Convex Optimization in Machine Learning and Computational Finance: Solutions to Exam

Michel Baes, Patrick Cheridito

February 8, 2019

Question 1

a) (2 pts) The set Q is convex. For every $k \geq 0$, we denote by Q'_k the set $\bigcap_{i\geq k} Q_i$; as an intersection of convex sets, Q'_k is convex itself. Also, $Q'_k \subseteq Q'_{k+1}$ for every $k \geq 0$.

Let $x, y \in Q$ and $\lambda \in [0,1]$. We need to verify that $z = x\lambda + (1-\lambda)y \in Q$. The point x belongs to a set Q'_{k_x} and y to a set Q'_{k_y} for some numbers k_x, k_y. Let $N = \max\{k_x, k_y\}$. Then x and y both belong to Q'_N, because $Q'_{k_x}, Q'_{k_y} \subseteq Q'_N$. Since Q'_N is convex, $z \in Q'_N$. Hence $z \in Q$ and Q is convex.

b) (2 pts) The set Q is not convex. Consider two disjoint balls B_1 and B_2 and define $Q_k = B_1$ for k odd and $Q_k = B_2$ for k even. Then $\bigcap_{i\geq k} Q_i = B_1 \cup B_2$ for every k. Hence $Q = B_1 \cup B_2$, which is not convex.

c) (2pts) A set $Q \subseteq \mathbb{R}^n$ is semidefinite representable iff there exists a linear function $\mathcal{A} : \mathbb{R}^{n+m} \to \mathbb{S}^N_+$ and a symetric matrix B for which $x \in Q$ iff there exists $y \in \mathbb{R}^m$ for which $\mathcal{A}\begin{pmatrix} x \\ y \end{pmatrix} + B \in \mathbb{S}^N_+$.

d) (2pts) The set Q is SDr. Let $Q_1 \subseteq \mathbb{R}^n$, $Q_2 \subseteq \mathbb{R}^m$ be two semidefinite representable sets, and let $\mathcal{A}_1 : \mathbb{R}^{n+m_1} \to \mathbb{S}^{N_1}_+$, $B_1 \in \mathbb{S}^{N_1}$, $\mathcal{A}_2 : \mathbb{R}^{m+m_2} \to \mathbb{S}^{N_2}_+$, $B_2 \in \mathbb{S}^{N_2}$ such that $x_i \in Q_i$ iff there exists $y_i \in \mathbb{R}^{m_i}$ for which $\mathcal{A}_i\begin{pmatrix} x_i \\ y_i \end{pmatrix} + B_i \in \mathbb{S}^{N_i}_+$ for $i = 1, 2$.

To show that $Q_1 \times Q_2$ is an SDr set, it suffices to define:

$$\mathcal{A}_3\begin{pmatrix} x_1 \\ x_2 \\ y_1 \\ y_2 \end{pmatrix} := \begin{pmatrix} \mathcal{A}_1\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} & 0 \\ 0 & \mathcal{A}_2\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} \end{pmatrix}, \qquad B_3 := \begin{pmatrix} B_1 & 0 \\ 0 & B_2 \end{pmatrix}.$$

Indeed, $(x_1; x_2) \in Q_1 \times Q_2$ iff there exists $(y_1; y_2) \in \mathbb{R}^{m_1+m_2}$ for which $\mathcal{A}_3(x_1; x_2; y_1; y_2) + B_3$ is positive semidefinite. This block-diagonal $N_1 + N_2$ by $N_1 + N_2$ matrix is positive semidefinite iff each of its diagonal block is positive semidefinite, that is iff $x_1 \in Q_1$ and $x_2 \in Q_2$.

e) (2pts) The set Q is SDr. With the same notation as above (and $m = n$), it suffices to verify that if $Q_1 \times Q_2$ if SDr, then $Q_1 + Q_2$ is also SDr, as the former was already verified in the previous item. Note that $z \in Q_1 + Q_2$ iff there exists $s_1 \in Q_1$, $s_2 \in Q_2$ for which $z = s_1 + s_2$; this equality constraint can be represented in an SDr form as:

$$\begin{pmatrix} \operatorname{diag}(z - s_1 - s_2) & 0 \\ 0 & \operatorname{diag}(s_1 + s_2 - z) \end{pmatrix} \in \mathbb{S}^{2n}_+.$$

