Due to the current COVID-19 pandemic, the final exam will be a take home final. In addition, your final course grade will be computed in a manner that deviates from the description given in the course syllabus. The final exam is now worth 20% of your grade and the midterm exam is worth 45% of your grade:

$$\text{quiz } 350 \;+\; \text{midterm } 450 \;+\; \text{final } 200 \;=\; 1000 \text{ points.}$$

No person-to-person contact is permitted for either retrieving or submitting the final exam.

I will email all students a PDF of the final exam at 10am on Sunday morning March 15. A PDF, scan, or photocopy of your worked exam is to be emailed back to me by 10am on Monday March 16.

The exam will be composed of 3 sections:
I Quadratic Optimization and Least Squares
II Optimality Conditions
III Convex Optimization

The number of questions and points for each section varies with an emphasis on Section III Convex Optimization since this is the material we studied most intensively since the midterm exam. A more detailed description of the material you will be tested on in each area is given below. This will be followed by sample questions for each area. The number of sample questions greatly exceeds the number of questions on the exam.

I Quadratic Optimization and Least Squares

The questions for this section are of three types: theory, linear algebra and computation. Theory questions refer to the basic theorems and properties of quadratic optimization and linear least squares problems. In particular, you should know how quadratic optimization and linear least squares problems are related, when they differ, when they are the same, and when solutions exist, what is the structure of the solutions sets, and when are solutions unique. The linear algebra questions refer to the linear algebraic properties of the objective functions and the matrix factorizations associated with computing solutions. The computational questions require you to compute the solution set to concrete problems of this type which includes quadratic optimization problems with affine constraints.

II Optimality Conditions

These question relate to first- and second-order optimality conditions for both constrained and unconstrained problems with particular emphasis on NLP problems in the constrained case. You will need to know what a tangent cone is and be able to compute one (under the assumption of regularity). You will **not** need to work with constraints qualifications other than to know that they imply regularity. You will be asked to to answer questions on the theory (statements of theorems), testing optimality, computing solutions, and computing KKT points.

III Convex Optimization

For these questions you will need to know the definitions of convex functions and sets and be able to show that a given function or set is convex. You will need to know the optimality conditions for convex optimization problems especially for convex NLP. You will need to be able to compute KKT points. You will also need to know the rudiments of Lagrangian duality and how to compute the Lagrangian dual for simple problems.

Sample Questions

I Quadratic Optimization and Least Squares

These questions were taken from the midterm guide whose solutions were provided.)

II Optimality Conditions

1. State the first- and second-order conditions for optimality for the following three problems:

   (a) Linear least squares: $\min_{x\in\mathbb{R}^n} \frac{1}{2}\|Ax-b\|_2^2$, where $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$.

   Solution
   First-order necessary and sufficient condition: $\bar{x}$ is a global (local) solution if and only if $A^T(A\bar{x} - b) = 0$. Moreover, such an $\bar{x}$ always exists.

   (b) Quadratic Optimization: $\min_{x\in\mathbb{R}^n} \frac{1}{2}x^TQx + g^Tx$, where $Q \in \mathbb{R}^{n\times n}$ is symmetric and $g \in \mathbb{R}^n$.

   Solution
   First-order: If $\bar{x}$ is a local solution, then $Q\bar{x} + g = 0$.
   Second-order: (Necessary) If $\bar{x}$ is a local solution, then $Q\bar{x} + g = 0$ and $Q \in \mathbb{S}^n_+$.
   (Sufficient) If $Q\bar{x} + g = 0$ and $Q \in \mathbb{S}^n_{++}$, then there is $\alpha > 0$, such that $f(x) \geq f(\bar{x}) + \alpha\|x-\bar{x}\|_2^2$, where $f(x) = \frac{1}{2}x^TQx + g^Tx$.

   (c) Quadratic Optimization with Affine Constraints: $\min_{x\in\mathbb{R}^n}\left\{\frac{1}{2}x^TQx + g^Tx \mid Ax = b\right\}$, where $Q \in \mathbb{R}^{n\times n}$ is symmetric, $A \in \mathbb{R}^{m\times n}$, $g \in \mathbb{R}^n$, and $b \in \mathbb{R}^m$.

   Solution
   First-order: If $\bar{x}$ is a local solution, then there is a vector $\bar{y} \in \mathbb{R}^m$ such that $(\bar{x}, \bar{y})$ solve the linear system

   $$\begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -g \\ b \end{pmatrix}.$$

   Second-order: (Necessary) Q is psd on $\mathrm{Null}(A)$, i.e., $z^TQz \geq 0$ for all $z \in \mathrm{Null}(A)$.
   (Sufficient) Q is pd on $\mathrm{Null}(A)$, i.e., $z^TQz > 0$ for all $z \in \mathrm{Null}(A) \setminus \{0\}$.

2. Consider the nonlinear programming (NLP) problem

   $$\min_{x\in\Omega} f(x), \tag{$\spadesuit$}$$

   where

   $$\Omega := \{x : c_i(x) \leq 0, i = 1, \ldots, s, c_i(x) = 0, i = s+1, \ldots, m\}. \tag{$\clubsuit$}$$

   (a) State the first-order optimally conditions for NLP.

   Solution
   If $\bar{x}$ is a local solution to NLP, then there exists $\bar{y} \in \mathbb{R}^m$ such that
   (i) $\bar{x} \in \Omega$ (primal feasibility)
   (ii) $0 \leq \bar{y}_i,\ i = 1, 2, \ldots, s$, (dual feasibility)
   (iii) $\bar{y}_i c_i(\bar{x}) = 0,\ i = 1, 2, \ldots, s$, (complementarity)
   (iv) $0 = \nabla_x L(\bar{x}, \bar{y})$, (stationarity of the Lagrangian)
   where $L(x, y) := f(x) + \langle y, c(x)\rangle$ is the Lagrangian for NLP.

   (b) What does it mean for $(\bar{x}, \bar{y}) \in \mathbb{R}, \times\mathbb{R}^m$ to be KKT pair for NLP?

   Solution: $(\bar{x}, \bar{y})$ satisfy (i)-(iv) above.

3. A critical point of a function $f : \mathbb{R}^n \to \mathbb{R}$ is any point x at which $\nabla f(x) = 0$. Compute all of the critical points of the following functions. Classify the critical points as local/global minima/maxima/saddle points. If no critical point exists, explain why.

(a) $f(x) = x_1^2 - 4x_1 + 2x_2^2 + 7$

Solution: This function is fully separable, $f(x) = f_1(x_1) + f_2(x_2)$, where $f_1(x_1) = x_1^2 - 4x_1$ and $f_2(x_2) = 2x_2^2 + 7$. Hence we need only optimize f_1 and f_2 separately.

$$f_1'(x_1) = 2x_1 - 4,\ f"_1(x_1) = 2,\ f_2'(x_2) = 4x_2,\ f"_2(x_2) = 4\ .$$

Hence the unique critical point of f_1 is $x_1 = 2$ which is a global minimizer since f_1 is a parabola with positive curvature. Similarly, the unique global minimizer of f_2 is $x_2 = 0$. Therefore, the unique global minimizer of f is $(x_1, x_2) = (2, 1)$, and f has no other critical points.

