1 Dual Convex Programs

1.1 Duality and Perturbation Function

1. Consider the program
\[
\begin{align*}
\text{(P)} & \quad \min x_2 \\
& \text{s.t. } x_1^2 + x_2^2 \leq 1 \\
& \quad x_1 \geq 1
\end{align*}
\]

(a) Verify that (P) is a convex program.

**Solution:** The objective function \( x_2 \) is a linear function, so is convex. The functions \( x_1^2 + x_2^2 - 1 \) and \( 1 - x_1 \) are also convex, so the problem is convex.

(b) Can you find a Slater point for (P)?

**Solution:** The only feasible point is \((1,0)\), which is clearly not a Slater point.

(c) Find the optimal value \( MP \) and optimum \( x^* \) for (P). Is \( x^* \) a KKT point? Does this contradict the KKT Theorem?

**Solution:** The only feasible point is \((1,0)\), so \( x^* = (1,0)^T \) and \( MP = 0 \). The Lagrangian is \( L(x,\lambda) = x_2 + \lambda_1(x_1^2 + x_2^2 - 1) - \lambda_2(1 - x_1) \). For the KKT condition, there must be \( \lambda^* \geq 0 \) such that \( \nabla L(x^*, \lambda^*) = 0 \), so we must have:
\[
\begin{pmatrix}
0 \\
1
\end{pmatrix} + \lambda_1^* \begin{pmatrix} 2 \\ 0 \end{pmatrix} + \lambda_2^* \begin{pmatrix} -1 \\ 0 \end{pmatrix} = 0
\]

It is clear that there is no such \( \lambda^* \). This does not contradict the KKT theorem, because we have no Slater point.

2. Rewrite (P) as
\[
\begin{align*}
\text{(P2)} & \quad \min f(x) := x_2 \\
& \text{s.t. } g(x) := 1 - x_1 \leq 0 \\
& \quad x \in C := \{x \in \mathbb{R}^2 : x_1^2 + x_2^2 \leq 1\}
\end{align*}
\]
(a) Derive the dual function $h(\lambda)$ for (P2) and plot it.

**Solution:** The Lagrangian is $L(x, \lambda) = x_2 + \lambda(1 - x_1)$. The Lagrangian dual function is

$$h(\lambda) = \inf_{x \in C} L(x, \lambda) = \inf_{x \in C} [(x_2 - \lambda x_1) + \lambda]$$

It is easy to check that for the optimal solution, $x_1^2 + x_2^2 = 1$, otherwise we can increase $x_1$ and so reduce the objective function. So our objective function is $-\sqrt{1 - x_1^2} - \lambda x_1 + \lambda$, where $-1 < x_1 < 1$. By taking derivative with respect to $x_1$ and putting equal to 0 we have $x_1 = \frac{\lambda}{\sqrt{\lambda^2 + 1}}$, and so $x_2 = \frac{-1}{\sqrt{\lambda^2 + 1}}$. Note that you have to choose the sign of $x_1$ and $x_2$ properly to make the objective function smaller. So we have:

$$h(\lambda) = -\frac{1 - \lambda^2}{\sqrt{\lambda^2 + 1}} + \lambda = \lambda - \sqrt{\lambda^2 + 1}$$

The plot is shown in Figure 1.

![Figure 1: h(\lambda).](image)

(b) Derive and solve the dual program. Is the dual optimal value MD equal to MP? Is the dual attained?
Solution: The dual problem is \( \sup(\lambda - \sqrt{\lambda^2 + 1}, \lambda \geq 0) \). It is easy to check \( \text{MD} = 0 = \text{MP} \), but it is not attained for the dual problem.

3. Now rewrite (P) as

\[
\begin{align*}
\text{(P3)} & \quad \min \ f(x) = x_2 \\
& \text{s.t.} \quad g_1(x) = 1 - x_1 \leq 0 \\
& \quad \quad g_2(x) := x_1^2 + x_2^2 - 1 \leq 0
\end{align*}
\]

(a) Plot the perturbation function \( MP(z), z \in \mathbb{R}^2 \).