To sum up, $Q_1 + Q_2$ is SDr as it admits the representation: $z \in Q_1 + Q_2$ iff there exists $(s_1; s_2; y_1; y_2) \in \mathbb{R}^{2n+m_1+m_2}$ for which:

$$\begin{pmatrix} \operatorname{diag}(z - s_1 - s_2) & 0 & 0 \\ 0 & \operatorname{diag}(s_1 + s_2 - z) & 0 \\ 0 & 0 & \mathcal{A}_3 \begin{pmatrix} s_1 \\ s_2 \\ y_1 \\ y_2 \end{pmatrix} \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & B_3 \end{pmatrix} \in \mathbb{S}^{2n+N_1+N_2}_+.$$

Question 2

a) (2pts) The function f is convex as a supremum of a collection of linear functions $x \mapsto \sum_{k=1}^n x_k t^k$. (Alternatively, but more painstakingly, we can prove the convexity of f by using the definition directly).

b) (3pts) We can always replace the "inf" by a "min", because the minimum of $y \mapsto g(x, y)$ is attained over S for every $x \in \mathbb{R}^m$ in view of the compactness of S. A justification is not needed, but here is one for the sake of completeness. Let us fix $x \in \mathbb{R}^m$. For every $k > 0$ there exists $y_k \in S$ such that $f(x) \geq g(x, y_k) - \frac{1}{k}$. Since S is compact, there exists a subsequence $\{y_{k_j} : j > 1\}$ that converges to some $y^* \in S$. Then $g(x, y^*) \geq \inf_{y \in S} g(x, y) \geq \liminf_{j \geq 0} g(x, y_{k_j}) - \frac{1}{k_j} = g(x, y^*)$, since g is closed. Thus $f(x) = g(x, y^*)$.

To prove that f is convex, let $x, x' \in \mathbb{R}^m$ and $\lambda \in [0, 1]$. Let $y_x, y'_x \in S$ be such that $f(x) = g(x, y_x)$ and $f(x') = g(x', y'_x)$. Then

$$\begin{aligned} f(\lambda x + (1 - \lambda)x') &= \inf_{y \in S} g(\lambda x + (1 - \lambda)x', y) \leq g(\lambda x + (1 - \lambda)x', \lambda y_x + (1 - \lambda)y'_x) \\ &\leq \lambda g(x, y_x) + (1 - \lambda)g(x', y'_x) = \lambda f(x) + (1 - \lambda)f(x'). \end{aligned}$$

c) (2pts) The support function of S is $\sigma_S(s) := \sup_{x \in S} s^T x$. It has finite values because S is bounded. Indeed, as there is an $R > 0$ for which $S \subseteq B_2[0, R]$, then

$$\sigma_S(s) := \sup_{x \in S} s^T x \leq \sup_{x \in B_2[0,R]} s^T x = ||s||_2 R,$$

which is finite for every s.

d) (3pts) To compute the conjugate function of σ_S, different avenues can be taken.

- The simplest consists in using that $\sigma_S = \chi_S^*$, where χ_S is the characteristic function of S. Then $\sigma_S^* = \chi_S^{**}$, whose epigraph is the closed convex hull of the epigraph of χ_S, that is, the epigraph of $\chi_{\operatorname{cl}(\operatorname{conv}(S))}$.

- While the above is a statement given in the lecture that can be used as such by the students, some might want to compute explicitly what the conjugate of σ_S is.

  There are different ways to do it. For this first solution, it is necessary to note that $\sigma_S(Ms) = M\sigma_S(s)$ for every $M > 0$, as it is clear from the definition of σ_S.

  Suppose first that z is not in the closed convex hull Q of S. Then there exists a hyperplane that separates strictly (and in fact strongly because Q is compact, we we will not need that here) z from Q: there exists a nonzero $\hat{s}$ for which $\hat{s}^T z > \hat{s}^T x$ for every $x \in Q$, that is, $\hat{s}^T z > \sup_{x\in Q} \hat{s}^T x \geq \sup_{x\in S} \hat{s}^T x = \sigma_S(\hat{s})$. Then

  $$\sigma_S^*(z) = \sup_s s^T z - \sigma_S(s) \geq \sup_M M\hat{s}^T z - \sigma_S(M\hat{s}) \geq \sup_{M\geq 0} M\hat{s}^T z - M\sigma_S(\hat{s}) = +\infty.$$

  Let us now take a point $z \in Q$. Then for every s, we have that $\sup_{x\in S} s^T x \geq s^T z \geq \inf_{x\in S} s^T x$. In particular, $\sigma_S(s) \geq s^T x$ for every s, thus

  $$\sigma_S^*(z) = \sup_s s^T z - \sigma_S(s) \geq 0$$

  Taking $s = 0$, we see that $\sigma_S^*(z) = 0$.