(b) $f(x) = \exp(-\|x\|^2)$

Solution: Write f as

$$f(x) = \mathsf{e}^{-x_1^2}\mathsf{e}^{-x_2^2}\cdots\mathsf{e}^{-x_n^2} = \prod_{j=1}^{n} \mathsf{e}^{-x_j^2},$$

then it is easily seen that

$$\frac{\partial f}{\partial x_k}(x) = -2x_k f(x), \quad j = 1, 2, \ldots, n.$$

Hence

$$\nabla f(x) = -2f(x)x \quad \text{and} \quad \nabla^2 f(x) = 2f(x)\left[2xx^T - I\right]\ .$$

An expression of the form xy^T $(x, y \in \mathbb{R}^n)$, as appears above, is called the *outer product* of x and y. It is an $n \times n$ matrix whose ijth entry is $x_i y_j$. In particular, we have $x^T y = \text{trace}(xy^T)$.
Clearly, $x = 0$ is the unique critical point of f and this critical point is a local maximum of f since the Hessian of f at the origin is $-2I$ which is negative definite. Indeed, the origin is a global maximizer since $f(0) = 1$ which is the largest possible value of e^{ξ} for $\xi < 0$.

(c) $f(x) = x_1^2 - 2x_1x_2 + \frac{1}{3}x_2^3 - 8x_2$

Solution:

$$\nabla f(x) = \begin{bmatrix} 2(x_1 - x_2) \\ x_2^2 - 2x_1 - 8 \end{bmatrix} \quad \text{and} \quad \nabla^2 f(x) = \begin{bmatrix} 2 & -2 \\ -2 & 2x_2 \end{bmatrix}.$$

Compute the critical points by setting $\nabla f(x) = 0$. Setting $\partial f(x)/\partial x_1 = 0$ gives $x_1 = x_2$. Plug this into the equation $\partial f(x)/\partial x_2 = 0$ to get $0 = x_2^2 - 2x_2 - 8 = (x_2 - 4)(x_2 + 2)$. This gives 2 critical points

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 4 \\ 4 \end{pmatrix}, \begin{pmatrix} -2 \\ -2 \end{pmatrix},$$

with

$$\nabla^2 f(4,4) = \begin{bmatrix} 2 & -2 \\ -2 & 8 \end{bmatrix} \quad \text{and} \quad \nabla^2 f(-2,-2) = \begin{bmatrix} 2 & -2 \\ -2 & -4 \end{bmatrix}.$$

It is easily shown that $\nabla^2 f(4,4)$ is positive definite and that $\nabla^2 f(-2,-2)$ has one positive and one negative eigenvalue. Hence $(4,4)$ is a local minimizer and $(-2,-2)$ is a saddle point. There are no global maximizers or minimizers since $f(0, x_2) = \frac{1}{3}x_2^3 - 8x_2$ which goes to $+\infty$ as $x_2 \uparrow +\infty$ and goes to $-\infty$ as $x_2 \downarrow -\infty$.

(d) $f(x) = (2x_1 - x_2)^2 + (x_2 - x_3)^2 + (x_3 - 1)^2$

Solution: Since f is a sum of squares, the smallest value f can take is zero. Hence any point $\bar{x}$ at which $f(\bar{x}) = 0$ is necessarily a global minimizer. To make $f(x) = 0$ each of the three squared terms in f must be zero. From the third term we get $x_3 = 1$. The second term gives $x_2 = x_3 = 1$, and the first term gives $2x_1 = x_2 = 1$ so $x_2 = 1/2$. Consequently, $(x_1, x_2, x_3) = (1/2, 1, 1)$ is the unique global minimizer of f.

Note that the function f is a convex function since we can write it in the form of a linear least squares objective:

$$f(x) = \|Ax - b\|_2^2, \quad \text{where } b = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \text{ and } A = \begin{bmatrix} 2 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}.$$

Hence $(x_1, x_2, x_3) = (1/2, 1, 1)$ is also the unique critical point.

(e) $f(x) = x_1^4 + 16x_1x_2 + x_2^4$

Solution:

$$\nabla f(x) = \begin{pmatrix} 4x_1^3 + 16x_2 \\ 4x_2^3 + 16x_1 \end{pmatrix} \qquad \nabla^2 f(x) = \begin{bmatrix} 12x_1^2 & 16 \\ 16 & 12x_2^2 \end{bmatrix}$$

Hence x is a critical point if

$$\begin{aligned} 0 &= 4x_1^3 + 16x_2 \\ 0 &= 4x_2^3 + 16x_1 \ . \end{aligned}$$

Multiply the first equation by x_1 and the second by x_2 and subtract to get the equation

$$0 = 4[x_1^4 - x_2^4] = 4(x_1^2 + x_2^2)(x_1 + x_2)(x_1 - x_2).$$

This implies that either $x_1 = x_2$ or $x_1 = -x_2$. Plug this information into the first equation above to get

$$0 = 4x_1^3 \pm 16x_1 = 4x_1(x_1^2 \pm 4).$$

Therefore, the only possible critical points are

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ -2 \end{pmatrix}, \begin{pmatrix} -2 \\ 2 \end{pmatrix}, \begin{pmatrix} -2 \\ -2 \end{pmatrix}.$$

Since the gradient must be zero, x_1 and x_2 must have opposite sign. Plugging these vectors into the gradient, we see that only

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ -2 \end{pmatrix}, \begin{pmatrix} -2 \\ 2 \end{pmatrix},$$

are critical points. Plugging these into the Hessian, we see that $x = 0$ is a saddle point (the eigenvalues of $\nabla^2 f(0)$ are ± 16), and the other two critical points are local minimizers. Furthermore, f is coercive since

$$\begin{aligned} f(x) &= x_1^4 + 16x_1x_2 + x_2^4 \\ &\geq x_1^4 - 16|x_1|\,|x_2| + x_2^4 \\ &\geq \begin{cases} x_1^4 - 16|x_1|^2 + x_2^4 & \text{if } |x_1| \geq |x_2|, \\ x_1^4 - 16|x_2|^2 + x_2^4 & \text{if } |x_2| \geq |x_1|, \end{cases} \end{aligned}$$

f is coercive (i.e. the right hand side of this inequality necessarily diverges to $+\infty$ as $\|x\|$ goes to infinity). Hence the critical points $(x_1, x_2) = (2, -2),\ (-2, 2)$ are global minimizers.

(f) $f(x) = (1 - x_1)^2 + \sum_{j=1}^{n-1} 10^j (x_j - x_{j+1}^2)^2$ (The Rosenbrock function)

Solution: Again, f is a sum of squares so any point at which the function takes the value zero is necessarily a global minimizer. The first term indicates that we should take $x_1 = 1$. The second term requires $x_2 = \pm x_1 = \pm 1$. The third term has $x_2 = x_3^2$, so $x_2 \geq 0$ implying that $x_2 = 1$. Moreover, $x_3 = \pm 1$. Continuing in this way we get $1 = x_1 = x_2 = \cdots = x_{n-1}$ and $x_n = \pm 1$, so that there are two global minimizers. There is only one other critical point for this function: $x_1 = 1/11, 0 = x_2 = x_2, \ldots, x_n$. One can show that it is a saddle point.

