Discussion 12

1. **Newton's Method for Quadratic Functions**

Give a symmetric positive defininte matrix $Q \in \mathbb{S}^n_{++}$ and $b \in \mathbb{R}^n$, consider minimizing

$$f(x) = \frac{1}{2}\vec{x}^\top Q\vec{x} - \vec{b}^\top \vec{x} \tag{1}$$

Let $\vec{x}^\star$ denote the point at which $f(\vec{x})$ is minimized, and define $\mathcal{B}(\vec{x}^\star)$ as the ball centered at $\vec{x}^\star$ with unit ℓ_2 norm:

$$\mathcal{B}(\vec{x}^\star) = \{\vec{x} \in \mathbb{R}^n : \|\vec{x} - \vec{x}^\star\|_2 \leq 1\} \tag{2}$$

Assume we use Newton's method to minimize f:

$$\vec{x}_{k+1} = \vec{x}_k - (\nabla^2 f(\vec{x}_k))^{-1}\nabla f(\vec{x}_k) \tag{3}$$

where the initial point is $\vec{x}_0 \in \mathcal{B}(\vec{x}^\star)$.

For any $k \in \mathbb{N}$, find

$$\max_{\vec{x}_0 \in \mathcal{B}(\vec{x}^\star)} \|\vec{x}_k - \vec{x}^\star\|_2. \tag{4}$$

2. **Generalized Linear Models**

A wide class of machine learning models (e.g. classification and regression) can be modelled in a common framework called generalised linear models (GLMs). In this problem, we'll talk about exponential families, generalized linear models and use Newton's method to perform maximum likelihood estimation (MLEs). Consider a special class of probability distributions known as exponential families whose density is of the form

$$f(\vec{y}; \vec{\theta}) = e^{\vec{y}^\top \vec{\theta} - b(\vec{\theta})} f_0(\vec{y}) \tag{5}$$

where $\vec{y}, \vec{\theta} \in \mathbb{R}^n$ and $b(\vec{\theta}) = \log\left(\int_{\mathbb{R}^n} e^{\vec{y}^\top \vec{\theta}} f_0(\vec{y}) d\vec{y}\right)$ is the normalizing constant which ensures f is a probability distribution over $\vec{y}$.

(a) Show that $b(\vec{\theta})$ is a convex function.

(b) We model $\vec{\theta} = X\vec{\beta}$ where $X \in \mathbb{R}^{n\times d}$ is the data matrix. Under this parameterization of $\vec{\theta}$, the exponential family is called a generalized linear model. Prove that $b(X\vec{\beta})$ is convex in $\vec{\beta}$.

(c) For a given exponential family/GLM model, MLE estimation for a data matrix X and corresponding output variables $\vec{y} \in \mathbb{R}^n$ corresponds to solving the following maximization problem:

$$\max_{\vec{\beta}} f(\vec{y}; X\vec{\beta}) \tag{6}$$

Prove that this maximization problem is equivalent to

$$\min_{\vec{\beta}} g(\vec{\beta}) := -\vec{y}^\top X\vec{\beta} + b(X\vec{\beta}) \tag{7}$$

Show that this is a convex optimization problem. Which choice of $b(\cdot)$ recovers linear regression?

(d) For the above convex minimization problem, find the undamped Newton's method (with step size 1) update. This update also goes by the name *iteratively reweighted least squares* (IRLS). Can you tell why? (For any iterate $\vec{\beta}$ and the Newton update on $\vec{\beta}$ denoted by $\vec{\beta}_+$, what optimization problem is $\vec{\beta}_+$ the optimum of?)

3. Robust Linear Programming

In this problem we will consider a version of linear programming under uncertainty.

Consider vector $\vec{x} \in \mathbb{R}^n$. Recall from the previous discussion that $\vec{x}^\top\vec{y} \leq \|\vec{x}\|_1$ for all $\vec{y}$ such that $\|\vec{y}\|_\infty \leq 1$. Further this inequality is tight, since it holds with equality for $\vec{y} = \operatorname{sgn}(\vec{x})$.

Let us focus now on a LP in standard form:

$$\min_{\vec{x}} \quad \vec{c}^\top\vec{x} \tag{8}$$

$$\text{s.t.} \quad \vec{a}_i^\top\vec{x} \leq b_i, \quad i = 1, ..., m. \tag{9}$$

Consider the set of linear inequalities in (9). Suppose you don't know the vectors $\vec{a}_i$ exactly. Instead you are given nominal values $\vec{\hat{a}}_i$, and you know that the actual vectors satisfy $\left\|\vec{a}_i - \vec{\hat{a}}_i\right\|_\infty \leq \rho$ for a given $\rho > 0$. In other words, the actual components a_{ij} can be anywhere in the intervals $[\hat{a}_{ij} - \rho, \hat{a}_{ij} + \rho]$. Or equivalently, each vector $\vec{a}_i$ can lie anywhere in a hypercube with corners $\vec{\hat{a}}_i + \vec{v}$ where $\vec{v} \in \{-\rho, \rho\}^n$. We desire that the set of inequalities that constrain problem (9) be satisfied for all possible values of $\vec{a}_i$; i.e., we replace these with the constraints

$$\vec{a}_i^\top \vec{x} \leq b_i \ \ \forall \vec{a}_i \in \{\vec{\hat{a}}_i + \vec{v} \ \mid \|\vec{v}\|_\infty \leq \rho\} \ \ i = 1, ..., m. \tag{10}$$

Note that the above defines an *infinite* number of constraints (of the form $\vec{\hat{a}}_i^\top \vec{x} + \vec{v}^\top \vec{x} \leq b_i, \quad \forall \vec{v}$ satisfying $\|\vec{v}\|_\infty \leq \rho, i = 1, 2, \ldots, m$).

(a) Argue why for our LP we can replace the infinite set of constraints as above to a finite set of $2^n m$ constraints of the form,

$$\vec{\hat{a}}_i^\top \vec{x} + \vec{v}^\top \vec{x} \leq b_i \ \ \forall \vec{v} \in \{-\rho, \rho\}^n \ \ i = 1, ..., m. \tag{11}$$

HINT: What do you know about the optimal solutions of LPs?

(b) Use result from part (a) to show that the constraint set in Equation (10) is in fact equivalent to the much more compact set of m nonlinear inequalities

$$\vec{\hat{a}}_i^\top \vec{x} + \rho \|\vec{x}\|_1 \leq b_i, \ \ i = 1, ..., m. \tag{12}$$

We now would like to formulate the LP with uncertainty introduced. We are therefore interested in situations where the vectors $\vec{a}_i$ are uncertain, but satisfy bounds $\left\|\vec{a}_i - \hat{\vec{a}}_i\right\|_\infty \leq \rho$ for given $\hat{\vec{a}}_i$ and ρ. We want to minimize $\vec{c}^\top \vec{x}$ subject to the constraint that the inequalities $\vec{a}_i^\top \vec{x} \leq b_i$ are satisfied for *all* possible values of $\vec{a}_i$.
We call this a *robust LP* :

$$\min_{\vec{x}} \quad \vec{c}^\top \vec{x} \tag{13}$$
$$\text{s.t.} \quad \vec{a}_i^\top \vec{x} \leq b_i, \;\; \forall \vec{a}_i \in \{\hat{\vec{a}}_i + \vec{v} \mid \|\vec{v}\|_\infty \leq \rho\} \quad i = 1, ..., m. \tag{14}$$

(c) Using the result from part (c), express the above optimization problem as an LP.

