

Caleb Logemann
MATH 565 Continuous Optimization
Final Exam

1. (a) *Proof.* Using Taylor Series for f we know that

$$f(x_k + \alpha p_k) = f(x_k) + \alpha p_k^T \nabla f(tx_k + (1-t)(x_k + \alpha p_k))$$

for some $t \in [0, 1]$. This can be simplified as follows

$$\begin{aligned} f(x_k + \alpha p_k) &= f(x_k) + \alpha p_k^T \nabla f(tx_k + (1-t)(x_k + \alpha p_k)) \\ &= f(x_k) + \alpha p_k^T \nabla f(x_k) + \alpha p_k^T \nabla f(tx_k + (1-t)(x_k + \alpha p_k)) - \alpha p_k^T \nabla f(x_k) \\ &= f(x_k) + \alpha p_k^T \nabla f(x_k) + \alpha p_k^T (\nabla f(tx_k + (1-t)(x_k + \alpha p_k)) - \nabla f(x_k)) \end{aligned}$$

By Cauchy-Schwarze

$$\leq f(x_k) + \alpha p_k^T \nabla f(x_k) + \alpha \|p_k\| \|\nabla f(tx_k + (1-t)(x_k + \alpha p_k)) - \nabla f(x_k)\|$$

By the Lipschitz continuity of ∇f

$$\begin{aligned} &\leq f(x_k) + \alpha p_k^T \nabla f(x_k) + \alpha L \|p_k\| \|tx_k + (1-t)(x_k + \alpha p_k) - x_k\| \\ &= f(x_k) + \alpha p_k^T \nabla f(x_k) + \alpha L \|p_k\| \|tx_k + (1-t)x_k + (1-t)\alpha p_k - x_k\| \\ &= f(x_k) + \alpha p_k^T \nabla f(x_k) + \alpha L \|p_k\| \|(1-t)\alpha p_k\| \\ &= f(x_k) + \alpha p_k^T \nabla f(x_k) + \alpha^2 (1-t) L \|p_k\|^2 \\ &\leq f(x_k) + \alpha p_k^T \nabla f(x_k) + \alpha^2 L \|p_k\|^2 \end{aligned}$$

This establishes the desired result. □

- (b) Prove that

$$\lim_{k \rightarrow \infty} (\nabla f(x_k)) = 0$$

Proof. First I will show that p_k is a descent direction. In order to do this consider $p_k^T \nabla f(x_k)$.

$$\begin{aligned} p_k^T \nabla f(x_k) &= -(\nabla f(x_k) \cdot e_{j_k})^T \nabla f(x_k) \\ &= -(0, \dots, 0, \frac{\partial f(x_k)}{\partial x_{j_k}}, 0, \dots, 0) \cdot \nabla f(x_k) \\ &= -\frac{\partial f(x_k)}{\partial x_{j_k}} < 0 \end{aligned}$$

This shows that p_k is a descent direction.

Given that f is continuously differentiable everywhere, that p_k is a descent direction and that f is bounded below and hence bounded below on the ray $\{x_k + \alpha p_k | \alpha > 0\}$, then by Lemma 3.1 there exists intervals of step lengths such that the Wolfe conditions are met.

Since there exists step lengths where the Wolfe conditions are met and since the step length α is chosen by an exact line search, then the α that is chosen must satisfy the Wolfe conditions.

Given that α_k satisfies the Wolfe conditions, f is continuously differentiable and bounded below, p_k is a descent direction, and ∇f is Lipschitz continuous, then Theorem 3.2 now states that

$$\sum_{k \geq 0} (\cos(\theta_k)^2 \|\nabla f(x_k)\|^2) < \infty$$

where

$$\cos(\theta_k) = \frac{-\nabla f(x_k)^T p_k}{\|\nabla f(x_k)\| \|p_k\|}.$$

This implies that

$$\lim_{k \rightarrow \infty} \left(\cos(\theta_k)^2 \|\nabla f(x_k)\|^2 \right) = 0.$$

Now if there exists $\delta > 0$ such that $\cos(\theta_k) > \delta$, then the desired limit can be verified.