(g) $f(x_1,\, x_2) := x_1^2 + 2x_2x_2 + x_2^2 - 8x_1 - 8x_2$

Solution: $f(x_1,\, x_2) := (x_1 + x_2)^2 - 8(x_1 + x_2) = z^2 - 8z$, where $z := x_1 + x_2$. As a function of z, $\hat{f}(z) := z^2 - 8z$ has $\hat{f}'(z) = 2z - 8$ and has a global minimizer at $z = 4$. Hence, $\{(x_1, x_2)\ |x_1 + x_2 = 4\}$ is the set of critical points for f and all critical points are global minimizers.

4. Compute the KKT points for the following problems.

Solution: Since all of these problems are convex, any KKT point is a global solution. In general, the solution procedure has the following steps:

Step 1 Consider the geometry of the problem and observe any obvious solution. In particular, if the global unconstrained optimal solution for the objective lies in the feasible region, then it is necessarily the solution to the constrained problem with all multipliers equal to zero.

Step 2 Write the equation for the Lagrangian $L(x, y)$. Two helpful tricks.

(a) I often separate the constraints into two groups, equalities and inequalities, writing a dual variable for each group, say $y = (u, v)$ where the u's are for the inequalities ($u \geq 0$) and the v's are for the equalities (free).

(b) Sometimes it is very useful to introduce a new primal variable for a complicated structure in the objective and then add the definition of the new variable as an equality constraint. For example, if the objective contains a composite term of the form $h(Ax - b)$, then one almost always defines a new variable $w = Ax - b$ adding this equation to the constraints, and then replace $h(Ax - b)$ by $h(w)$. Always use this trick for terms of the form $\frac{1}{2}\|Ax - b\|_2^2$ or, more generally, for any term of the form $\|Ax - b\|$ regardless of the norm involved.

Step 3 Write $0 = \nabla_x L(x, y)$. That is, take the derivative with respect to all of the primal variables. The goal here is to locate the minima of L with respect to the primal variables.

Step 4 Use the full set of KKT conditions to compute KKT points and the multipliers. In this process, one needs to either make an educated guess as to the activities at the solution (which inequality constraints are active), or to run through all possible cases of actives and non-actives. This step is the hard step and requires multiple tricks and years of experience. I will only give "easy" problems. But you will still struggle with them as they are still *very hard* for the beginner.

(a) $\min \left\{ \frac{1}{2}\|x\|_2^2 \ \middle| a^T x \leq \alpha \right\}$ where $a \in \mathbb{R}^n \setminus \{0\}$ and $\alpha \in \mathbb{R}$.

Solution: There are two cases to consider; $\alpha \geq 0$ and $\alpha < 0$. For $\alpha \geq 0$, $0 \in \left\{ x \ \middle| a^T x \leq \alpha \right\}$ so $\bar{x} = 0$ is the solution and $\bar{y} = 0$ is an associated KKT multiplier. For $\alpha < 0$, the Lagrangian is $L(x, y) = \frac{1}{2}\|x\|_2^2 + y(a^T x - \alpha)$ and at the solution $\bar{x}$ the constraint $a^T x \leq \alpha$ is active. Therefore the KKT pair solves the system

$$\begin{aligned} 0 &= x + ya = \nabla_x L(x, y) \\ \alpha &= a^T x. \end{aligned}$$

Hence $x = -ya$. Plugging this into the second equation gives $\alpha = -y\,\|a\|_2^2$ or $y = -\alpha/\,\|a\|_2^2 > 0$. Therefore the KKT pair is $(\bar{x}, \bar{y}) = (\frac{\alpha}{\|a\|_2^2}a,\ -\frac{\alpha}{\|a\|_2^2})$.

(b) $\min\left\{\frac{1}{2}\|x\|_2 \;\middle|\; a^T x \le \alpha\right\}$ where $a \in \mathbb{R}^n \setminus \{0\}$ and $\alpha \in \mathbb{R}$.

Solution: The solution and the two cases are the same as in the previous problem only the multiplier changes in the second case. For $\alpha \ge 0$, the solution is $\bar{x} = 0$ since $0 \in \left\{x \;\middle|\; a^T x \le \alpha\right\}$. In this case the multiplier is zero. Next assume that $\alpha < 0$ and the Lagrangian is $L(x, y) = \frac{1}{2}\|x\|_2 + y(a^T x - \alpha)$ with $\|x\|_2$ differentiable at the solution since it cannot be zero. Again the constraint is active at the solution. Hence the KKT conditions are

$$
\begin{aligned}
0 &= \frac{1}{2}\frac{x}{\|x\|_2} + ya = \nabla L(x, y) \\
\alpha &= a^T x.
\end{aligned}
$$

Consequently, $x = \mu a$ for some $\mu \le 0$ so that $\alpha = \mu\,\|a\|_2^2$, or $\mu = \alpha/\,\|a\|_2^2$ and $\bar{x} = \frac{\alpha}{\|a\|_2^2}a < 0$. Therefore,

$$
\frac{-a}{\|a\|_2} = \frac{\bar{x}}{\|\bar{x}\|_2} = -2\bar{y}a,
$$

so $\bar{y} = 1/(2\,\|a\|_2)$.

(c) $\min\left\{\frac{1}{2}\|x\|_2^2 \;\middle|\; a^T x = \alpha,\ b^T x \le \beta\right\}$ where $a, b \in \mathbb{R}^n \setminus \{0\}$ with a and b linearly independent and $\alpha, \beta \in \mathbb{R}$.

Solution: The Lagrangian is $L(x, \lambda, \mu) = \frac{1}{2}\|x\|_2^2 + \lambda(a^T x - \alpha) + \mu(b^T x - \beta)$, and the KKT conditions are

$$
\begin{aligned}
\alpha &= a^T x \\
\beta &\ge b^T x \\
0 &\le \mu \\
0 &= \mu(b^T x - \beta) \\
0 &= x + \lambda a + \mu b.
\end{aligned}
$$

We consider the cases $b^T x = \beta$ and $b^T x < \beta$ separately, computing KKT points for each case. First suppose $b^T x = \beta$ at the solution. This yields the equations

$$
\begin{aligned}
0 &= x + \lambda a + \mu b \\
\alpha &= a^T x \\
\beta &= b^T x.
\end{aligned}
$$

By multiplying the first equation through by a and then b we get the two equations $0 = \alpha + \lambda\,\|a\|_2^2 + \mu a^T b$ and $0 = \beta + \lambda a^T b + \mu\,\|b\|_2^2$. Since a and b are linearly independent $\|a\|_2\,\|b\|_2 > |a^T b|$ so that this system has the unique solution given by

$$
\begin{pmatrix} \hat{\lambda} \\ \hat{\mu} \end{pmatrix} = \frac{1}{\|a\|_2^2\,\|b\|_2^2 - \langle a, b\rangle^2}\begin{pmatrix} \beta\,\langle a, b\rangle - \alpha\,\|b\|_2^2 \\ \alpha\,\langle a, b\rangle - \beta\,\|a\|_2^2 \end{pmatrix}. \tag{1}
$$

