KKT Conditions

Video: 6A-C

Reading (review from the last module):

Chong, An Introduction to Optimization, 2011, Sections 19.3-19.4, 20.1

https://onlinelibrary-wiley-com.proxy1.library.jhu.edu/doi/book/10.1002/9781118033340

Rao, Engineering Optimization: Theory and Practice, 2009, Section 2.5

https://onlinelibrary-wiley-com.proxy1.library.jhu.edu/doi/book/10.1002/9780470549124

In the last module, we derived the Lagrangian

$$L(x, \lambda_1, \lambda_2, \ldots, \lambda_q) = f(x) + \sum_{i=1}^{q} \lambda_i h_i(x)$$

for the equality constrained problem

$$\begin{aligned}
&\min f(x) \\
&\text{s.t.} \\
&h_i(x) = 0,\ 1 \le i \le m
\end{aligned}$$

At the optimal solution x^*,

$$\nabla L(x^*, \lambda_1, \lambda_2, \ldots, \lambda_q) = 0,$$

which is equivalent to $\begin{cases} \dfrac{\partial L}{\partial x_j}(x^*, \lambda_1, \lambda_2, \ldots, \lambda_q) = 0 \\ \dfrac{\partial L}{\partial \lambda_j}(x^*, \lambda_1, \lambda_2, \ldots, \lambda_q) = 0 \end{cases}$

We also derived the Lagrangian

$$L(x, \mu_1, \mu_2, \ldots, \mu_m) = f(x) + \sum_{i=1}^{m} \mu_i g_i(x)$$

for an inequality constrained problem,

$$\begin{aligned}
&\min f(x) \\
&\text{s.t.} \\
&g_i(x) \le 0, 1 \le i \le q
\end{aligned}$$

At the optimal solution,

$$\frac{\partial L}{\partial x_j}(x^*, \mu_1, \mu_2, ..., \mu_m) = 0$$
$$\mu_i \geq 0$$
$$\mu_i g_i(x^*) = 0$$

Combining these two results, for a problem with both equality and inequality constraints, we have

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{q} \lambda_i h_i(x) + \sum_{i=1}^{m} \mu_i g_i(x)$$
$$\frac{\partial L}{\partial x_j}(x^*, \lambda, \mu) = 0$$
$$\frac{\partial L}{\partial \lambda_j}(x^*, \lambda, \mu) = 0$$
$$\mu_i \geq 0$$
$$\mu_i g_i(x^*) = 0$$

These are the *KKT conditions* (the $\frac{\partial L}{\partial x_j}(x^*, \lambda, \mu) = 0$ may be omitted or included as $h_i(x) = 0$).

As demonstrated in videos 6B-C, the KKT conditions are useful on their own, but they also help to define the *Lagrangian dual* problem.

Lagrangian Duality

Reading: (optional) Boyd, Convex Optimization, 2004, Chapter 5
http://web.stanford.edu/~boyd/cvxbook/bv_cvxbook.pdf

We have already seen duality in linear programming. Often, it is easier to solve the dual LP than the primal LP. The utility of duality increases with the complexity of the optimization problem because the more complex the problem, the fewer approaches are available for solving it.

For example, a non-convex optimization problem, in general, doesn’t have an efficient solution and may have to be solved heuristically, so the obtained solution is not guaranteed to be optimal. In such case, any information on how close the solution may be to the optimal solution is valuable and so it helps to have a bound on the value of the objective function, even if it is not a tight bound. A dual, in general (there are duals other than the Lagrangian dual), is an optimization problem that gives such a bound. The logic of building a Lagrangian dual is described below.

Consider a quadratic optimization problem,

$$\min c^T x + \frac{1}{2} x^T Q x$$

s.t. (1)

Ax=b

$x \geq 0$

We can enforce the $x \geq 0$ constraint explicitly, by requiring that x is non-negative. We can also enforce the same constraint by adding a step function to the objective function,

$$\text{step function}(x) = \begin{cases} 0, & x \geq 0 \\ \infty, & x < 0 \end{cases}$$

However, the above step function is not very helpful because it's not differentiable and it may return a ∞ instead of a finite number. A better step function is $\sup_{s \geq 0} \left(-s^T x \right)$.

