Problem Sheet 5

Problems in Part A will be discussed in class. The Problem in Part B can be tried at home.

Part A

(5.1) Let $(\Delta x, \Delta y, \Delta s) \in \mathbb{R}^{2n+m}$ be given as the solution of the system of equations

$$
\begin{pmatrix} \mathbf{0} & A^\top & I \\ A & \mathbf{0} & \mathbf{0} \\ S & \mathbf{0} & X \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta s \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ -XSe + \sigma\mu \end{pmatrix},
$$

with $A \in \mathbb{R}^{m\times n}$, $x \in \mathbb{R}^n$, $s \in \mathbb{R}^n$, $X = \mathrm{diag}(x_1, \ldots, x_n)$, $S = \mathrm{diag}(s_1, \ldots, s_n)$, $\sigma \in (0, 1)$, and $\mu = 1/n \sum_{i=1}^n x_i s_i$. Show that $\langle \Delta x, \Delta s \rangle = 0$.

(5.2) Consider the linear programming problem

$$\text{minimize} \quad x_1 \quad \text{subject to} \quad x_1 + x_2 = 1, \ x_1 \geq 0, \ x_2 \geq 0.$$

(a) Write down the equation $F(x, y, s)$ and the Jacobian JF, and write down Newton's method for this system.

(b) Solve the optimization problem graphically, or by any other means.

(c) Compute the solution to $F(x, y, s) = \mathbf{0}$ (either using Newton's method or by any other means) and conclude that it is unrelated to the solution of the optimization problem.

(5.3) A **linear programming feasibility problem** is the problem of finding a point that satisfies the constraints of the problem (not necessarily optimal with respect to the objective function).

Suppose we are given p data points $x_1, \ldots, x_p$ with associated labels $y_i \in \{-1, 1\}$. We would like to fine a linear hyperplane $h(x) = w^\top x + b$ so that $h(x_i) > 0$ if $y_i = 1$ and $h(x_j) < 0$ if $y_j = -1$. Show that this problem can be formulated as a linear programming feasibility problem.

Part B

(5.4) (Practical considerations.) Most of the computational cost of interior point methods is taken up by solving systems of linear equations. Luckily, these systems are sparse and can be solved efficiently in practice. A few tricks can make the tasks easier.

Show that the system of equations

$$\begin{pmatrix} \mathbf{0} & \mathbf{A}^\top & \mathbf{I} \\ \mathbf{A} & \mathbf{0} & \mathbf{0} \\ \mathbf{S} & \mathbf{0} & \mathbf{X} \end{pmatrix} \begin{pmatrix} \Delta \mathbf{x} \\ \Delta \mathbf{y} \\ \Delta \mathbf{s} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ -\mathbf{X}\mathbf{S}\mathbf{e} + \sigma\mu \end{pmatrix},$$

is equivalent to the *normal equations*

$$\mathbf{A}\mathbf{D}^2\mathbf{A}^\top \Delta \mathbf{y} = \mathbf{b} - \sigma\mu\mathbf{A}\mathbf{S}^{-1}\mathbf{e},$$
$$\Delta \mathbf{s} = -\mathbf{A}^\top \Delta \mathbf{y},$$
$$\Delta \mathbf{x} = -\mathbf{x} + \sigma\mu\mathbf{S}^{-1}\mathbf{e} - \mathbf{D}^2\Delta \mathbf{s},$$

where $\mathbf{D} = \mathbf{S}^{-1/2}\mathbf{X}^{1/2}$, and $\mathbf{X}^{1/2}$ denotes the diagonal matrix whose entries are the square roots of those of $\mathbf{X}$. Argue how this form can be used to increase performance.