For this solution to yield multipliers for the problem, we must have $0 \le \hat{\mu} = \alpha\,\langle a, b\rangle - \beta\,\|a\|_2^2$. If this is not the case, the inequality $\beta \ge b^T x$ must be inactive (or perhaps active, but not "binding", i.e., "binding= active with positive multiplier", "non-binding = multiplier is zero"), in which case

$\mu = 0$. In this case $\lambda = -\alpha/\|a\|_2^2$ and $\bar{x} = \frac{\alpha}{\|a\|_2^2}a$ as was seen in part (a) above. Putting this all together, we have

$$\begin{pmatrix}\bar{x}\\ \bar{\lambda}\\ \bar{\mu}\end{pmatrix} = \begin{cases}\begin{pmatrix}\frac{\alpha}{\|a\|_2^2}a\\ \frac{-\alpha}{\|a\|_2^2}\\ 0\end{pmatrix} & \text{, if } \alpha\frac{\langle a,b\rangle}{\|a\|_2^2} < \beta,\\ \begin{pmatrix}-(\hat{\lambda}a + \hat{\mu}b)\\ \hat{\lambda}\\ \hat{\mu}\end{pmatrix} & \text{, otherwise,}\end{cases} \tag{2}$$

where $\hat{\lambda}$ and $\hat{\mu}$ are given in (1). Notice that in the case $\alpha\frac{\langle a,b\rangle}{\|a\|_2^2} < \beta$ we have $b^T\bar{x} = \frac{\alpha}{\|a\|_2^2}b^Ta < \beta$ as required.

(d) $\min\left\{\frac{1}{2}\|x\|_2^2 \,|\langle a,x\rangle \le \alpha,\ \langle b,x\rangle \le \beta\right\}$ where $a, b \in \mathbb{R}^n \setminus \{0\}$ with a and b linearly independent and $\alpha, \beta \in \mathbb{R}$.

Solution: The solution to this problem is embedded in the solutions to the previous problems. First observe that if both $\alpha \ge 0$ and $\beta \ge 0$, then $\bar{x} = 0$ solves the problem and the mulitpliers both take the value zero. So we now assume that at least one of α and β is negative. We now run through the various cases. Let $\bar{x}$ denote the solution, and let $\bar{\lambda} \ge 0$ and $\bar{\mu} \ge 0$ be the multipliers for the constraints $\langle a,x\rangle \le \alpha$ and $\langle b,x\rangle \le \beta$, respectively.

Case 1: $\langle a,\bar{x}\rangle < \alpha$ and $\langle b,\bar{x}\rangle < \beta$. Complementarity implies that both multipliers must be zero. In this case the solution must be $\bar{x} = 0$ which can only occur if both $0 \le \alpha$ and $0 \le \beta$, and so $\bar{x} = 0$, $\bar{\lambda} = 0$, and $\bar{\mu} = 0$.

Case 2: $\langle a,\bar{x}\rangle = \alpha$ and $\langle b,\bar{x}\rangle = \beta$. If both constraints are active, then the multipliers are given by (1) above and $\bar{x}$ is given by (2). Since both multipliers are non-negative, we have at least one of α and β is negative and

$$\begin{aligned}
\beta\langle a,b\rangle &\ge \alpha\|b\|_2^2 && (3)\\
\alpha\langle a,b\rangle &\ge \beta\|a\|_2^2 && (4)\\
\bar{\lambda} &= (\beta\langle a,b\rangle - \alpha\|b\|_2^2)/(\|a\|_2^2\|b\|_2^2 - \langle a,b\rangle^2) && (5)\\
\bar{\mu} &= (\alpha\langle a,b\rangle - \beta\|a\|_2^2)/(\|a\|_2^2\|b\|_2^2 - \langle a,b\rangle^2) && (6)\\
\bar{x} &= -(\bar{\lambda}a + \bar{\mu}b). && (7)
\end{aligned}$$

To use this solution, first check that (3) and (4) are satisfied. If they are, then the solution is given by (5)-(7).

Case 3: $\langle a,\bar{x}\rangle = \alpha$ and $\langle b,\bar{x}\rangle < \beta$. By the previous problem, at least one of α and β is negative and

$$\begin{aligned}
\beta\|a\|_2^2 &> \alpha\langle a,b\rangle && (8)\\
\alpha\|b\|_2^2 &\le \beta\langle a,b\rangle && (9)\\
\bar{\lambda} &= -\alpha/\|a\|_2^2 \quad \text{(in particular, } \alpha < 0\text{)} && (10)\\
\bar{\mu} &= 0 && (11)\\
\bar{x} &= (\alpha/\|a\|_2^2)a. && (12)
\end{aligned}$$

To use this solution, first check that (8) and (9) are satisfied. If they are, then the solution is given by (10)-(12).

Case 4: $\langle a, \bar{x}\rangle < \alpha$ and $\langle b, \bar{x}\rangle = \beta$. Here we just interchange the roles of the inequalities in the previous case to obtain the following: at least one of α and β is negative and

$$\begin{aligned}
\beta \|a\|_2^2 &\le \alpha \langle a, b\rangle \\
\alpha \|b\|_2^2 &> \beta \langle a, b\rangle \\
\bar{\lambda} &= 0 \\
\bar{\mu} &= -\beta / \|b\|_2^2 \quad \text{(in particular, } \beta < 0) \\
\bar{x} &= (\beta / \|b\|_2^2) b.
\end{aligned}$$

(e) $\min \left\{ \frac{1}{2} x^T H x \;\middle|\; \mathsf{e}^T x = 1 \quad \text{and} \quad r^T x = \mu \right\}$ where $H \in \mathbb{S}^n_{++}$, $\mathsf{e} := (1, 1, \ldots, 1)^T \in \mathbb{R}^n$ is the vector of all ones, e and $r \in \mathbb{R}^n$ are linearly independent, and $\mu \in \mathbb{R}$.

Solution: Since H is positive definite and, by hypotheses, $A = \begin{bmatrix} \mathsf{e}^T \\ r^T \end{bmatrix}$ has full rank, there is a unique solution-multiplier pair $\begin{pmatrix} \bar{x} \\ \bar{y} \end{pmatrix}$ obtained as the solution to the system $\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -g \\ b \end{pmatrix}$, or equivalently,

$$\begin{aligned}
Hx + ey_1 + ry_2 &= 0 \\
e^T x &= 1 \\
r^T x &= \mu.
\end{aligned}$$

Consequently, $x = -H^{-1}(ey_1 + ry_2)$, and so

$$\begin{aligned}
1 = e^T x &= -e^T H^{-1}(ey_1 + ry_2) = -e^T H^{-1} e \; y_1 - e^T H^{-1} r \; y_2 \\
\mu = r^T x &= -r^T H^{-1}(ey_1 + ry_2) = -r^T H^{-1} e \; y_1 - r^T H^{-1} r \; y_2.
\end{aligned}$$

That is,

$$\begin{pmatrix} 1 \\ \mu \end{pmatrix} = - \begin{bmatrix} e^T H^{-1} e & e^T H^{-1} r \\ e^T H^{-1} r & r^T H^{-1} r \end{bmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}.$$