Function sup(y) is the same as function max(y) except when y is bounded from the above but doesn't have a finite maximum value. In this class, we won't see any examples when sup(y) is not equal to max(y). However, it may help to still use sup instead of max because max(y) as a function that returns a value can be confused with max(y) as an optimization problem. max(y) as function stands for *maximum*(y), while max(y) as an optimization problem stands for *maximize*(y).

Let's see why the $\sup_{s \geq 0} \left(-s^T x \right)$ qualifies as a step function that prohibits negative values of x:

Question 1: If x=(3, 5), $\sup_{s} \left(-s^T x \right)$ is

Note: There is no $s \geq 0$ restriction.

- $-\infty$
- $+\infty$
- 0
- none of the above

Question 2: If x=(-3, 5), $\sup_{s} \left(-s^T x \right)$ is

- $-\infty$
- $+\infty$
- 0
- none of the above

Question 3: If x=(-3, -5), $\sup_{s}\left(-s^T x\right)$ is

- $-\infty$
- $+\infty$
- 0
- none of the above

Question 4: If x=(3, 5), $\sup_{s \geq 0}\left(-s^T x\right)$ is

Note: There is a $s \geq 0$ restriction now.

- $-\infty$
- $+\infty$
- 0
- none of the above

Question 5: If x=(-3, 5), $\sup_{s \geq 0}\left(-s^T x\right)$ is

- $-\infty$
- $+\infty$
- 0
- none of the above

Question 6: If x=(-3, -5), $\sup_{s \geq 0}\left(-s^T x\right)$ is

- $-\infty$
- $+\infty$
- 0
- none of the above

Question 7: If $x \geq 0$, $\sup_{s \geq 0}\left(-s^T x\right)$ is

- $-\infty$
- $+\infty$
- 0
- none of the above

Question 8: If x has at least one negative component, $\sup_{s \geq 0}\left(-s^T x\right)$ is

- $-\infty$
- $+\infty$

- 0
- none of the above

We now have an equivalent problem

$$\min_x \left(c^T x + \frac{1}{2} x^T Q x + \sup_{s \geq 0} - s^T x \right)$$
s.t.
Ax=b

or

$$\min_x \sup_{s \geq 0} c^T x + \frac{1}{2} x^T Q x - s^T x$$
s.t.
Ax=b

According to the famous minimax theorem (it was the starting point of the game theory), if $f(z_1, z_2)$ is convex for fixed z_2 and concave for fixed z_1,

$$\max_{z_2} \min_{z_1} f(z_1, z_2) = \min_{z_1} \max_{z_2} f(z_2, z_1)$$

provided the min and max are finite or

$$\sup_{z_2} \inf_{z_1} f(z_1, z_2) = \inf_{z_1} \sup_{z_2} f(z_2, z_1)$$

If the convex/concave conditions are not necessarily met,

$$\max_{z_2} \min_{z_1} f(z_1, z_2) \leq \min_{z_1} \max_{z_2} f(z_2, z_1)$$

$$\sup_{z_2} \inf_{z_1} f(z_1, z_2) \leq \inf_{z_1} \sup_{z_2} f(z_2, z_1)$$

Question 9: Which of the c^T, Q and s^T define whether $c^T x + \frac{1}{2} x^T Q x - s^T x$ is convex for fixed s? Check all that apply.

- c^T
- Q
- s^T
- none of the above

Question 10: Is $c^T x + \frac{1}{2} x^T Q x - s^T x$ convex if $Q = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}$?

- yes
- no
- depends on c and/or s

Question 11: Is $c^T x + \frac{1}{2} x^T Q x - s^T x$ concave for fixed x?

- yes
- no
- depends on the value of Q
- depends on the values of Q, c and/or s

If we assume convexity of the original objective function then, according to the minimax theorem, the solution to the dual

$$\max_{s \geq 0} \inf_x c^T x + \frac{1}{2} x^T Q x - s^T x$$

s.t.

Ax=b

gives a tight lower bound on the value of the objective function of the primal.

The same logic can be used to bring the equality constraint, Ax=b into the objective function as $\sup_y y^T (b - Ax)$.