Solution: The perturbation problem is

\[
\text{(P3)(z)} \quad \min \ f(x) = x_2 \\
\text{s.t.} \quad g_1(x) = 1 - z_1 \leq x_1 \\
\quad \quad g_2(x) := x_1^2 + x_2^2 - 1 \leq 1 + z_2
\]

The feasible set is non-empty if \( z_2 \geq -1 \), and \( (1 - z_1) \leq \sqrt{1 + z_2} \). You can use Matlab to solve this problem, but it can also be solved by using the figure. The feasible region is the space inside the shpere such that \( 1 - z_1 \leq x_1 \). The minimum value of \( x_2 \) is achieved for a point on the circle created by the intersection of \( x_1^2 + x_2^2 = 1 + z_2 \) and \( x_1 = 1 - z_1 \) if \( z_1 \leq 1 \), or \( x_1 = 0 \) if \( z_1 > 1 \). So we have \( M(z) = -\sqrt{1 + z_2} - (1 - z_1)^2 \) if \( z_1 \leq 1 \), and \( M(z) = -\sqrt{1 + z_2} \) if \( z_1 > 1 \). The plot if \( M(z) \) is shown in Figure 2.

(b) Is there a nontrivial/nonvertical supporting hyperplane to the epigraph of \( MP(z) \)?

Solution: It can be seen from the figure, and also can be checked easily by considering the gradient of \( M(z) \) close to \( (z_1, z_2) = (0, 0) \) that there is no nontrivial/nonvertical supporting hyperplane to the epigraph of \( MP(z) \).

1.2 Quadratic Programs

Consider the convex quadratic program

\[
\begin{align*}
\text{(P3)} & \quad \min \ & \frac{1}{2} x^T H x + d^T x \\
& \text{s.t.} & A x & \leq b
\end{align*}
\]

where \( x \in \mathbb{R}^n \) and \( b \in \mathbb{R}^m \) and \( H, A, d \) are appropriate matrices/vectors with \( H \) positive definite.

1. Find/derive the Lagrangian dual function \( h(\lambda) \) and Lagrangian dual program.

Solution: The lagrangian function is \( L(x, \lambda) = \frac{1}{2} x^T H x + d^T x + \lambda^T (A x - b) \) for a vector \( \lambda \geq 0 \). Note that lagrangian is a convex function of \( x \), so to find
the inf, we take the derivative and put it equal to zero: $Hx + d + A^T\lambda = 0$, so $x^* = -H^{-1}(d + A^T\lambda)$ The lagrangian dual function is:

$$h(\lambda) = \frac{1}{2}(d^T + \lambda^T A)H^{-1}(d + A^T \lambda) - d^T H^{-1}(d + A^T \lambda) + \lambda^T (-AH^{-1}(d + A^T \lambda) - b)$$

$$= -\frac{1}{2}(d^T + \lambda^T A)H^{-1}(d + A^T \lambda) - \lambda^T b$$

The dual problem is $\max\{h(\lambda) : \lambda \geq 0\}$.

2 Trust Region Methods

1. Let $f(x) := 10(x_2 - x_1^2)^2 + (1 - x_1)^2$.

   (a) Draw the contour lines of the quadratic model/approximation of $f$ at the point $x = (0, -1)^T$.

   Solution: If you calculate the gradient and Hessain of $f$ at $x_c = (0, -1)^T$, we have $\nabla f(x_c) = (-2 - 20)^T$ and $\nabla^2 f(x_c) = \begin{pmatrix} 42 & 0 \\ 0 & 20 \end{pmatrix}$. By substitu-
tion in the formula we have

\[ m(x) = f(x_c) + \nabla f(x_c)^T (x - x_c) + \frac{1}{2} (x - x_c)^T \nabla^2 f(x_c) (x - x_c) = 21x_1^2 + 10x_2^2 - 2x_1 + 1 \]

The contours \( m(x) = k \) can be easily drawn by using Matlab for different \( k \).

(b) Draw the family of solutions of the trust region subproblem, TRS, as the trust region varies from \( \delta = 0 \) to \( \delta = 2 \). (Hint: Apply/use the optimality conditions for TRS.)

**Solution:** \( \bar{x} \) is an optimal solution for \( \min \{ m(x) : \|x - x_c\| \leq \delta \} \), if and only if \( (d = \bar{x} - x_c) \)

\[
\begin{align*}
(1) & \quad 0 = \nabla f(x_c) + \nabla^2 f(x_c) d + \lambda d, \quad \lambda \geq 0 \\
(2) & \quad \|d\| \leq \delta \\
(3) & \quad \lambda(\|d\| - \delta) = 0 \\
(4) & \quad \nabla^2 f(x_c) + \lambda I \geq 0
\end{align*}
\]

(1)

Here as \( \lambda \geq 0 \) and \( \nabla^2 f(x_c) \geq 0 \), (4) is always satisfied. If we solve (1) we have \( d_1 = \frac{2}{42 + \lambda} \) and \( d_2 = \frac{20}{20 + \lambda} \). If \( \lambda = 0 \), then \( (d_1, d_2) = (\frac{1}{21}, 1) \).