Therefore,

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = - \begin{bmatrix} e^T H^{-1} e & e^T H^{-1} r \\ e^T H^{-1} r & r^T H^{-1} r \end{bmatrix}^{-1} \begin{pmatrix} 1 \\ \mu \end{pmatrix},$$

where

$$\begin{bmatrix} e^T H^{-1} e & e^T H^{-1} r \\ e^T H^{-1} r & r^T H^{-1} r \end{bmatrix}^{-1} = \frac{1}{e^T H^{-1} e \, r^T H^{-1} r - (e^T H^{-1} r)^2} \begin{bmatrix} r^T H^{-1} r & -e^T H^{-1} r \\ -e^T H^{-1} r & e^T H^{-1} e \end{bmatrix}.$$

Note that the Cauchy-Schwarz inequality tells us that

$$e^T H^{-1} e \, r^T H^{-1} r - (e^T H^{-1} r)^2 \ge 0. \tag{13}$$

Consequently

$$\begin{aligned}
\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} &= \frac{1}{e^T H^{-1} e \, r^T H^{-1} r - (e^T H^{-1} r)^2} \begin{bmatrix} -r^T H^{-1} r & e^T H^{-1} r \\ e^T H^{-1} r & -e^T H^{-1} e \end{bmatrix} \begin{pmatrix} 1 \\ \mu \end{pmatrix} \\
&= \frac{1}{e^T H^{-1} e \, r^T H^{-1} r - (e^T H^{-1} r)^2} \begin{pmatrix} \mu e^T H^{-1} r - r^T H^{-1} r \\ e^T H^{-1} r - \mu e^T H^{-1} e \end{pmatrix},
\end{aligned}$$

which gives

$$
\begin{aligned}
x &= -H^{-1}(ey_1 + ry_2) = -H^{-1}\begin{bmatrix} e & r \end{bmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \\
&= H^{-1}\begin{bmatrix} e & r \end{bmatrix}\begin{bmatrix} e^T H^{-1} e & e^T H^{-1} r \\ e^T H^{-1} r & r^T H^{-1} r \end{bmatrix}^{-1}\begin{pmatrix} 1 \\ \mu \end{pmatrix} \\
&= \frac{1}{e^T H^{-1} e\, r^T H^{-1} r - (e^T H^{-1} r)^2} H^{-1}\begin{bmatrix} e & r \end{bmatrix}\begin{pmatrix} r^T H^{-1} r - \mu e^T H^{-1} r \\ -e^T H^{-1} r + \mu e^T H^{-1} e \end{pmatrix} \\
&= \frac{1}{e^T H^{-1} e\, r^T H^{-1} r - (e^T H^{-1} r)^2}\begin{bmatrix} H^{-1} e & H^{-1} r \end{bmatrix}\begin{pmatrix} r^T H^{-1} r - \mu e^T H^{-1} r \\ -e^T H^{-1} r + \mu e^T H^{-1} e \end{pmatrix}.
\end{aligned}
$$

(f) $\min\left\{\frac{1}{2}x^T H x \,\middle|\, \mathsf{e}^T x = 1 \quad \text{and} \quad r^T x \geq \mu\right\}$ where $H \in \mathcal{S}^n_{++}$, $\mathsf{e} := (1, 1, \ldots, 1)^T \in \mathbb{R}^n$ is the vector of all ones, e and $r \in \mathbb{R}^n$ are linearly independent, and $\mu \in \mathbb{R}$.

Solution: Let $\bar{x}$ denote the solution, and let $\bar{y}_1 \geq 0$ and $\bar{y}_2 \geq 0$ be the multipliers for the constraints $\langle \mathsf{e}, x\rangle = 1$ and $\langle r, x\rangle \geq \mu$, respectively.

There are only two cases to consider: (i) $\langle r, \bar{x}\rangle = \mu$ or (ii) $\langle r, \bar{x}\rangle > \mu$. If (i) occurs, then we are in the situation described in part (e). The difference from (e) is that the constraint in this problem is $r^T x \geq \mu$ which must now be written as $-r^T x \leq -\mu$ in order to interpret the sign condition on the multiplier correctly as $y_2 \geq 0$. This requires us to replace (r, μ) in (e) by its negative $(-r, -\mu)$.

Case (i): From part (e) the we have

$$
\bar{y}_2 = \frac{\mathsf{e}^T H^{-1}(-r) - (-\mu)\mathsf{e}^T H^{-1}\mathsf{e}}{\mathsf{e}^T H^{-1}\mathsf{e}\,(-r)^T H^{-1}(-r) - (\mathsf{e}^T H^{-1}(-r))^2} = \frac{\mu \mathsf{e}^T H^{-1}\mathsf{e} - \mathsf{e}^T H^{-1} r}{\mathsf{e}^T H^{-1}\mathsf{e}\, r^T H^{-1} r - (\mathsf{e}^T H^{-1} r)^2}.
$$

Consequently, the inequality (13) tells us that Case (i) occurs if and only if $\mu \geq \frac{\mathsf{e}^T H^{-1} r}{\mathsf{e}^T H^{-1}\mathsf{e}}$. This makes sense since the larger μ is the more likely the constraint $r^T x \geq \mu$ is active. In this case,

$$
\begin{aligned}
\bar{y}_1 &= \frac{r^T H^{-1} r - \mu \mathsf{e}^T H^{-1} r}{\mathsf{e}^T H^{-1}\mathsf{e}\, r^T H^{-1} r - (\mathsf{e}^T H^{-1} r)^2} \\
\bar{y}_2 &= \frac{\mu \mathsf{e}^T H^{-1}\mathsf{e} - \mathsf{e}^T H^{-1} r}{\mathsf{e}^T H^{-1}\mathsf{e}\, r^T H^{-1} r - (\mathsf{e}^T H^{-1} r)^2} \\
\bar{x} &= H^{-1}(r y_2 - \mathsf{e} y_1).
\end{aligned}
$$

Case (ii): This follows the same pattern, but now follow the pattern of Case (3) in part (d) above with $\bar{x} = -H^{-1} y_1 \mathsf{e}$.

(g) $\min\left\{4x_1^2 + 4x_1 x_2 + x_2^2 - 8x_1 - 4x_2 \,\middle|\, x_1 + x_2 \leq 0,\ x_1 - x_2 \leq 0\right\}$

Solution: We have $f(x) := 4x_1^2 + 4x_1 x_2 + x_2^2 - 8x_1 - 4x_2 = (2x_1 + x_2)^2 - 4(2x_1 + x_2) = z^2 - 4z$, where $z := 2x_1 + x_2$. Considered as a function of z, f is convex with the unique global minimizer of $\bar{z} = 2$. Consequently, $(\bar{x}_1, \bar{x}_2)$ is a global minimizer of f if and only if $2x_1 + x_2 = 2$. If we add the two constraints we see that $x_1 \leq 0$ and rewriting them as $x_1 \leq x_2 \leq -x_1$ tells us that $x_1 \leq -|x_2|$. The line $2x_1 + x_2 = 2$ does not intersect this constraint region so the solution does not lie on this line.

Since this is a convex optimization problem we need only locate a KKT point to obtain a global minimizer. After graphing the constraint region and the line $2 = 2x_1 + x_2$ we conjecture that the solution is $(\bar{x}_1, \bar{x}_2) = (0, 0)$. The KKT conditions yield the multipliers $(\bar{y}_1, \bar{y}_2) = (14, 2)$.