- A possible way to simplify the reasoning above is to argue first that $\sigma_S = \sigma_Q$, then the use the standard $\chi_Q^{**} = \chi_Q$.

Question 3

a) (2pts) The existence of a solution can be deduced from an argument similar to 2.b): the feasible set is $Q_1 \times Q_2$, which is compact because Q_1 and Q_2 are both compact by assumption. Thus, there exists at least a solution (x^*, y^*).

   This solution is not necessarily unique. For instance (insert a picture), suppose that the first satellite is a unit cube shifted so that its center coincides with the point $(-2, 0, 0)$, and that the second satellite is a unit cube shifted so that its center coincides with the point $(2, 0, 0)$. Then their faces $(-1.5, x_1, x_2)$ and $(1.5, y_1, y_2)$, $1/2 \leq x_i, y_i \leq 1/2$ are parallel. Any point of these faces with $y_1 = x_1$ and $y_2 = x_2$ solves the problem.

b) (4pts) Observe that the original problem is convex. With $K := \mathbb{R}_+^{m_1} \times \mathbb{R}_+^{m_2}$, we shall take for $\mathcal{F}$ the set of affine functions $x \mapsto u^T x + u_0$ that are nondecreasing with respect to the ordering induced by K, i.e., for which $u \in K^*$. Note that $K^* = K$, i.e. u must have all its component nonnegative.

   Since the original problem has only six variables, the dual problem will only have at most six constraints, in addition to the nonnegativity of the dual variables u: the dual problem will satisfy the requirements of our solver.

   The dual will be exact, that is, there will not be any duality gap: indeed, the original problem is convex and the constraints $g(x, y) \succeq_K b$ are convex (g is actually a linear function). Since each satellite has an interior point, Slater's conditions are satisfied, and there is no duality gap between the primal and the dual.

   Let us write the dual of the original problem[1]:

   $$\max\{u_1^T b_1 + u_2^T b_2 + u_0 : ||x - y||_2^2 \geq u_1^T A_1 x + u_2^T A_2 y + u_0 \text{ for all } x, y, \quad u_1, u_2 \succeq 0\}$$

[1]one can also use the Lagrangian, it will be actually faster.

$$= \max\{u_1^T b_1 + u_2^T b_2 + \min_{x,y} ||x-y||_2^2 - u_1^T A_1 x - u_2^T A_2 y : u_1, u_2 \succeq 0\}$$

Let us fix $u_1, u_2 \succeq 0$ and let us focus on the subproblem. Taking $x = y$, we can see that $A_1^T u_1 + A_2^T u_2 = 0$ for otherwise the infimum equals $-\infty$.

The optimality conditions for the subproblem read:

$$2(x_u - y_u) = A_1^T u_1, \qquad 2(y_u - x_u) = A_2^T u_2.$$

(We see that the necessary constraint $A_1^T u_1 + A_2^T u_2 = 0$ is satisfied.) As $y_u = A_2^T u_2/2 + x_u$, the subproblem becomes:

$$||A_2^T u_2/2||_2^2 - (u_1^T A_1 + u_2^T A_2) x_u - u_2^T A_2 A_2^T u_2/2 = -||A_2^T u_2||_2^2/4 - 0.$$

The dual problem then takes the form:

$$\max\{-\frac{1}{4}||A_2^T u_2||_2^2 + u_1^T b_1 + u_2^T b_2 : A_1^T u_1 + A_2^T u_2 = 0,\ u_1 \succeq_{\mathbb{R}^{m_1}} 0,\ u_2 \succeq_{\mathbb{R}^{m_2}} 0\}.$$

To prove that the duality gap is null between both problems, we can use Slater's conditions: they are satisfied by a point $(\hat{x}, \hat{y})$ in the interior of $Q_1 \times Q_2$.

