UNIVERSITY OF CALIFORNIA, BERKELEY

HW 6: EE127

Oscar Ortega

March 15, 2019

1 ABOUT GENERAL OPTIMIZATION

- a: This statement is true:
  Proof:
  Consider the following convex problem:

  $$\min_{x \in \mathscr{X}} f_0(x)$$

  $$Ax = b$$

  and assume one of the inequalities is both strict at the optimum and the constraint was unable to be removed. This implies that the the optimal solution lies in the interior of the convex set of solutions.

  $$x^* \in \text{Interior}\mathscr{X}$$

  Here, we arrive at a contradiction as by lemma 8.6 in the textbook it must be the case that the if a feasible point of a convex problem is an optimal solution to an optimization problem, then it must be on the boundary. This would either imply that we could further minimize the value of the function by moving to the boundary. Or that if we are on the boundary of the convex set, then we could remove the constraint.

- b: This is true:
  Proof:

  $$\text{Let}(x_0, y_0) \in X \times Y$$

  $$F(x_0, y_0) \geq \min_x F(x, y_0) \geq \min_x \min_y F(x, y)$$

  $$\forall (x_0, y_0) : F(x_0, y_0) \geq \min_x \min_y F(x, y)$$

$$\min_x \min_y F(x,y) \geq \min_y \min_x F(x,y)$$

By symmetry:

$$\min_x \min_y F(x,y) \leq \min_y \min_x F(x,y)$$

Therefore:

$$\min_x \min_y F(x,y) = \min_y \min_x F(x,y)$$

- c:

$$\text{Let } d^* = \max_y \min_x F(x,y)$$

$$\text{Let } p^* = \min_x \max_y F(x,y)$$

$$h(y_0) = \min_x F(x,y_0) \leq F(x_0,y_0) \leq \max_y F(x_0,y) = g(x_0)$$

$$h(y_0) \leq g(x_0) : \forall (x_0, y_0)$$

$$\min_x F(x,y) \leq \min_x \max_y F(x,y)$$

$$d^* \leq p^*$$

2 Minimizing a Sum of Logarithms

- a:

We can re-express the following problem as follows:

$$p^* = \min_{x \in \mathbb{R}^n} - \sum_{i=1}^{n} \alpha_i \ln x_i$$

$$\text{subject to: } x \geq 0, 1^T x \leq c$$

The equality constraint can be reformulated as an inequality constraint because the function is monotonically decreasing, the constraint is monotonically increasing, and both functions can obtain their respective optimums ('the sets are closed').

- b:

the lagrangian of the following problem is as follows:

$$\mathcal{L}(x,\mu)_{x\geq 0} = -\sum_{i=1}^{n} a_i \ln(x_i) + \mu \sum_{i=1}^{n} x_i - c$$

$$g(\mu) = \min_x \mathcal{L}(x,\mu)$$

- c:

$$g(\mu) = \min_x \mathscr{L}(x, \mu)$$
$$= \min_x - \sum_{i=1}^{n} a_i \ln(x_i) + \mu \sum_{i=1}^{n} x_i - c$$

Taking derivatives and setting equal to 0:

$$\nabla \mathscr{L}(x, \mu)_x - \sum_{i=1}^{n} \frac{a_i}{x_i} + \mu = 0$$

$$\mu = \sum_{i=1}^{n} \frac{a_i}{x_i}$$

$$\mu^{-1} = \sum_{i=1}^{n} \frac{x_i}{a_i}$$

$$x_i^* = \frac{\alpha_i}{\mu} \rightarrow x^* = \frac{\alpha}{\mu}$$

$$g(\mu) = -\sum_{i=1}^{n} a_i \ln(\frac{a_i}{\mu}) + \sum_{i=1}^{n} a_i - \mu c$$

$$d^* = \max_{\mu \geq 0} g(\mu) = \max_{\mu \geq 0} - \sum_{i=1}^{n} a_i \ln(\frac{a_i}{\mu}) + \sum_{i=1}^{n} a_i - \mu c$$

$$\frac{\partial g}{\partial \mu} = \sum_{i=1}^{n} \frac{a_i}{\mu} - \sum_{i=1}^{n} c = 0$$

$$\sum_{i=1}^{n} - \frac{a_i}{\mu} = \sum_{i=1}^{n} c$$
$$\mu^* = \frac{\alpha}{c}$$