III Convexity

1. Which of the following functions are convex? Which are strictly convex?

   (a) $f(x) = x_1^2 - 4x_1 + 2x_2^2 + 7$

   Solution: $\nabla^2 f(x) = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$ which is positive definite, so f is strictly convex.

   (b) $f(x) = e^{-\|x\|^2}$

   Solution: It was shown above that $\nabla^2 f(x) = 2f(x)\left[2xx^T - I\right]$. Therefore, if $n > 1$, there is a $u \in \mathbb{R}^n$ is such that $\|u\|_2 = 1$ and $\langle u, x\rangle = 0$ with $u^T\nabla^2 f(x)u = -2f(x) < 0$. If $n \geq 1$ and $0 < \|x\|_2 < 1$, then $x^T\nabla^2 f(x)x = -2f(x)\|x\|_2(1 - \frac{\|x\|_2}{2}) < 0$. So f is not convex.

   (c) $f(x) = x_1^2 - 2x_1x_2 + \frac{1}{3}x_2^3 - 8x_2$

   Solution: $\nabla^2 f(0,0) = \begin{bmatrix} 2 & -2 \\ -2 & 0 \end{bmatrix}$, so $\begin{pmatrix} 1 \\ 1 \end{pmatrix}^T \nabla^2 f(0,0) \begin{pmatrix} 1 \\ 1 \end{pmatrix} = -2 < 0$ which implies that f cannot be convex.

   (d) $f(x) = (2x_1 - x_2)^2 + (x_2 - x_3)^2 + (x_3 - 1)^2$

   Solution: $f(x) = \left\| \begin{bmatrix} 2 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} - \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix} \right\|_2^2$ is a linear least squares function and so is convex.

   Moreover, since $\text{Nul}\left(\begin{bmatrix} 2 & -1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}\right) = \{0\}$, it is strictly convex.

   (e) $f(x_1, x_2) := x_1^2 + 2x_2x_2 + x_2^2 - 8x_1 - 8x_2$

   Solution: Since $\nabla^2 f(x) = 2\begin{pmatrix} 1 \\ 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \end{pmatrix}$ is positive semidefinite, f is convex. But f is not strictly convex since $\nabla^2 f(x)$ is not positive definite.

   (f) $f(x_1, x_2) = \ln(e^{x_1} + e^{x_2})$, where $\ln(\mu)$ is the natural log of μ for $\mu > 0$ and is $+\infty$ for $\mu \leq 0$.

   Solution: $\nabla^2 f(x) = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} - \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}\begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}^T$, where $\lambda_1 := \frac{e^{x_1}}{e^{x_1}+e^{x_2}} \leq 1$ and $\lambda_2 =: \frac{e^{x_2}}{e^{x_1}+e^{x_2}} \leq 1$. Hence,

   $$\begin{pmatrix} \alpha \\ \beta \end{pmatrix}^T \nabla^2 f(x) \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \lambda_1\alpha^2 + \lambda_2\beta^2 - (\lambda_1\alpha)^2 - (\lambda_2\beta)^2 = \lambda_1(1-\lambda_1)\alpha^2 + \lambda_2(1-\lambda_2)\beta^2 \geq 0$$

   with equality if and only if $\alpha = \beta = 0$. Hence f is strictly convex.

2. Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is a differentiable convex function and let $\Omega \subset \mathbb{R}^n$ is a closed convex set. Use the subdifferential inequality to show that $\bar{x}$ is a global solution to the problem $\min_{x\in\Omega} f(x)$ if and only if $f'(\bar{x}; x - \bar{x}) \geq 0$ for all $x \in \Omega$.

   Solution: The subdifferential inequality states that

   $$f(x) \geq f(x) + f'(\bar{x}; x - \bar{x}) \quad \forall\, y \in \mathbb{R}^n \quad \text{and} \quad \bar{x} \in \text{dom}\,(f).$$

   Consequently, is $\bar{x} \in \Omega$ is such that $f'(\bar{x}; x - \bar{x}) \geq 0$ for all $x \in \Omega$, then $\bar{x}$ is necessarily a global minimizer of f over Ω. Conversely, if $\bar{x}$ is a global minimizer of f over Ω, then, for all $x \in \Omega$, $0 \leq t^{-1}(f(x) - f(\bar{x}))$ for all $t > 0$ so that $f'(\bar{x}; x - \bar{x}) \geq 0$ for all $x \in \Omega$.

3. Compute the Lagrangian dual to each of the following problems.

Solution: Since all of these problems are convex, they all have Lagrangian duals. In general, the solution procedure has the following steps:

Step 1 Write the equation for the Lagrangian $L(x, y)$. Two helpful tricks.

(a) It is often helpful separate the constraints into two groups, equalities and inequalities, writing a dual variable for each group, say $y = (u, v)$ where the u's are for the inequalities ($u \geq 0$) and the v's are for the equalities (free).

(b) Sometimes it is very useful to introduce new primal variables for complicated structures in the objective and then add the definition of the new variables as equality constraints. For example, if the objective contains a composite term of the form $h(Ax - b)$, then set $w = Ax - b$ adding this equation to the constraints. Then replace $h(Ax - b)$ by $h(w)$. Always use this trick for terms of the form $\frac{1}{2}\|Ax - b\|_2^2$ or, more generally, for any term of the form $\|Ax - b\|$ regardless of the norm involved.

Step 2 The goal here is to obtain an expression for the dual objective $\psi(y) := \min_x L(x, y)$. The first-order conditions for this problem are $0 = \nabla_x L(x, y)$, where we take the derivative with respect to all of the primal variables. These conditions identify the minima of L with respect to the primal variables. Then use these conditions to eliminate the primal variables from L so the dual objective can be written entirely in terms of the dual variables. *This is the hard step.* But keep going until all primal variables are eliminated. This may require starting again in Step 1 by introducing new primal variables that help simplify Step 2. (In a graduate course you learn about *conjugate duality* also known as *Fenchel-Rockafellar duality* which both dramatically simplifies and extends this technique to the non-differentiable setting. Terry Rockafellar is a UW emeritus Professor. He is known as the *Father of Convex Analysis.* At 85 years of age he still backpacks and back-country skis in the Cascade mountains with his pack dog. He also still publishes ground-breaking research at a regular rate.)

Step 3 Write the dual problem. Then consider ways to simplify your expression for the dual by (i) combining terms, (ii) eliminating variables, and (iii) partial optimization (a procedure I do not expect student to perform in an introductory course).

(a) $\min\left\{\frac{1}{2}\|x\|_2^2 \;\middle|\; a^T x \leq \alpha\right\}$ where $a \in \mathbb{R}^n \setminus \{0\}$ and $\alpha \in \mathbb{R}$.

Solution: The condition $0 = \nabla_x L(x, y)$ tells us that $x = -ya$. Hence the dual is the one dimensional problem
$\max\left\{-\frac{\|a\|_2^2}{2}y^2 - \alpha y \;\middle|\; 0 \leq y\right\}$, or $\min\left\{\frac{\|a\|_2^2}{2}y^2 + \alpha y \;\middle|\; 0 \leq y\right\}$ whose obvious solution is $y = \max\{-\alpha/\|a\|_2^2,\ 0\}$.