Question 12: If $b - Ax < 0$ then $\sup_{y \geq 0} y^T (b - Ax)$ is

- $-\infty$
- $+\infty$
- 0
- none of the above

Question 13: If at least one component of $b - Ax$ is negative then $\sup_y y^T (b - Ax)$ is

- $-\infty$
- $+\infty$
- 0
- none of the above

Question 14: If at least one component of $b - Ax$ is positive then $\sup_y y^T (b - Ax)$ is

- $-\infty$
- $+\infty$
- 0
- none of the above

We now have an unconstrained problem

$$\max_{s\geq 0,y}\inf_{x} c^T x+\frac{1}{2}x^T Qx+y^T(b-Ax)-s^T x \quad \text{or} \quad \max_{s\geq 0,y}\min_{x} c^T x+\frac{1}{2}x^T Qx+y^T(b-Ax)-s^T x.$$

The $L(x,y,s)=c^T x+\frac{1}{2}x^T Qx+y^T(b-Ax)-s^T x$ is the same expression as we earlier defined as the Lagrangian (with λ and μ renamed into y and s, correspondingly) and the above unconstrained problem is the *Lagrangian dual* of the original problem. The fact that the primal and the dual share the optimal solution (provided the primal is convex) allows one to get rid of the $\inf_{x}$ part. Rearranging the terms of the sum,

$$L(x,y,s)=c^T x+\frac{1}{2}x^T Qx+y^T(b-Ax)-s^T x=b^T y+x^T(Qx+c-A^T y-s)-\frac{1}{2}x^T Qx$$

and, according to one of the KKT conditions, $\frac{\partial L}{\partial x_j}(x,y,s)$ must be zero at the optimal solution, so $Qx+c-A^T y-s=0$. This means that, at the optimal solution,

$$L(x,y,s)=b^T y-\frac{1}{2}x^T Qx \text{ and } x=Q^{-1}\left(s+A^T y-c\right).$$

Therefore, at the optimal solution, s and y fully define x, so $\max_{s\geq 0,y}\inf_{x}$ becomes simply $\max_{s\geq 0,y}$.

Therefore, the Lagrangian dual is

$$\begin{aligned}&\max b^T y-\frac{1}{2}x^T Qx\\ &\text{s.t.} \qquad\qquad\qquad\qquad (2)\\ &A^T y-Qx+s=c\\ &s\geq 0\end{aligned}$$

When the conditions for the $\sup_{z_2}\inf_{z_1} f(z_1,z_2)=\inf_{z_1}\sup_{z_2} f(z_2,z_1)$ are met, the primal optimal objective and the dual optimal objective are equal, this is called *strong duality*. When the conditions are not met, $\sup_{z_2}\inf_{z_1} f(z_1,z_2)\leq\inf_{z_1}\sup_{z_2} f(z_2,z_1)$ still holds and the maximization dual gives a lower bound for the minimization primal, but this may be not a tight bound. The difference between the primal and the dual objective functions at the optimal solutions is the *duality gap*.

The paper we will discuss in the next section also defines a *complementarity gap*, which is the difference between the primal and the dual objective functions at a given (not necessarily optimal) solution. In certain conditions, this complementarity gap is equal to the duality gap.

Question 15: Find the difference between the values of the primal (1) and the dual (2) objective functions at a solution that is feasible both for the primal and the dual (but is not necessarily an optimal solution), $c^T x + \frac{1}{2} x^T Qx - \left(b^T y - \frac{1}{2} x^T Qx \right) =$

- (A) $x^T s$
- (B) $s^T x$
- (C) $-x^T s$
- (D) $-s^T x$
- (E) $x^T s + y^T x$
- (F) $s^T x + x^T y$
- (A) and (B) are correct
- (C) and (D) are correct
- (E) and (F) are correct
- none of the above

Question 16: Express the Lagrangian dual of

$$\min c^T x + \frac{1}{2} x^T Qx$$

s.t.

$x \geq 0$

in terms of s, c and Q if Q is symmetric and invertible.

- $$\max_s - \frac{1}{2} s^T Q^{-1} s - c^T Q^{-1} s - \frac{1}{2} c^T Q^{-1} c$$
  s.t.
  $s \geq 0$

- $$\max_s - \frac{1}{2} s^T Q^{-1} s - c^T Q^{-1} s + \frac{1}{2} c^T Q^{-1} c$$
  s.t.
  $s \geq 0$

$$\max_s -\frac{1}{2}s^T Q^{-1} s + c^T Q^{-1} s - \frac{1}{2}c^T Q^{-1} c$$

- s.t.

