Reexam Continuous Optimization

19 February 2024, 14.00–17.00

This closed-book exam consists of 5 questions. Please start each question on a new page, write legibly, and hand in your work with the solutions in the correct order. In total, you can obtain 90 points. The final grade is $1 + \#\text{points}/10$ rounded to the nearest integer. Good luck!

1. Consider the optimization problem

$$\begin{array}{ll} \text{minimize} & -\log(c^\mathsf{T} x) \\ \text{subject to} & Ax = b \\ & x^\mathsf{T} B x + d^\mathsf{T} x \leq 1. \end{array}$$

   where we assume B to be symmetric.

   (a) (10 points) Work out the KKT conditions for this problem.

   (b) (5 points) Under what condition(s) on c, A, b, B, and d is this optimization problem convex? Explain why.

   (c) (5 points) Show that if this problem is convex, the primal and dual optimal values are the same.

2. (15 points) Consider the optimization problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m, \end{array}$$

   where all functions are convex and continuously differentiable. Assume strong duality holds and that there are primal and dual optimal solutions x^* and λ^*. Show that

$$\nabla f_0(x^*)^\mathsf{T}(x - x^*) \geq 0$$

   for all feasible x.

3. (a) (10 points) Consider the function

$$F(x_1, x_2) = \log(x_1^2 + x_2^2)$$

   where we view addition, taking the square, and the logarithm as elementary functions. Show how $\nabla F(2, -4)$ is computed using reverse-mode automatic differentiation by drawing the appropriate diagrams.

   (b) (5 points) Explain why we usually use nonlinear activation functions in neural networks.

   (c) (5 points) Explain why we usually do not use Newton's method to train a neural network. overfitting second-derivative functions

4. In this exercise, we consider the primal-dual interior point method as discussed in class for the optimization problem

$$\begin{aligned} \text{minimize} \;\; & x_1 + x_2^2 \\ \text{subject to} \;\; & 1 - x_1 - x_2 \leq 0. \end{aligned}$$

   (a) (5 points) Recall that for a general minimization problem of the form

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq 0, \quad i = 1, \dots, m, \\ & Ax = b, \end{aligned}$$

   the residual vector is

$$r_t(x, \lambda, \nu) = \begin{pmatrix} \nabla f_0(x) + Df(x)^{\mathsf{T}} \lambda + A^{\mathsf{T}} \nu \\ -\operatorname{diag}(\lambda) f(x) - t1 \\ Ax - b \end{pmatrix}.$$

   Write down the residual vector for the specific optimization problem in this exercise.

   (b) (15 points) Suppose the current primal-dual iterate is (x, λ) with $x = (1, 1)$ and $\lambda = 1$. Compute the corresponding surrogate duality gap $\widehat{\eta}(x, \lambda)$ and primal-dual search direction $(\Delta x, \Delta \lambda)$.

5. Fix $\gamma > 0$ and consider the primal SVM problem

$$\begin{aligned} \text{minimize} \;\; & \gamma 1^{\mathsf{T}} u + \|a\|_2^2 \\ \text{subject to} \;\; & y_i(a^{\mathsf{T}} x_i - b) \geq 1 - u_i, \; i = 1, \dots, N \\ & u \geq 0 \end{aligned}$$

   in the variables $a \in \mathbb{R}^n$, $b \in \mathbb{R}$, and $u \in \mathbb{R}^N$.

   (a) (5 points) What is the computational bottleneck when applying the barrier method to solve this problem for large n and/or N?

   (b) (5 points) Consider the point $x^*(t)$ on the central path of the barrier method. Give an upper bound in terms of t, n, and N on how far the objective value of $x^*(t)$ is to the optimal objective value p^*.

   (c) (5 points) In the lecture we derived the following dual formulation:

$$\begin{aligned} \text{minimize} \;\; & \frac{1}{2} \alpha^{\mathsf{T}} Q \alpha - 1^{\mathsf{T}} \alpha \\ \text{subject to} \;\; & y^{\mathsf{T}} \alpha = 0 \\ & 0 \leq \alpha \leq \gamma 1. \end{aligned}$$

   Explain why we expect the optimal solution to this dual problem to be sparse when N is much larger than n.

End of test