Homework on convex problems

To be handed in before March 15 AoE (included)

1 Convex sets

Exercise 1: Is this convex? Using the definition of convex sets, discuss whether the following sets are convex:

a) $\mathcal{A} = \{x \in \mathbb{R}^n | Ax = b\}$ with $A \in \mathbb{R}^{n \times n}, b \in \mathbb{R}^n$

b) $\mathcal{S} = \{x \in \mathbb{R}^n | \|x - x_c\|_2 = r\}$ with $x_c \in \mathbb{R}^n, r > 0$

c) $\mathcal{H} = \{x \in \mathbb{R}^n | \|x - x_1\|_2 \leq \|x - x_2\|_2\}$ with $x_1, x_2 \in \mathbb{R}^n$

Exercise 2: Set of polynomial coefficients Show that $\mathcal{S} = \{x \in \mathcal{R}^n | |p_x(t)| \leq 1 \text{ for } |t| \leq 1\}$ with $p_x(t) = x_1 t + ... + x_n t^n$ is convex.

2 Convex functions

Exercise 3: Is this convex? Using the definition of convex functions, discuss whether the following functions are convex:

a) $f : \mathbb{R}^n \to \mathbb{R}, x \mapsto \|x\|_2$

b) $f : \mathbb{R} \to \mathbb{R}, x \mapsto x^3$

Exercise 4: Sum of biggest absolute value coordinates Let $f : \mathbb{R}^n \to \mathbb{R}, x \mapsto \sum_{i=1}^r |x|_{[i]}$ with $|x|$ the elementwise absolute value of x, and $|x|_{[i]}$ the i-th biggest coordinate of $|x|$. Is f convex?

Exercise 5: Epigraph and convex functions Show that f: $\mathbb{R}^n \to \mathbb{R}$ is convex if and only if its epigraph is convex.

3 Convex problems

Exercise 6: Convex problems Are these problems convex? Can you find an equivalent convex problem?

a)
$$\begin{array}{ll} \text{Min} & x_1^2 + x_2^2 \\ \text{s.t.} & \frac{x_1}{1+x_2^2} \leq 0 \\ & (x_1 + x_2)^2 = 0 \end{array}$$

b)
$$\begin{array}{ll} \text{Min} & \frac{x_1}{x_2} + \frac{x_3}{x_1} \\ \text{s.t.} & \frac{x_2}{x_3} + x_1 \leq 1 \\ & (x_1 + x_2)^2 = 0 \end{array}$$

Exercise 7: Solving Linear Problems Solve the following problem:

a)
$$\begin{array}{ll} \text{Min} & -3x_1 + x_2 \\ \text{s.t.} & x_1 + x_2 \leq 5 \\ & 2x_1 + x_2 \leq 8 \\ & x_1, x_2 \geq 0 \end{array}$$

4 Duality

Exercise 8: Dual analysis of a simple example We consider the following optimization problem over $x \in \mathbb{R}$:
$$\begin{array}{ll} \text{Min} & x^2 + 1 \\ \text{s.t.} & (x-2)(x-4) \leq 0 \end{array}$$

a) What is the solution of this problem?

b) What is the dual problem?

c) Compare the solutions of the dual and primal problems. Does strong duality hold?

d) Could you have known this without computing the solutions?

Exercise 9: Let's look at a kernel SVM! The general Kernel SVM problem is a simple regularized minimization problem over $f \in \mathcal{H}$ ($\mathcal{H}$ a RKHS with the positive definite kernel k, with labeled samples (x_i, y_i) drawn over the corresponding set) that can be written as $\text{Min}_f \frac{1}{n} \sum_{i=1}^n \phi(y_i f(x_i)) + \lambda \|f\|_2$ where $\phi(u) = max(1-u, 0)$ is the hinge loss.

a) Show that the primal kernel SVM problem can be rewritten a minimization over $\alpha, \xi \in \mathbb{R}^n$:
$$\begin{array}{ll} \text{Min}_{\alpha \in \mathbb{R}^n, \xi \in \mathbb{R}^n} & \frac{1}{n} \sum_{i=1}^n \xi_i + \lambda \alpha^T K \alpha \\ \text{s.t.} & \forall i, \xi_i + y_i \sum_{j=1}^n \alpha_j k(x_j, x_i) - 1 \geq 0 \\ & \forall i, \xi_i \geq 0 \end{array}$$
where K is the kernel's similarity matrix ($K_{i,j} = k(x_i, x_j)$).

Hint: Recall the representer theorem tells you the solution over a RKHS satisfies $f(x) = \sum_{i=1}^n \alpha_i k(x_i, x)$ for some $\alpha \in \mathbb{R}^n$

b) Show the Lagrangian can be written

$$\mathcal{L}(\alpha, \xi, \mu, \nu) = \frac{\xi^T \mathbf{1}}{n} + \lambda \alpha^T K \alpha - (diag(y)\mu)^T K \alpha - (\mu + \nu)^T \xi + \mu^T \mathbf{1}$$

in matricial form ($\mathbf{1}$ is the vector with all coordinates equal to 1).

Note: μ, ν are both Lagrange multipliers for inequality constraints here. You might want to start expressing the Lagrangian with respect to the μ_i and ν_i, and factorize afterwards.

c) Show the dual function can be written as

$$g(\mu, \nu) = \mu^T \mathbf{1} - \frac{1}{4\lambda} \mu^T diag(y) K diag(y) \mu \text{ if } \mu + \nu = \frac{\mathbf{1}}{n}, -\infty \text{ otherwise}$$

Hint: Minimize over α, then ξ

d) Show the dual problem can be written as

$$\begin{array}{ll} Max_{\alpha \in \mathbb{R}^n} & 2\alpha^T y - \alpha^T K \alpha \\ s.t. & 0 \leq y_i \alpha_i \leq \frac{1}{2\lambda n} \end{array}$$

e) Show the KKT lead to the following conditions

$$\forall i, \alpha_i [y_i f(x_i) + \xi_i - 1] = 0, (\alpha_i - \frac{y_i}{2\lambda n}) \xi_i = 0$$

Note: Start by writing out the KKT normally, then express the lagrange multipliers as a function of α. Remember you can remove non-null terms from a product equal to 0.

f) Analyze these conditions (e.g. what happens when $\alpha_i = 0$?). What is a support vector? Why is it very efficient to optimize on α?

g) Traditional kernel SVM solvers like Sequential Minimal Optimization (SMO) operates on α (or on the Lagrange multipliers, the correspondence between the two being immediate). They find α_i that do not fit the KKT conditions and find an analytical form of α_i that does. This process is repeated until α fits the KKT (up to some margin of tolerance). Can you explain why this is a good way to solve the optimization problem?