$$\begin{aligned} \cos(\theta_k) &= \frac{-\nabla f(x_k)^T p_k}{\|\nabla f(x_k)\| \|p_k\|} \\ &= \frac{\frac{\partial f(x_k)}{\partial x_{j_k}}}{\|\nabla f(x_k)\| \|p_k\|} \\ &= \frac{\frac{\partial f(x_k)}{\partial x_{j_k}}}{\|\nabla f(x_k)\| \left| \frac{\partial f(x_k)}{\partial x_{j_k}} \right|} \\ &= \frac{\left| \frac{\partial f(x_k)}{\partial x_{j_k}} \right|}{\|\nabla f(x_k)\|} \end{aligned}$$

Since j_k is chosen such that $\left| \frac{\partial f(x_k)}{\partial x_{j_k}} \right| \geq \frac{1}{10} \|\nabla f(x_k)\|_\infty$, then

$$\cos(\theta_k) \geq \frac{1}{10} \frac{\|\nabla f(x_k)\|_\infty}{\|\nabla f(x_k)\|}$$

Since $\|\cdot\|_\infty$ and $\|\cdot\|_2$ are equivalent there exists $C > 0$ such that $\|\nabla f(x_k)\|_\infty > C \|\nabla f(x_k)\|_2$, thus

$$\cos(\theta_k) \geq \frac{1}{10} C > 0$$

Now this combined with the fact that $\lim_{k \rightarrow \infty} \left(\cos(\theta_k)^2 \|\nabla f(x_k)\|^2 \right) = 0$ shows that

$$\lim_{k \rightarrow \infty} (\|\nabla f(x_k)\|) = 0$$

and also

$$\lim_{k \rightarrow \infty} (\nabla f(x_k)) = 0.$$

□

2. (a) Without the constraint $x \geq 0$, the Lagrangian for this problem is

$$\mathcal{L}(x, \lambda) = \frac{1}{2} x^T Q x + c^T x - \lambda(a^T x - 1).$$

The gradient of the Lagrangian is then

$$\nabla_x \mathcal{L}(x, \lambda) = Qx + c - \lambda a.$$

The KKT conditions for this problem are

$$\begin{aligned} Qx + c - \lambda a &= 0 \\ a^T x - 1 &= 0 \\ \lambda(a^T x - 1) &= 0 \end{aligned}$$

The last equation is redundant with the second equation. These equations can be solved by noting that Q is invertible as it is diagonal with strictly positive entries on the diagonal. Specifically Q^{-1} is a diagonal matrix whose diagonal entries are the reciprocals of the entries of Q .

$$\begin{aligned} Qx + c - \lambda a &= 0 \\ Qx &= \lambda a - c \\ x &= Q^{-1}(\lambda a - c) \\ a^T x - 1 &= 0 \\ a^T(Q^{-1}(\lambda a - c)) - 1 &= 0 \\ \lambda a^T Q^{-1} a - a^T Q^{-1} c - 1 &= 0 \\ \lambda a^T Q^{-1} a &= a^T Q^{-1} c + 1 \\ \lambda &= \frac{a^T Q^{-1} c + 1}{a^T Q^{-1} a} \end{aligned}$$

Note that λ is well defined as $a^T Q^{-1} a > 0$ as $a > 0$ and Q^{-1} is diagonal with positive entries on the diagonal. Therefore the full solution to this problem is

$$x = Q^{-1} \left(\frac{a^T Q^{-1} c + 1}{a^T Q^{-1} a} a - c \right).$$

(b) Back in the full problem the Lagrangian is

$$\mathcal{L}(x, \lambda) = \frac{1}{2} x^T Q x + c^T x - \lambda(a^T x - 1) - \mu^T x.$$

The gradient of this Lagrangian is

$$\nabla_x \mathcal{L}(x, \lambda) = Qx + c - \lambda a - \mu$$

The KKT conditions for this problem are now

$$\begin{aligned} Qx + c - \lambda a - \mu &= 0 \\ a^T x - 1 &= 0 \\ x &\geq 0 \\ \mu &\geq 0 \\ \mu_i x_i &= 0 \quad \forall i \end{aligned}$$

(c) The KKT conditions from part (b) can be solved as follows.

test

(d)

3. (a) *Proof.* Clearly M is symmetric as

$$M^T = (I + ee^T)^T = I^T + (ee^T)^T = I + ee^T = M.$$

Next, to see that M is positive definite let $x \in \mathbb{R}^n$ such that $x \neq 0$, then

$$\begin{aligned} x^T Mx &= x^T (I + ee^T)x \\ &= (x^T + x^t ee^T)x \\ &= x^T x + x^T ee^T x \\ &= \|x\|^2 + \left(\sum_{i=1}^n (x_i) \right)^2. \end{aligned}$$

