x\|_1$ is:

$$f^*(y) = \begin{cases} 0 & \text{if } \|y\|_\infty \le 1 \\ \infty & \text{otherwise} \end{cases}$$

Rewriting the infimum as a supremum, we have:

$$\inf_w(\lambda\|w\|_1 - \mu^T Xw) = -\sup_w(\mu^T Xw - \lambda\|w\|_1) = \lambda f^*\left(\frac{X^T\mu}{\lambda}\right) \tag{6}$$

Thus, the dual function becomes:

$$g(\mu) = -\frac{1}{2}\mu^T\mu + \lambda f^*\left(\frac{X^T\mu}{\lambda}\right) + \mu^T y \tag{7}$$

Finally, the dual problem is formulated as:

$$\begin{gathered}\max_\mu \mu^T y - \frac{1}{2}\mu^T\mu \\ \text{subject to } \|X^T\mu\|_\infty \le \lambda\end{gathered} \tag{8}$$

This is equivalent to the following quadratic programming (QP) problem:

$$\min_{\mu} \frac{1}{2}\mu^T\mu - \mu^T y \quad \text{subject to } \|X^T y\|_\infty \leq \lambda \tag{9}$$

The terms in this formulation are as follows:

- $Q = \frac{1}{2} I_{d\times n}$
- $p = y$
- $b \in \mathbb{R}^{2d}$, $b_i = \lambda \forall i$
- $A \in \mathbb{R}^{2d\times n} = (X, -X)^T$

It is important to note that the dimensionality $2d$ arises due to the infinity norm, which enforces two constraints per coordinate of $X^T v$: $(X^T)_i v \leq \lambda$ and $(X^T)_i v \geq -\lambda$.

2 Question 2

To solve the optimization problem, we define the function f, its gradient ∇f, and its Hessian $H(f)$ as follows:

2.1 Definition of f

The function f is defined as:

$$f(v, Q, p, b, A, t) = t\left(v^\top Q v + p^\top v\right) - \sum_{i=1}^{2n} \log\left(b_i - (A^\top v)_i\right),$$

where:

- $A \in \mathbb{R}^{n\times 2d}$
- $Q \in \mathbb{R}^{n\times n}$
- $v, p \in \mathbb{R}^n$
- $b \in \mathbb{R}^{2n}$
- $t \in \mathbb{R}$

2.2 Gradient of f

The gradient ∇f is given by:

$$\nabla f(v, Q, p, b, A, t) = t\,(2Qv + p) - A\left(\frac{1}{b - A^\top v}\right),$$

where $\frac{1}{b-A^\top v}$ is a vector of dimension $2n$ with each element:

$$\left[\frac{1}{b - A^\top v}\right]_i = \frac{1}{b_i - (A^\top v)_i}.$$

2.3 Hessian of f

The Hessian $H(f)$ is expressed as:

$$H(f)(v, Q, b, A, t) = t \cdot 2Q + A \cdot \text{diag}\left(\frac{1}{(b - A^\top v)^2}\right) \cdot A^\top,$$

where $\text{diag}\left(\frac{1}{(b-A^\top v)^2}\right)$ is a diagonal matrix with diagonal elements:

$$\text{diag}\left(\frac{1}{(b - A^\top v)^2}\right)_{ii} = \frac{1}{(b_i - (A^\top v)_i)^2}.$$