  $s \geq 0$

$$\max_s -\frac{1}{2}s^T Q^{-1} s + c^T Q^{-1} s + \frac{1}{2}c^T Q^{-1} c$$

- s.t.

  $s \geq 0$

- none of the above

Question 17: (continued from the previous question) Find the optimal solution to the primal

$$\min c^T x + \frac{1}{2}x^T Q x$$

s.t.

$x \geq 0$

if $Q = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ $c = \begin{bmatrix} -2 \\ -3 \end{bmatrix}$. The optimal value of the objective function is -2.3333

A hint: function quadprog() may be useful.

Question 18: (continued from the previous question) Find the optimal solution to the dual. The optimal value of the objective function is 0

Question 19: (continued from the previous question) The duality gap is 2.3333

Question 20: (continued from the previous question) The complementarity gap at the optimal solution to the primal is 2.3333

Question 21: (continued from the previous question) Assuming that (x, s) is feasible both in the primal and the dual, the complementarity gap at x = (1, 1) is 2

Interior Point Quadratic Programing

Video: 6D-G

Reading: Gondzio et. al., Interior Point Methods 25 Years Later, 2012 – Part 2 till equation (7).

https://www-sciencedirect-com.proxy1.library.jhu.edu/science/article/pii/S0377221711008204

In the previous section, we have not applied one of the KKT conditions, $x^T s = 0$. In the matrix form, $x^T s = 0$ can be written as $XSe = 0$ where X and S are the diagonal matrixes of x and s and e is a vector of 1s.

In linear programming, the simplex algorithm is based on enforcing $XSe = 0$. In quadratic programming, the interior point method forces XSe *towards* zero. The simplex method walks along the perimeter of the feasible region, while the interior point QP algorithm attempts to go across the feasible region.

Consider the primal,

$$\begin{aligned} &\min c^T x + \frac{1}{2} x^T Q x \\ &\text{s.t.} \\ &Ax=b \\ &x \geq 0 \end{aligned} \tag{3}$$

Q is a symmetric n by n matrix and A is an m by n (m rows, n columns) matrix.

The Lagrangian is

$$L(x, y, s) = c^T x + \frac{1}{2} x^T Q x + y^T (b - Ax) - s^T x$$

and, applying the KKT conditions, we get

$$\begin{aligned} &Ax=b \\ &A^T y - Qx + s = c \\ &s \geq 0 \end{aligned}$$

Question 22: What is the size of vector y (the number of elements in y)?
- m
- n
- m+n
- n-m
- none of the above

Question 23: What is the size of vector s?
- m
- n
- m+n
- n-m
- none of the above

If we ignore the non-negativity constraint, we can solve the system of equations

$$\begin{cases} \text{Ax=b} \\ A^T y - Qx + s = c \\ XSe = 0 \end{cases}$$

using the Newton method.

The Newton method, in the form presented in module 4, gives the search direction Δz towards $\nabla f(z) = 0$,

$$\Delta z = z^{(k+1)} - z^{(k)} = -\left(\nabla^2 f(z^{(k)})\right)^{-1} \nabla f(z^{(k)})$$

The above equation can be also written as $\left(\nabla^2 f(z^{(k)})\right)\Delta z = -\nabla f(z^{(k)})$ to avoid inverting the Hessian $\nabla^2 f(z^{(k)})$. In our example, z is concatenated vectors x, y and s.

Let's define $f(x, y, s)$ to be a function with the gradient

$$\nabla f(x, y, s) = \begin{bmatrix} Ax - b \\ A^T y - Qx + s - c \\ XSe \end{bmatrix}$$

Question 24: What is the size of vector $\nabla f(x, y, s)$?

- m
- n
- m+n
- 2m+n
- m+2n
- 3m
- 3n
- none of the above

The Hessian is the Jacobian of the gradient,

$$\nabla^2 f(x, y, s) = \begin{bmatrix} \frac{\partial \nabla f(x, y, s)}{\partial x_1} & \frac{\partial \nabla f(x, y, s)}{\partial x_2} & \cdots & \frac{\partial \nabla f(x, y, s)}{\partial y_1} & \frac{\partial \nabla f(x, y, s)}{\partial y_2} & \cdots & \frac{\partial \nabla f(x, y, s)}{\partial s_1} & \cdots \end{bmatrix}$$

$$= \begin{bmatrix} A & 0 & 0 \\ -Q & A^T & I \\ S & 0 & X \end{bmatrix}$$

where I is the identity matrix, x_1, x_2, y_1, y_2, s_1, etc. are the components of x, y and s.