This satisfies constraint (2) for \( \delta \geq \sqrt{\frac{1}{21} + 1} = 1.0235 \). If \( \lambda > 0 \), then by (3) we have \( \|d\| = \delta \). In this case, we can find \( \lambda \) by solving

\[
\sqrt{\left(\frac{2}{42 + \lambda}\right)^2 + \left(\frac{20}{20 + \lambda}\right)^2} = \delta.
\]

The points are shown in Figure 3 when we change \( \delta \) from 2 to 0.
3 Penalty and Barrier Methods

1. Consider the program

\[(P) \quad \min f(x) = x_1 + x_2 \\
\text{s.t.} \quad x_1^2 - x_2 \leq 2\]

(a) Use the penalty function method with the Courant-Beltrami penalty term to solve (P).

**Solution:** The penalty function is \(P_k(x) = x_1 + x_2 + k[(x_1^2 - x_2 - 2)^+]^2\).
As it is proved in the notes we have

\[
\nabla P_k(x) = \left( \begin{array}{c}
1 + 2k(2x_1)(x_1^2 - x_2 - 2) + \\
1 - 2k(x_1^2 - x_2 - 2) +
\end{array} \right)
\]

\(\nabla P_k(x) = 0\) doesn’t have a solution when \(x_1^2 - x_2 - 2 \leq 0\). For \(x_1^2 - x_2 - 2 > 0\), the solution for \(\nabla P_k(x) = 0\) is \((-\frac{1}{2}, -\frac{7}{4} - \frac{1}{2k})^T\), which converges to \(x^* = (-\frac{1}{2}, -\frac{7}{4})^T\). For this point we have \(MP = \frac{9}{4}\).

(b) Show that the objection function \(F_k(x)\) corresponding to the absolute value penalty term has no critical points off the parabola \(x_1^2 - x_2 = 2\), for \(k > 1\), and compute the minimizer of \(F_k(x)\).
Consider the program

\( \text{min } f(x) = x_1 + x_2 + k(x_1^2 - x_2 - 2)^+ \) for the points
\( x_1^2 - x_2 > 2 \), we have

\[ \nabla f_k(x) = \begin{pmatrix} 1 + 2kx_1 \\ 1 - k \end{pmatrix} \]

It is clear that the second component is always nonzero for \( k > 1 \). This means for \( k > 1 \), the minimizer of \( f_k(x) \) occurs on the parabola, so we have \( x_2 = x_1^2 - 2 \). By substituting this in \( x_1 + x_2 \), the objective function becomes \( x_1^2 - 2 + x_1 \). By putting derivative equal to zero we have \( x_1 = \frac{-1}{2} \), and by using \( x_2 = x_1^2 - 2 \) we have \( x_2 = \frac{7}{4} \). This means \( x = (\frac{-1}{2}, \frac{7}{4})^T \) is the optimal solution for all \( f_k(x) \).

(c) Solve (P) using the log-barrier method and compare your solution with the one obtained from the penalty function method above.

**Solution:** For log-barrier function we have \( B_k(x) = x_1 + x_2 - \frac{1}{k} \log(2 + x_2 - x_1^2) \). Then we have:

\[ \nabla B_k(x) = \begin{pmatrix} 1 + \frac{2x_1}{k(2 + x_2 - x_1^2)} \\ 1 + \frac{1}{k(2 + x_2 - x_1^2)} \end{pmatrix} \]

By solving \( \nabla B_k(x) = 0 \), we get the solution \( (\frac{-1}{2}, \frac{7}{4} + \frac{1}{k})^T \), which again converges to \( x^* = (\frac{-1}{2}, \frac{7}{4})^T \).

(d) Confirm that you have the optimal solution using the KKT conditions.

**Solution:** Use \( \lambda^* = 1 \) and \( x^* = (-1/2, -7/4)^T \) and apply sufficiency of the KKT conditions for this convex program.

2. Consider the program

\[ \min \ f(x) = x_1^2 + 1 \]

s.t. \( 2 \leq x \leq 4, x \in \mathbb{R} \)

Plot the objective function \( f(x) \) and plot the barrier function \( B_t(x) = f(x) - (1/t)(\log(x-2) + \log(4-x)) \) for various values of \( t > 0 \). Include a plot of the optima \( x'(t) \).