(b) $\min\left\{\frac{1}{2}\|x\|_2^2 \;\middle|\; a^T x = \alpha,\ b^T x \leq \beta\right\}$ where $a, b \in \mathbb{R}^n \setminus \{0\}$ with a and b linearly independent and $\alpha, \beta \in \mathbb{R}$.

Solution: $\max\left\{-\frac{1}{2}\|y_1 a + y_2 b\|_2^2 - \alpha y_1 - \beta y_2 \;\middle|\; 0 \leq y_2\right\}$.

(c) $\min\left\{\frac{1}{2}\|x\|_2^2 \;\middle|\; a^T x \leq \alpha,\ b^T x \leq \beta\right\}$ where $a, b \in \mathbb{R}^n \setminus \{0\}$ with a and b linearly independent and $\alpha, \beta \in \mathbb{R}$.

Solution: $\max\left\{-\frac{1}{2}\|y_1 a + y_2 b\|_2^2 - \alpha y_1 - \beta y_2 \;\middle|\; 0 \leq y_1,\ 0 \leq y_2\right\}$.

(d) $\min\left\{\frac{1}{2}x^T H x \;\middle|\; \mathsf{e}^T x = 1 \quad \text{and} \quad r^T x = \mu\right\}$ where $H \in \mathbb{S}^n_{++}$, $\mathsf{e} := (1, 1, \ldots, 1)^T \in \mathbb{R}^n$ is the vector of all ones, e and $r \in \mathbb{R}^n$ are linearly independent, and $\mu \in \mathbb{R}$.

Solution: $\max -\frac{1}{2}(y_1\mathsf{e} + y_2 r)^T H^{-1}(y_1\mathsf{e} + y_2 r) - y_1 - y_2\mu$.

(e) $\min\left\{\frac{1}{2}x^T Hx \;\middle|\; \mathsf{e}^T x = 1 \quad\text{and}\quad r^T x \geq \mu\right\}$ where $H \in \mathcal{S}^n_{++}$, $\mathsf{e} := (1, 1, \dots, 1)^T \in \mathbb{R}^n$ is the vector of all ones, e and $r \in \mathbb{R}^n$ are linearly independent, and $\mu \in \mathbb{R}$.

Solution: $\max\left\{-\frac{1}{2}(y_1\mathsf{e} - y_2 r)^T H^{-1}(y_1\mathsf{e} - y_2 r) - y_1 + y_2\mu \;|0 \leq y_2\right\}$.

(f) $\min\left\{4x_1^2 + 4x_1x_2 + x_2^2 - 8x_1 - 4x_2 \;|x_1 + x_2 \leq 0,\ x_1 - x_2 \leq 0\right\}$

Solution: Observe that $4x_1^2 + 4x_1x_2 + x_2^2 = z^2$ where $z = 2x_1 + x_2$. Hence we can rewrite this problem as

$$\min\left\{\frac{1}{2}z^2 - 4z \;|z = 2x_1 + x_2,\ x_1 + x_2 \leq 0,\ x_1 - x_2 \leq 0\right\}.$$

The Lagrangian is

$$L(z, x_1, x_2, y_1, y_2, y_3) = \frac{1}{2}z^2 - 4z + y_1(2x_1 + x_2 - z) + y_2(x_1 + x_2) + y_3(x_1 - x_2),$$

where $0 \leq y_2, y_3$. The dual objective is $\psi(y) = \min_x L(x, y)$. Due to convexity, the minimum is attained for any x at which $0 = \nabla_x L(x, y)$. For this problem, $0 = \nabla_x L(x, y)$ if and only if

$$\begin{aligned}
0 &= z - 4 - y_1 \\
0 &= 2y_1 + y_2 + y_3 \\
0 &= y_1 + y_2 - y_3.
\end{aligned}$$

Plugging this information into L gives

$$\psi(y) = \frac{1}{2}(y_1 + 4)^2 - 4(y_1 + 4) - y_1(y_1 + 4) = -\frac{1}{2}(y_1 + 4)^2$$

Hence the dual is

$$\begin{aligned}
&\text{minimize } \frac{1}{2}(y_1 + 4)^2 \\
&\text{subject to } \begin{bmatrix} 2 & 1 & 1 \\ 1 & 1 & -1 \end{bmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\
&\qquad 0 \leq y_2,\ 0 \leq y_3.
\end{aligned}$$

(g) Let $Q \in \mathbb{S}^n_{++}$, $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$, and $b \in \mathbb{R}^m$ and compute the Lagrangian dual to the problem

$$\begin{array}{lll}
\mathcal{P} & \text{minimize} & \frac{1}{2}x^T Qx + c^T x \\
 & \text{subject to} & Ax \leq b,\ 0 \leq x\ .
\end{array}$$

Solution: The Lagrangian is $L(x, u, v) = \frac{1}{2}x^T Qx + c^T x + u^T(Ax - b) - v^T x$ with $0 \leq u,\ 0 \leq v$. Stationarity of the Lagrangian gives

$$0 = \nabla_x L(x, u, v) = Qx + c + A^T u - v,$$

so that $\bar{x} = -Q^{-1}(c + A^T u - v)$. Plugging this into L gives

$$\begin{aligned}
L(\bar{x}, u, v) &= \frac{1}{2}(c + A^T u - v)^T Q^{-1}(c + A^T u - v) - (c + A^T u - v)^T Q^{-1}(c + A^T u - v) - b^T u \\
&= -\left[\frac{1}{2}(c + A^T u - v)^T Q^{-1}(c + A^T u - v) + b^T u\right].
\end{aligned}$$

Hence the dual is

$$\begin{aligned}
&\text{maximize } -\left[\frac{1}{2}(c + A^T u - v)^T Q^{-1}(c + A^T u - v) + b^T u\right] \\
&\text{subject to } 0 \leq u,\ 0 \leq v.
\end{aligned}$$

Moreover, if $(\bar{u}, \bar{v})$ solve the dual, then $\bar{x} = -Q^{-1}(c + A^T\bar{u} - \bar{v})$ solves the primal.

4. Let $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$. Consider the optimization problem

$$\mathcal{P} \quad \begin{array}{ll} \text{minimize} & \frac{1}{2}\|Ax-b\|_2^2 \\ \text{subject to} & \|x\|_1 \le 1\ . \end{array}$$

(a) Show that this problem is equivalent to the problem

$$\hat{\mathcal{P}} \qquad \begin{array}{ll} \text{minimize}_{(x,z,w)} & \frac{1}{2}\|w\|_2^2 \\ \text{subject to} & Ax-b=w, \\ & -z \le x \le z \ \text{ and } \ e^T z \le 1, \end{array}$$

where e is the vector of all ones.

Solution:

Observe that $\{x| -z \le x \le z \ \text{ and } \ e^T z \le 1\} = \{x| \|x\|_1 \le 1\}$.

(b) What is the Lagrangian for $\hat{\mathcal{P}}$?