c) (2pts) The KKT conditions for the original problem read:

$$2(x^* - y^*) = A_1^T u_1^*, \qquad 2(y^* - x^*) = A_2^T u_2^*, \tag{1}$$

$$A_1 x^* \succeq b_1, \qquad A_2 y^* \succeq b_2, \qquad u_1^* \succeq 0, \qquad u_2^* \succeq 0$$

$$[u_1^*]_i [A_1 x^* - b_1]_i = 0, \qquad [u_2^*]_j [A_2 y^* - b_2]_j = 0 \tag{2}$$

for $1 \leq i \leq m_1$, $1 \leq j \leq m_2$. Note that u_1^*, u_2^* are solutions of the dual problem (and in particular $A_1^T u_1^* + A_2^T u_2^* = 0$).

The KKT for the dual problem are:

$$-\frac{1}{2} A_2 A_2^T u_2^* + b_2 - A_2 \mu^* + \lambda_2^* = 0, \qquad b_1 - A_1 \mu^* + \lambda_1^* = 0 \tag{3}$$

$$A_1^T u_1^* + A_2^T u_2^* = 0, \qquad u_1^* \succeq 0, \qquad u_2^* \succeq 0, \qquad \lambda_1^* \succeq 0, \qquad \lambda_2^* \succeq 0,$$

$$[u_1^*]_i [\lambda_1^*]_i = 0, \qquad [u_2^*]_j [\lambda_2^*]_j = 0 \tag{4}$$

for $1 \leq i \leq m_1$, $1 \leq j \leq m_2$. (Here, the multipliers μ^* are unconstrained.) Denoting $\nu^* := \mu^* + \frac{1}{2} A_2^T u_2^*$, the conditions (3) can be simplified into:

$$-A_2 \nu^* + b_2 + \lambda_2^* = 0, \qquad b_1 - A_1 \mu^* + \lambda_1^* = 0. \tag{5}$$

We can eliminate λ_1^* and λ_2^* to get:

$$A_1^T u_1^* + A_2^T u_2^* = 0, \qquad u_1^* \succeq 0, \qquad u_2^* \succeq 0, \qquad A_1 \mu^* \succeq b_1, \qquad A_2 \nu^* \succeq b_2,$$

$$[u_1^*]_i [A_1 \mu^* - b_1]_i = 0, \qquad [u_2^*]_j [A_2 \nu^* - b_2]_j = 0 \tag{6}$$

for $1 \leq i \leq m_1$, $1 \leq j \leq m_2$. With the definition of ν^*, that is $2(\nu^* - \mu^*) = A_2^T u_2^*$, we see that the KKT constraints for the primal problem are identical than that for the dual problem. As they should.

d) (2pts) Computing y^* from u^* and x^* can be easily done using the KKT condition (1). Using a solution $u^* = (u_1^*, u_2^*)$ to the dual problem to compute x^* is a bit more complicated in general. Observe first that if $u_1^* = 0$ or $u_2^* = 0$, then $x^* = y^*$, and the satellites touch each other, a contradiction. Hence u_1^* and $u_2^* = 0$ must have some non-zero components. Let I_1 be the nonzero components i of u_1^* and I_2 be those of u_2^*. According to the complementarity conditions (2), $[A_1 x^*]_i = [b_1]_i$ for $i \in I_1$ and, substituting y^* with $x^* + A_2^T u_2^*/2$ by (1), $[A_2 x^*]_j = [b_2]_j - [A_2 A_2^T u_2^*/2]_j$ for $j \in I_2$. If these equations form a non-degenerated system, we are done: we can solve it for x^*, and the feasibility comes from the existence of a solution proved in 2a.

Otherwise, our problem has not a unique solution and things get (even) more complicated. We only know that x^* belongs to an affine space M. To get a solution to our problem, we can for instance rewrite the original problem as $\min\{||x||^2 : x \in M, x \in Q_1, y = x + A_2^T u_2^*/2 \in Q_2\}$. We can again write a strong dual for this problem (again by Slater's conditions) and proceed as above. The condition $x \in M$ ensures that the x^* we find is optimal for the original problem.