- d:

  Because we now have an optimal value for μ we can substitute this value for μ in our expression for x^*

$$x_i^* = \frac{\alpha_i}{\frac{\alpha}{c}}$$

  Plugging into the primal:

$$p^* = \sum_{i=1}^{n} a_i \ln(x_i)$$

$$= \sum_{i=1}^{n} a_i \ln(\frac{\alpha_i}{\frac{\alpha}{c}})$$

$$= \sum_{i=1}^{n} a_i \ln(\alpha_i) - \sum_{i=1}^{n} a_i \ln(\frac{a}{c})$$

$$= \sum_{i=1}^{n} a_i \ln(\frac{c}{a}) + \sum_{i=1}^{n} a_i \ln(\alpha_i)$$

$$\alpha \ln(\frac{c}{\alpha}) + \sum_{i=1}^{n} a_i \ln(\alpha_i)$$

3 Trust Region Subproblem

- a: Using the second order conditions of this problem, and considering the hessian, which is just the matrix Q, one cannot assume convexity the matrix Q because the objective function's Hessian is not positive semi-definite. However, even if we assume Q is positive semi-definite, the constraints are not linear, therefore the problem is not convex.

- b: The spectral decomposition theorom tells us that given a matrix $Q \in \mathbf{S_n}$ There exists an orthogonal matrix V and a diagonal matrix Λ s.t $Q = V\Lambda V^T$

  Now consider our optimization problem:

  $$p^* = \min_x x^T Qx + 2c^T x : \|x\|_2 = 1$$

  Expanding the matrices and the vectors:

  $$= \min_x \sum_{i=1}^{n} (x_i Q x_i) + 2c_i x_i) : \sum_{i=1}^{n} x_i^2 = 1$$

  $$= \min_x \sum_{i=1}^{n} (x_i(v_i \lambda_i v_i^T)x_i) + 2c_i x_i) : \sum_{i=1}^{n} x_i^2 = 1$$

  Because these matrix V is orthonormal, we can reduce to the following

  $$= \min_x \sum_{i=1}^{n} \lambda_i x_i^2 + 2c_i x_i : \sum_{i=1}^{n} x_i^2 = 1$$

- c:

  Starting from our reformulation from part c:

  $$= \min_x \sum_{i=1}^{n} \lambda_i x_i^2 + 2c_i x_i : \sum_{i=1}^{n} x_i^2 = 1$$

  Let $z = \sqrt{x}$

  $$= \min_z \sum_{i=1}^{n} \lambda_i z_i + 2c_i \sqrt{z_i} : \sum_{i=1}^{n} z_i = 1$$

Because the domain of the variable change is only defined on the positive reals, we now need to add this constraint to the problem, take note how the variable change is strictly increasing.

$$= \min_z \sum_{i=1}^{n} \lambda_i z_i + 2c_i \sqrt{z_i} : \sum_{i=1}^{n} z_i = 1, z_i \geq 0 : i = 1, 2, ..., n$$

To show that this is equivalent to the following:

$$= \min_z \sum_{i=1}^{n} \lambda_i z_i - 2|c_i| \sqrt{z_i} : \sum_{i=1}^{n} z_i = 1, z_i \geq 0 : i = 1, 2, ..., n$$

Consider the minimized objective value for the reformulation at a given index i and let $c_i \geq 0$:

$$\lambda_i z_i - 2|c_i| \sqrt{z_i} = d, d \in \mathbb{R}$$

$$|c_i| = \frac{-d + \lambda_i z_i}{2\sqrt{z_i}} = d'$$

$$|c_i| = |-c_i| = d'$$

$$\to c_i = d'$$

$$\to -c_i = -d' : \forall c_i$$

Because this expression is minimized we can assume $-d' \geq d'$

This implies that the two reformulations are equivalent as they can both achieve the same minimized value.