**Solution:** The barrier function for \( t = .01, .1, .5 \) is shown in Figure 4. The objective function is shown by dashed line. The derivative set to 0 is:

\[
\begin{align*}
0 &= 2x - (1/t)(\frac{1}{x-2} - \frac{1}{4-x}) \\
0 &= 2tx(x-2)(4-x) - (4-x) + (x-2) \\
0 &= tx(x^2 - 6x + 8) + 3 - x \\
0 &= tx^3 - 6tx^2 - (8t-1)x + 3
\end{align*}
\]

For any \( t \), we can find a solution for the above inequality in the interval \([2,4]\). This solution as a function of \( t \) is shown in Figure 5. The minimizer of the barrier function is also shown in Figure 4 by star. From the figure, \( x_t \) converges to \( x = 2 \), which is clearly the optimal value of the problem, because \( x^2 + 1 \) is increasing on \([2,4]\).
3. Consider the convex program \( \text{min } f(x) \text{ s.t. } g(x) \leq 0 \) where \( f, g \) are sufficiently smooth/differentiable and \( f \) is coercive.

(a) Prove that the associated unconstrained program \( \text{min } F_k(x) := f(x) + kg^+(x) \) has a minimizer \( x_k \) for each positive integer \( k \).

**Solution:** First note that \( f(x) \) is a coercive convex function, so it has a minimizer. Let’s show the minimizer of \( f(x) \) by \( \bar{x} \). We have \( g^+(x) \geq 0 \), this means \( F_k(x) \) is also coercive. Note that \( g^+(x) \) is not smooth, but it is continuous, so \( F_k(x) \) is continuous. Any continuous coercive function attains its minimum. We also have \( f(\bar{x}) \) as lower bound for \( F_k(x) \):

\[
F_k(x) = f(x) + kg^+(x) \geq f(x) \geq f(\bar{x}), \quad \forall x
\]

(b) Prove that if the gradient of \( \phi_k(x) = f(x) + kg(x) \) is nonzero for all nonfeasible points for (P), then \( x_k \) must be feasible for (P).

**Solution:** Note that for the points \( g(x) > 0 \), we have \( F_k(x) = \phi_k(x) \), and \( F_k(x) \) is smooth. If \( x_k \) is infeasible, this means \( g(x) > 0 \), so \( F_k(s) \) is smooth at \( x_k \). By first order necessary condition, we must have \( \nabla F_k(x_k) = \nabla \phi_k(x) = 0 \). This is a contradiction to the gradient of \( \phi_k(x) \) is nonzero for all nonfeasible points.
Figure 5: $x_t$ as a function of $t$.

(c) Show by example that the sequence $\{x_k\}$ may converge to a point $x^*$ that is not a solution of (P). (Hint: Try a simple inconsistent program (P).)

**Solution:** Consider the minimization problem $\min \{x^2 : x^2 + 1 \leq 0\}$. This problem is infeasible. We have $F_k(x) = (k + 1)x^2 + k$. $x = 0$ is the minimizer for all $F_k(x)$, but it is clearly not the solution for our problem.
1. Consider the program

\[
\begin{align*}
\text{(P)} & \quad \min \ f(x) := x^2 + y^2 \\
& \text{s.t.} \ h(x) := (x - 2)^3 - y^2 = 0.
\end{align*}
\]

(a) Show that (P) admits no Lagrange multipliers and explain why.

**Solution** The Lagrangian would be

\[
L(x,y,\mu) = x^2 + y^2 + \mu ((x - 2)^3 - y^2).
\]

Taking the gradient and set it equal to zero we have:

\[
\begin{align*}
\nabla L_x(x,y,\mu) &= 2x + 3\mu(x - 2)^2 = 0 \quad (1) \\
\nabla L_y(x,y,\mu) &= 2y - 2\mu y = 0 \quad (2)
\end{align*}
\]

(2) implies that $\mu = 1$. Substituting this value of $\mu$ into (1) gives us $3x^2 - 10x + 12 = 0$, which has no real solution. Thus, there cannot be any real value of $\mu$ that simultaneously solves this system.

(b) Solve this problem graphically.

**Solution:** In the feasible region, we have $x \geq 2$. The feasible region is shown by red in Figure [6]. The circle centered at the origin with the minimum radius is also shown that has radius 2 and touches the feasible region at point $\begin{pmatrix} x, y \end{pmatrix}^T = (2, 0)$.

(c) What happens if you apply the Beltrami-Courant quadratic penalty function method? (I.e. use the quadratic penalty function $F_k(x) = f(x) + \frac{1}{2}k h(x)^T h(x)$.