Question 4

a) (2pts) Newton's Method requires a function f that is twice differentiable. Given a starting point x_0, Newton's method constructs the sequence $x_{k+1} = x_k - f''(x_k)^{-1} f'(x_k)$ as long as $f''(x_k)$ is an invertible matrix.

b) (2pts) Kantorovitch's Theorem states the following:

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice differentiable function minimized in x^* for which:

(a) $f''(x^*) \succeq lI$ for some $l > 0$

(b) there exists $M > 0$ for which $|||f''(x) - f''(y)||| \leq M||x - y||_2$ for every $x, y \in \mathbb{R}^n$. Here, the matrix norm $||| \cdot |||$ is the norm induced by $|| \cdot ||_2$.

Suppose that $||x_0 - x^*||_2 \leq \frac{2l}{3M}$. Then the Newton method is well defined and

$$||x_{k+1} - x^*||_2 \leq \frac{3M||x_2 - x^*||_2^2}{2(l - M||x_2 - x^*||_2)}.$$

Hence, the convergence of Newton's method is quadratic. (Even though the convergence result was not required)

c) Newton's method, as given above, can only be applied as such for unconstrained problems. When the problem is convex and constrained, we must distinguish two cases

Linear equality constraints (3pts) Let us recall the Newton's Method for unconstrained minimization of a function f can be seen as a method to solve the optimality condition equation system $f'(x^*) = 0$ by linearizing these equations around a point x_k and taking for x_{k+1} the exact solution of the linearized system.

Suppose that we need to solve $\min\{f(x) : Ax = b\}$, where f is a twice differentiable convex function and the matrix A has full row rank (that is, no constraint is redundant). The KKT conditions of this problem read:

$$f'(x^*) + A^T u^* = 0, \quad Ax^* - b = 0.$$

Linearizing these conditions around a point (x_k, u_k), we get:

$$f''(x_k)(x - x_k) + A^T(u - u_k) + f'(x_k) + A^T u_k = 0, \quad A(x - x_k) + Ax_k - b = 0,$$

which gives us:

$$\begin{pmatrix} f''(x_k) & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x \\ u \end{pmatrix} = \begin{pmatrix} f''(x_k)x_k - f'(x_k) \\ b \end{pmatrix}.$$

As the matrix A has full row rank (no redundant constraint), the matrix of the above linear system is invertible iff $f''(x_k)$ is itself invertible. Hence, Newton's method on the constrained optimization problem is well defined iff it is well-defined for the constrained problem.

Incidentally, this scheme coincides with projecting $x_k - f''(x_k)^{-1}f'(x_k)$ on the affine space $S = \{x \in \mathbb{R}^n : Ax = b\}$. As this was not in the course, a careful justification of this assertion is in order.

First, we need to prove that:

$$\arg\min_{y \in S} ||y - x_k + f''(x_k)^{-1}f'(x_k)||_2 = \arg\min_{y \in S} f(x_k) + f'(x_k)^T(y - x_k) + \frac{1}{2}||x - x_k||^2,$$

which can be done by comparing the KKT condition of either problem.

General convex constraints (3pts) Consider the problem $\min\{f(x) : x \in Q\}$, where Q is a set with a nonempty interior. This problem is equivalent to the unconstrained problem $\min f(x) + \chi_Q(x)$. However, the non-differentiability of χ_Q, especially at the boundary of Q forbids us to use Newton's Method to minimize this function. Suppose that we have a barrier F_Q for the set Q, that is, a strongly convex function (even though in some examples a strictly convex function might be used instead of a strongly convex one; the idea is to ensure that Newton's method can be applied to the resulting problem) that is twice differentiable on its domain int$Q$, that is bounded from below, and for which every sequence $\{x_n : n \geq 0\} \subseteq Q$ that converges to ∂Q is such that $\lim_{n\to\infty} F_Q(x_n) = +\infty$. Then, for any positive parameter $\mu > 0$, we consider the problem:

$$\min f(x) + \mu F_Q(x).$$

Since this problem is convex, unconstrained, and with an objective function that is twice differentiable, Newton's Method can be applied, provided that the Hessian is invertible. Note that the above optimization problem is only an approximation of the original problem $\min f(x) + \chi_Q(x)$.

It can be proven that the optimum $x^*(\mu)$ of the approximated problem converges to an optimum x^* of the original problem.

Further discussions on Interior-Point Methods are welcomed, but not indispensable.