Therefore, the minimization is equivalent to the following:

$$= \min_z \sum_{i=1}^{n} \lambda_i z_i - 2|c_i| \sqrt{z_i} : \sum_{i=1}^{n} z_i = 1, z_i \geq 0 : i = 1, 2, ..., n$$

4 Fenchel Conjugate

- a: Consider the definition of the Fenchel Conjugate:

$$f^*(x) = \sup_{y \in \mathbf{R}} \{\langle y, x \rangle - f(y)\}$$

  This can immediately be seen to be a convex function as it is the pointwise superemum of an affine transformation of convex functions of x.

- a: given

$$f(x) = (x-2)^2$$

$$f^*(x) = \sup_{y \in \mathbb{R}} \{\langle y, x \rangle - (y-2)^2$$

$$f^*(x) = \sup_{y \in \mathbb{R}} xy - (y-2)^2$$

Taking derivatives, one can maximize this expression in terms of y.

$$\frac{\partial f}{\partial y} xy - (y-2)^2$$

$$= x - 2(y-2) = 0$$

$$x = 2y + 4$$

$$y^* = \frac{4+x}{2}$$

$$f^*(x) = x\frac{4+x}{2} - (\frac{4+x}{2} - 2)^2$$

$$= 2x + \frac{x^2}{2} - (\frac{x}{2})^2$$

$$= \frac{x^2}{4} + 2x$$

- c: Starting from our formulation of the fenchel conjugate.

$$f^*(x) = \sup_{y\in\mathbb{R}} xy - (\frac{y^2}{4} + 2y)$$

We can similarily take derivatives as follows to maximize the expression in terms of y.

$$\frac{\partial f}{\partial y} xy - (\frac{y^2}{4} + 2y)$$

$$= x - \frac{1}{2}y - 2 = 0$$

$$2x - 4 = y^*$$

$$f^*(x) = x(2x-4) - (\frac{(2x-4)^2}{4} + 2(2x-4))$$

$$2x^2 - 4x - (\frac{4x^2 - 16x + 16}{4} + 4x - 4)$$

$$2x^2 - 4x - (x^2 - 4x + 4 + 4x - 8)$$

$$2x^2 - 4x - (x^2 - 4)$$

$$x^2 - 4x + 4 = (x-2)^2$$

$$(f^*)^*(x) = f(x)$$

- d: We now consider the following piecewise quantity.

$$f(x) = \begin{cases} -x & x < 0 \\ x & 0 \leq x \leq 0.5 \\ 1-x & 0.5 \leq x \leq 1 \\ x-1 & x \geq 1 \end{cases} \tag{4.1}$$

$$f^*(x) = \sup_{y \in \mathbf{R}}\{\langle y, x\rangle - f(y)\}$$

Rewrote the function to be easier to analyse:

$$f^*(x) = \max \begin{cases} \sup(x+1)y & x < 0 \\ \sup(x-1)y & 0 \leq x \leq 0.5 \\ \sup(x+1)y - 1 & 0.5 \leq x \leq 1 \\ \sup(x-1)y + 1 & x \geq 1 \end{cases} \tag{4.2}$$

$$f^*(x) = \begin{cases} 1 & x = 1 \\ \infty & else \end{cases} \tag{4.3}$$

$$(f^*)^*(x) = \max \begin{cases} \sup_y xy & x = -1 \\ \sup xy - \infty & else \end{cases} \tag{4.4}$$

We know this is true because the conjugate of the conjugate is the original function. =

$$f(x) = \begin{cases} -x & x < 0 \\ x & 0 \leq x \leq 0.5 \\ 1-x & 0.5 \leq x \leq 1 \\ x-1 & x \geq 1 \end{cases} \tag{4.5}$$

- We can see that there is a one to one correspondence

5 SVD

We know that in a general optimization problem:

$$\min f_0(x) = \max - f_0(x)$$

We know that that

$$max_{x,y \in \mathscr{B}_2} x^T Ay = \sqrt{\lambda}_{\max} A^T A = \sigma_{\max} A = 3$$
$$\min x^T Ay = -3$$

