Discussion 12

We might not cover all the problems on the worksheet, as discussion worksheets are not always designed to be finished within an hour. However, this is totally fine since they are deliberately made slightly longer so they can serve as resources you can use to practice, reinforce, and build upon concepts covered in lectures and homework.

1. **Squaring SOCP Constraints**

   When considering a second-order cone (SOC) constraint, you might be tempted to square it to obtain a classical convex quadratic constraint. This problem explores why that might not always work, and how to introduce additional constraints to maintain equivalence and convexity.

   (a) For $\vec{x} \in \mathbb{R}^2$, consider the constraint

   $$x_1 - 2x_2 \geq \|\vec{x}\|_2\,, \tag{1}$$

   and its squared counterpart

   $$(x_1 - 2x_2)^2 \geq \|\vec{x}\|_2^2\,. \tag{2}$$

   Are the two sets equivalent? Are they both convex?

   (b) What additional constraint must be imposed alongside the squared constraint to enforce the same feasible set as the unsquared SOC constraint?

2. **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

   $$f(x) = \frac{1}{2}\vec{x}^\top Q\vec{x} - \vec{b}^\top\vec{x} \tag{3}$$

   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:

   $$\mathcal{B}(\vec{x}^\star) = \{\vec{x} \in \mathbb{R}^n : \|\vec{x} - \vec{x}^\star\|_2 \leq 1\} \tag{4}$$

   Assume we use Newton's method to minimize f:

   $$\vec{x}_{k+1} = \vec{x}_k - (\nabla^2 f(\vec{x}_k))^{-1}\nabla f(\vec{x}_k) \tag{5}$$

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

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

$$\max_{\vec{x}_0 \in \mathcal{B}(\vec{x}^\star)} \|\vec{x}_k - \vec{x}^\star\|_2. \tag{6}$$

3. **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

$$f(\vec{y}; \vec{\theta}) = e^{\vec{y}^\top \vec{\theta} - b(\vec{\theta})} f_0(\vec{y}) \tag{7}$$

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) **(OPTIONAL)** 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:

$$\max_{\vec{\beta}} f(\vec{y}; X\vec{\beta}) \tag{8}$$

Prove that this maximization problem is equivalent to

$$\min_{\vec{\beta}} g(\vec{\beta}) := -\vec{y}^\top X\vec{\beta} + b(X\vec{\beta}) \tag{9}$$

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?)