Question 25: The size of the identity matrix above is

- mxm
- nxn
- (m+n)x(m+n)
- (n-m)x(n-m)
- none of the above

Hence,

$$\begin{bmatrix} A & 0 & 0 \\ -Q & A^T & I \\ S & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix} = \begin{bmatrix} b - Ax \\ c + Qx - A^T y - s \\ -XSe \end{bmatrix}$$

Question 26: The size of vector $\begin{bmatrix} \Delta x & \Delta y & \Delta s \end{bmatrix}$ is

- m
- n
- m+n
- 2m+n
- m+2n
- 3m
- 3n
- none of the above

We start at some arbitrary chosen point $(x^{(0)}, y^{(0)}, s^{(0)})$ and, at each iteration,

$$\begin{bmatrix} \Delta x^{(k+1)} \\ \Delta y^{(k+1)} \\ \Delta s^{(k+1)} \end{bmatrix} = \begin{bmatrix} \Delta x^{(k)} \\ \Delta y^{(k)} \\ \Delta s^{(k)} \end{bmatrix} + \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix}$$

where

$$\begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix} \quad \text{is the solution to} \quad \begin{bmatrix} A & 0 & 0 \\ -Q & A^T & I \\ S^{(k)} & 0 & X^{(k)} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix} = \begin{bmatrix} b - Ax^{(k)} \\ c + Qx^{(k)} - A^T y^{(k)} - s^{(k)} \\ -X^{(k)} S^{(k)} e \end{bmatrix}$$

This iterative procedure doesn't guarantee $s \ge 0$. Recall that when we built the Lagrangian, we moved the $x \ge 0$ into the objective function and $x \ge 0$ was replaced by the combination of $XSe = 0$ and $s \ge 0$. If the $s \ge 0$ is not guaranteed, so is $x \ge 0$.

$x, s \ge 0$ is not guaranteed for two reasons. First, a step of Newton method may happen to be long enough to step outside of the $x, s \geq 0$ region. Second, the above procedure doesn't have anything in it that favors non-negative values of x and s. There is not much we can do about the first problem except to verify that $x, s \geq 0$ at each iteration and shorten the step if needed. However, if we fix the second problem, the search will be at least *biased* towards $x, s \geq 0$.

Therefore, let's start again from scratch and redefine the primal (3) as

$$\min c^T x + \frac{1}{2} x^T Q x - \mu \sum_{j=1}^{n} \ln x_j$$
s.t. (4)
Ax=b

where n is the size of vector x. Earlier in this module, we used $\max_{s \geq 0} -s^T x$ as a step function to enforce the non-negativity of x. $\max_{s \geq 0} -s^T x$ created a barrier at $x = 0$. Here, instead of the $\max_{s \geq 0} -s^T x$, we have $-\mu \sum_{j=1}^{n} \ln x_j$,which creates a barrier at a very small positive x.

Question 27: (not for forum discussion) The $-\mu \sum_{j=1}^{n} \ln x_j$ becomes a nearly vertical $+\infty$ barrier (the function is nearly L shaped) as μ approaches

- $-\infty$
- $+\infty$
- 0 from below
- 0 from above
- none of the above

With this barrier, the $x \geq 0$ constraint is taken care of and so only the Ax=b constraint is moved to the objective function to form the Lagrangian of the primal,

$$L(x, y, s) = c^T x + \frac{1}{2} x^T Q x + y^T (b - Ax) - \mu \sum_{j=1}^{n} \ln x_j$$

μ is an arbitrary constant for the barrier, it's not a Lagrange multiplier.

Applying the KKT conditions and then the Newton method,

$$\begin{bmatrix} A & 0 \\ 1? & 2? \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} b - Ax \\ c + Qx - A^T y - 3? \end{bmatrix}$$

Question 28: (not for forum discussion) Replace the 1? with the correct expression.