6 Online Least Squares

- a: This is just the optimizer for least squares.

$$w_t^* = (X_t^T X_t)^{-1} X_t^T y$$

- b: This is just the optimizer for least squares.

$$v_t^* = (Q_t^T Q_t)^{-1} Q_t^T y$$

- c:

$$X_t w_t^* = X_t((X_t^T X_t)^{-1} X_t^T) y$$

  Because Q_t is the orthonormalized matrix of X_t we know that the spans of the column spaces of the two matrices are equal. Furthermore, because the parenthesized matrix is a projection matrix, projecting y onto the column space of X is the same as projecting y onto the column space of Q. This implies that the $X_t w_t^* = Q_t v_t^*$

- d: By the orthogonal decomposition theorem:

$$\forall y \in \mathbb{R}^n : \exists e_t \in \mathrm{Col}(X_t)^\perp : y = X_t w_t^* + e_t$$

$$w_t^* = (X_t^T X_t)^{-1} X_t^T y$$
$$w_t^* = (X_t^T X_t)^{-1} X_t^T (X_t w_t^* + e_t)$$
$$= (X_t^T X_t)^{-1} X_t^T X_t w_t^* + X_t^T e_t$$
$$= (X_t^T X_t)^{-1} X_t^T X_t w_t^*$$
$$= (X_t^T X_t)^{-1} X_t^T Q_t v_t^*$$

- e:

$$v_t^* = (Q_t^T Q_t)^{-1} y$$
$$Q_t^T y$$

  Because we constructed Q orthonormal. For t = 2,3,...n

$$v_t^* = (v_{t-1}^* + q_t^T y)^T$$

7 Errors in the Measurement Apparatus

- a:

  We can show that Z as defined is invertible if we can prove the matrix is positive-definite.

$$
\begin{gathered}
x^T Z x = x^T (w w^T + \lambda I) x \\
= x^T w w^T x + x^T \lambda I x \\
(x^T w)^2 + \lambda \|x\|_2^2 \\
\rightarrow x^T Z x = 0 \rightarrow x = 0
\end{gathered}
$$

  Therefore, the matrix is PD.

$$
\begin{gathered}
f(Q) = \|Qw - y\|_2^2 + \lambda \|X - Q\|_F^2 \\
= (Qw - y)^T (Qw - y) + \lambda \mathrm{Tr}((X - Q)^T (X - Q)) \\
w^T Q^T Q w - w^T Q^T y - y^T Q w + y^T y + \lambda \mathrm{Tr}(Q^T Q) - 2\lambda \mathrm{Tr}(X^T Q) + y^T y + \lambda \mathrm{Tr}(X^T X) \\
\mathrm{Tr}(w^T Q^T Q w) - 2\mathrm{Tr}(y^T Q w) + \lambda \mathrm{Tr}(Q^T Q) - 2\lambda \mathrm{Tr}(X^T Q) + y^T y + \lambda \mathrm{Tr}(X^T X) \\
\mathrm{Tr}(Q^T Q w w^T) - 2\mathrm{Tr}(w y^T Q) + \lambda \mathrm{Tr}(Q^T Q) - 2\lambda \mathrm{Tr}(X^T Q) + y^T y + \lambda \mathrm{Tr}(X^T X)
\end{gathered}
$$

  Using the definitions of the gradient provided:

$$
\nabla_Q = 2Qww^T - 2yw^T + 2\lambda Q - 2\lambda X
$$

$$
\begin{gathered}
2(Q(ww^T + \lambda I) - (yw^T + \lambda X)) = 0 \\
Q(ww^T + \lambda I) = (yw^T + \lambda X) \\
Q^* = (yw^T + \lambda X)(ww^T + \lambda I)^{-1}
\end{gathered}
$$

- c:

$$
\begin{gathered}
\lim_{\lambda \to \infty} Q^* = (yw^T + \lambda X)(ww^T + \lambda I)^{-1} \\
= \lambda X (\lambda I)^{-1} \\
= X
\end{gathered}
$$