**Solution:** Applying the penalty function method, you get the following function

\[
F_k(x,y) = x^2 + y^2 + \frac{1}{2}k ((x - 2)^3 - y^2)^2
\]

Differentiating this, we get

\[
\begin{align*}
\nabla F_x(x,y) &= 2x + k ((x - 2)^3 - y^2) (3(x - 2)^2) = 0 \\
\nabla F_y(x,y) &= 2y + k ((x - 2)^3 - y^2) (-2y) = 0
\end{align*}
\]

From the second equation, we have $y = 0$, or $k ((x - 2)^3 - y^2) = 1$. It is easy to check that the second one does not work. By substituting $y = 0$ in the first equation, we have $2x + k(x - 2)^3 = 0$. It can be solved numerically for different values of $k$, and it can be seen that the value of $x_k$ converges to 2, but the rate of convergence in really slow. We can see that $x_{100} = 1.5969$, $x_{1000} = 1.7412$, and $x_{10000} = 1.8350$. 
2. Determine all maxima and minima of \( f(x, y, z) = xz + y^2 \) on the sphere \( x^2 + y^2 + z^2 = 4 \).

**Solution:** We can start by taking the Lagrangian function:

\[
L(x, y, z, \mu) = xz + y^2 + \mu(x^2 + y^2 + z^2 - 4)
\]

If we take the gradient and set equal to zero we have:

\[
\begin{align*}
\nabla L_x(x, y, z, \mu) &= z + 2\mu x = 0 \quad (1) \\
\nabla L_y(x, y, z, \mu) &= 2y + 2\mu y = 0 \quad (2) \\
\nabla L_z(x, y, z, \mu) &= x + 2\mu z = 0 \quad (3)
\end{align*}
\]

From (2) we have \( y = 0 \) or \( \mu = -1 \). By \( \mu = -1 \), from (1) and (3) we have \( x = z = 0 \), and then from \( x^2 + y^2 + z^2 = 0 \) we have \( y = \pm 2 \). By using \( y = 0 \), by substitution we can get \( x = \pm \sqrt{2} \) and \( z = \pm \sqrt{2} \). Hence we get 4 points \( (\pm \sqrt{2}, 0, \pm \sqrt{2})^T \) which are minimum with objective value 2, and 2 points \( (0, \pm 2, 0)^T \) which are maximum with objective value 4.

3. Consider the Beltrami-Courant quadratic penalty function applied to

\[
\min -x_1 x_2 x_3 \text{ s.t. } 72 - x_1 - 2x_2 - 2x_3 = 0.
\]
(a) Solve the problem using the quadratic penalty function method.

**Solution:** The penalty function is

\[ F_k(x_1, x_2, x_3) = -x_1 x_2 x_3 + \frac{1}{2} k (72 - x_1 - 2x_2 - 2x_3)^2 \]

We can solve this numerically using Matlab to see that the optimal solution is \( x^* = (24, 12, 12)^T \).

(b) Verify that the explicit expression for \( x(k) \) is given by \( x_2 = x_3 = 24 / \left( 1 + \sqrt{1 - 8/k} \right) \) and \( x_1 = 2x_2 \), i.e. verify that this point satisfies stationarity for \( F_k \).

**Solution:** By taking the gradient of \( F_k \) and setting equal to zero we have:

\[
\begin{align*}
\nabla_{x_1} F_k(x_1, x_2, x_3) &= -x_2 x_3 - k (72 - x_1 - 2x_2 - 2x_3) = 0 \\
\nabla_{x_2} F_k(x_1, x_2, x_3) &= -x_1 x_3 - 2k (72 - x_1 - 2x_2 - 2x_3) = 0 \\
\nabla_{x_3} F_k(x_1, x_2, x_3) &= -x_1 x_2 - 2k (72 - x_1 - 2x_2 - 2x_3) = 0
\end{align*}
\]

By simple substitution we can see that \( x_2 = x_3 = 24 / \left( 1 + \sqrt{1 - 8/k} \right) \) and \( x_1 = 2x_2 \) satisfy the above equations.

(c) Verify that \( x(k) \to x^* \) as \( k \to \infty \).

**Solution:** By the relations given in part (b), it is clear that \( x(k) \to (24, 12, 12)^T \)

(d) Find \( x(k) \) when \( k = 9 \) and verify that the Hessian of \( F_9 \) is positive definite.

**Solution:** For \( k = 9 \), we have \( x(k) = (36, 18, 18)^T \). For the Hessian we have

\[
\nabla^2 F_k(x) = \begin{bmatrix}
    k & -x_3 + 2k & -x_2 + 2k \\
    -x_3 + 2k & 4k & -x_1 + 4k \\
    -x_2 + 2k & -x_1 + 4k & 4k
\end{bmatrix}
\]

\[
\nabla^2 F_9(x(9)) = 9 \begin{bmatrix}
    1 & 0 & 0 \\
    0 & 4 & 0 \\
    0 & 0 & 4
\end{bmatrix}
\]

which is positive definite.