Solution:

$$L(w,x,z,y,u,v,\lambda) = \frac{1}{2}w^T w + y^T(Ax-b-w) - u^T(x+z) + v^T(x-z) + \lambda(e^T z - 1)$$

(c) Show that the Lagrangian dual for $\hat{\mathcal{P}}$ is the problem

$$\mathcal{D} \qquad \max -\frac{1}{2}\|y\|_2^2 - y^T b - \left\|A^T y\right\|_\infty \qquad = \qquad -\min \frac{1}{2}\|y-b\|_2^2 + \left\|A^T y\right\|_\infty - \frac{1}{2}\|b\|_2^2\ .$$

This is also the Lagrangian dual for $\mathcal{P}$.

Solution:

$$\begin{aligned} L(w,x,z,y,u,v,\lambda) &= \frac{1}{2}w^T w + y^T(Ax-b-w) - u^T(x+z) + v^T(x-z) + \lambda(e^T z - 1) \\ g(y,u,v,\lambda) &= \min_{w,x,z} L(w,x,z,y,u,v,\lambda) \\ 0 &= \nabla_w L = w - y \\ 0 &= \nabla_x L = A^T y - u + v \\ 0 &= \nabla_z L = -u - v + \lambda e, \end{aligned}$$

hence $g(y,u,v,\lambda) = -\frac{1}{2}y^T y - y^T b - \lambda$ and the Lagrangian dual is

$$(\tilde{\mathcal{D}}) \quad \begin{aligned} \max \ & -\frac{1}{2}y^T y - y^T b - \lambda \\ \text{s.t. } & A^T y = u - v \\ & \lambda e = u + v \\ & \lambda \ge 0, u \ge 0, v \ge 0. \end{aligned}$$

Observe that

$$\{(y,\lambda)|A^T y = u-v, \lambda e = u+v, u \ge 0, v \ge 0, \lambda \ge 0\} = \{(y,\lambda)|A^T y + \lambda e \ge 0, -A^T y + \lambda e \ge 0, \lambda \ge 0\},$$

hence $(\tilde{\mathcal{D}})$ is equivalent to

$$(\tilde{\mathcal{D}}) \quad \begin{aligned} \max \ & -\frac{1}{2}y^T y - y^T b - \lambda \\ \text{s.t. } & -\lambda e \le A^T y \le \lambda e \end{aligned}$$

which is equivalent to $\mathcal{D}$.

5. Consider the functions

$$f(x) = \frac{1}{2}x^T Qx - c^T x$$

and

$$f_t(x) = \frac{1}{2}x^T Qx - c^T x + t\phi(x),$$

where $t > 0$, $Q \in \mathbb{S}^n_+$, $c \in \mathbb{R}^n$, and $\phi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is given by

$$\phi(x) = \begin{cases} -\sum_{i=1}^n \ln x_i & \text{, if } x_i > 0,\ i = 1, 2, \ldots, n, \\ +\infty & \text{, otherwise.} \end{cases}$$

(a) Show that ϕ is a convex function.

(b) Show that both f and f_t are convex functions.

(c) Show that the solution to the problem $\min f_t(x)$ always exists and is unique.

(d) Let $\{t_i\}$ be a decreasing sequence of positive real scalars with $t_i \downarrow 0$, and let x^i be the solution to the problem $\min f_{t_i}(x)$. Show that if the sequence $\{x^i\}$ has a cluster point $\bar{x}$, then $\bar{x}$ must be a solution to the problem $\min\{f(x) : 0 \le x\}$.
Hint: Use the KKT conditions for the QP $\min\{f(x) : 0 \le x\}$.

Solution

(a) $\nabla^2\phi(x) = \text{diag}(x)^{-2}$ which is positive definite on $\mathbb{R}^n_{++}$.

(b) $\nabla^2 f(x) = Q$ which is positive definite. The result follows from (a) and the fact that the sum of two convex functions is convex.

(c) The sum of a symmetric positive definite and symmetric positive semi-definite matrix is symmetric and positive definite. So ∇f_t is everywhere positive definite which implies that f_t is strictly convex. Hence if a solution exists, it must be unique.

(d) For all $i = 1, 2, \ldots$, we have $0 = \nabla f_{t_i}(x^i) = Qx^i - c - t_i\text{diag}(x^i)^{-1}\mathsf{e}$, where e is the vector of all ones. Without loss of generality we can assume that $x^i \to \bar{x}$. Set $\bar{v} := Q\bar{x} - c$ so that $t_i\text{diag}(x^i)^{-1}\mathsf{e} = Qx^i - c \to \bar{v} \ge 0$. Observe that $\bar{v}^T\bar{x} = \lim_i t_i\mathsf{e}^T\text{diag}(x^i)^{-1}x^i = \lim_i t_i n = 0$. Hence $(\bar{x}, \bar{v})$ is a KKT pair for the convex optimization problem $\min\{f(x)\ |0 \le x\}$ which implies that $\bar{x}$ must be a solution to this problem.

6. Let $Q \in \mathbb{R}^{n\times n}$ be symmetric and positive definite. Consider the optimization problem

$$\begin{array}{lll} \mathcal{P} & \text{minimize} & \frac{1}{2}x^T Qx + c^T x \\ & \text{subject to} & \|x\|_\infty \le 1\ . \end{array}$$

(a) Show that this problem is equivalent to the problem

$$\hat{\mathcal{P}} \quad \begin{array}{ll} \text{minimize} & \frac{1}{2}x^T Qx + c^T x \\ \text{subject to} & -e \le x \le e\ , \end{array}$$

where e is the vector of all ones.

Solution:

$\{x|\, \|x\|_\infty \le 1\} = \{x|\ |x_i| \le 1 \quad \forall i = 1, \cdots, n\} = \{x| - e \le x \le e\}$.

(b) What is the Lagrangian for $\hat{\mathcal{P}}$?

Solution:

$$L(x, u, v) = \frac{1}{2}x^T Qx + c^T x - u^T(x + e) + v^T(x - e)$$

(c) Show that the Lagrangian dual for $\hat{\mathcal{P}}$ is the problem

$$\mathcal{D} \qquad \max -\frac{1}{2}(y-c)^T Q^{-1}(y-c) - \|y\|_1 \qquad = \qquad -\min \frac{1}{2}(y-c)^T Q^{-1}(y-c) + \|y\|_1 \ .$$

This is also the Lagrangian dual for $\mathcal{P}$.

Solution:

$$\begin{aligned} g(u,v) &= \min_x L(x,u,v) = \min_x \frac{1}{2}x^T Qx + c^T x - u^T(x+e) + v^T(x-e) \\ 0 &= \nabla_x L(x,u,v) = Qx + c - u + v \\ x &= Q^{-1}(u-v-c) \\ g(u,v) &= -\frac{1}{2}(u-v-c)^T Q^{-1}(u-v-c) - (u+v)^T e \end{aligned}$$

then the Lagrangian dual of $\hat{\mathcal{P}}$ is

$$\max_{u\geq 0, v\geq 0} g(u,v) = -\frac{1}{2}(u-v-c)^T Q^{-1}(u-v-c) - (u+v)^T e.$$

Set $y = u - v$, then we have

$$\max_{y, v\geq 0} g(y) = -\frac{1}{2}(y-c)^T Q^{-1}(y-c) - \|y\|_1 - 2e^T v,$$

it is easily seen that $v = 0$ when optimal, hence the dual becomes

$$\max_y g(y) = -\frac{1}{2}(y-c)^T Q^{-1}(y-c) - \|y\|_1 \ .$$