- $-Q-\mu X$
- $-Q-\mu X^{-1}$
- $-Q-\mu X^{-2}$
- $-Q+\mu X$
- $-Q+\mu X^{-1}$
- $-Q+\mu X^{-2}$
- none of the above

Question 29: (not for forum discussion) Replace the 2? with the correct expression.

- A
- A^T
- Q
- I
- none of the above

Question 30: (not for forum discussion) Replace the 3? with the correct expression.
- μX
- μX^{-1}
- μX^{-2}
- none of the above

Question 31: (not for forum discussion) What is the size of vector $\begin{bmatrix}\Delta x & \Delta y\end{bmatrix}^T$ above?

- m
- n
- m+n
- 2m
- 2n
- none of the above

Let's illustrate how the above works by performing one iteration of the Newton method (pure form) for the portfolio selection problem from module 1 with an equality instead of the inequality, so the expected return is exactly 0.2 and the objective is to minimize the risk,

$$\min \sum_{i=1}^{3}\sum_{j=1}^{3}\sigma_{ij}x_i x_j$$

subject to

$$\sum_{i=1}^{3} \mu_i x_i = 0.2$$

$$\sum_{i=1}^{3} x_i = 1$$

$$x_i \geq 0,\ i = 1, 2, 3$$

The means and covariances are the same as they were in module 1,

Stock	Expected Value, μ	Covariance with Stock 1	Covariance with Stock 2	Covariance with Stock 3
1	0.1	0.1	0.03	0.07
2	0.3	0.03	0.35	0.01
3	0.4	0.07	0.01	0.5

the starting point is $x^{(0)} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ $y^{(0)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\mu = 0.5$. Note that μ_i s are not related to μ, μ_i are the constants from the problem in module 1 and μ is the barrier constant from this module.

After the first iteration, $x^{(1)} = (1?, ?, ?)$.

Question 32: (not for forum discussion) Replace the 1? with the correct number. 1

Question 33: (not for forum discussion) How close is $x^{(1)}$ to the optimal solution x^* (which can be found with quadprog() as in module 1)? $\left\| x^{(1)} - x^* \right\|_2 =$ 2.3355

The above procedure doesn't guarantee $x \geq 0$, but the barrier makes it *biased* towards $x \geq 0$. It converges to the optimal solution provided $x \geq 0$ is verified at each iteration (and the step size is decreased until $x \geq 0$). μ has to be decreased at each iteration.

The algorithm based on

$$\begin{bmatrix} A & 0 & 0 \\ -Q & A^T & I \\ S & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix} = \begin{bmatrix} b - Ax \\ c + Qx - A^T y - s \\ -XSe \end{bmatrix}$$

didn't have the bias towards $x \geq 0$, but it had another important quality, it solved the primal and the dual together, taking advantage of the $x^T s = 0$ condition, which provided for fast convergence.

Hence, let’s combine the two approaches into one. We’ll use the Lagrangian from the second algorithm,

$$L(x, y, s) = c^T x + \frac{1}{2} x^T Q x + y^T (b - Ax) - \mu \sum_{j=1}^{n} \ln x_j$$

The KKT conditions are

$$\begin{aligned} &Ax=b \\ &A^T y - Qx + \mu X^{-1} e = c \end{aligned}$$

Next, we denote $\mu X^{-1} e$ by s so that $XSe = \mu e$.

The KKT conditions become

$$\begin{aligned} &Ax=b \\ &A^T y - Qx + s = c \end{aligned}$$

and we also have the $XSe = \mu e$ equation because this is how s was defined.

Applying the Newton method as we did before, we get a perturbed first version of the algorithm,

$$\begin{bmatrix} A & 0 & 0 \\ -Q & A^T & I \\ S & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix} = \begin{bmatrix} b - Ax \\ c + Qx - A^T y - s \\ \mu e - XSe \end{bmatrix}$$

μ is reduced by factor σ at each iteration, $\mu^{(k+1)} = \sigma \mu^{(k)}$, so, we can write it as

$$\begin{bmatrix} A & 0 & 0 \\ -Q & A^T & I \\ S & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta s \end{bmatrix} = \begin{bmatrix} b - Ax \\ c + Qx - A^T y - s \\ \sigma \mu e - XSe \end{bmatrix}$$

to match the notation in the paper.