Since $\|x\|^2 > 0$ and $(\sum_{i=1}^n (x_i))^2 \geq 0$, this shows that

$$x^T Mx > 0.$$

Therefore M is positive definite as well. □

(b) The Lagrangian for this problem is

$$\mathcal{L}(x, \lambda) = \frac{1}{2} x^T Mx + (r - ae)^T x - \lambda^T x.$$

Now the gradient of the Lagrangian is

$$\nabla_x \mathcal{L}(x, \lambda) = Mx + r - ae - \lambda.$$

The KKT conditions are now

$$\begin{aligned} Mx + r - ae - \lambda &= 0 \\ x &\geq 0 \\ \lambda &\geq 0 \\ \lambda_i x_i &= 0 \quad \forall i. \end{aligned}$$

(c) *Proof.* Suppose that x^* is an optimal solution and that $x_i = 0$ for some $1 \leq i < n$. Assume to the contrary that there exists some $j > i$ such that $x_j > 0$. This implies that $\lambda_j = 0$, so that $\lambda_j x_j = 0$. Consider the equation for the j th entry of the Lagrangian.

$$\sum_{k \neq j} (x_k) + 2x_j + r_j - a + \lambda_j = 0$$

This can be solved for x_j also noticing that $\lambda_j = 0$

$$x_j = \frac{a - r_j - \sum_{k \neq j} (x_k)}{2}$$

Since $x_j > 0$ this implies that

$$a > r_j + \sum_{k \neq j} (x_k).$$

Also since $r_j > r_i$,

$$a > r_i + \sum_{k \neq j} (x_k).$$

Now consider the i th equation for Lagrangian condition.

$$\sum_{k \neq i} (x_k) + 2x_i + r_i - a - \lambda_i = 0.$$

Since $x_i = 0$ this can be solved for λ_i .

$$\lambda_i = \sum_{k \neq i} (x_k) + r_i - a.$$

The value of x_j can be pulled out of the sum, and the value can be replaced using the equation found earlier.

$$\lambda_i = \frac{a - r_j - \sum_{k \neq j} (x_k)}{2} + \sum_{k \neq i, k \neq j} (x_k) + r_i - a$$

Since $x_i = 0$ these two sums are equivalent, and this expression can be simplified.

$$\begin{aligned} \lambda_i &= \frac{a - r_j - \sum_{k \neq j} (x_k)}{2} + \sum_{k \neq i, k \neq j} (x_k) + r_i - a \\ \lambda_i &= \frac{a - r_j - \sum_{k \neq j} (x_k) + 2 \sum_{k \neq i, k \neq j} (x_k) + 2r_i - 2a}{2} \\ \lambda_i &= \frac{r_i - r_j + \sum_{k \neq j} (x_k) + r_i - a}{2} \end{aligned}$$

Since $r_i < r_j$ and $a > \sum_{k \neq j} (x_k) + r_i$ this implies that $\lambda_i < 0$, which is a contradiction.

Therefore this shows that $x_j = 0$. Hence if $x_i = 0$, then $x_j = 0$ for all $j > i$. \square

(d) Since M is symmetric positive definite it is also invertible. Consider the matrix

$$B = \frac{1}{n+1} \begin{bmatrix} n & -1 & -1 & \cdots & -1 \\ -1 & n & -1 & \cdots & -1 \\ \vdots & & \ddots & \vdots & \vdots \\ -1 & \cdots & -1 & n & -1 \\ -1 & \cdots & -1 & -1 & n \end{bmatrix}$$

The matrix M can be expressed as

$$M = \begin{bmatrix} 2 & 1 & 1 & \cdots & 1 \\ 1 & 2 & 1 & \cdots & 1 \\ \vdots & & \ddots & \vdots & \vdots \\ 1 & \cdots & 1 & 2 & 1 \\ 1 & \cdots & 1 & 1 & 2 \end{bmatrix}$$

Now consider the matrix product MB . The ij entry of this matrix where $i \neq j$ can be expressed as

$$MB_{ij} = \frac{1}{n+1} (-2 + n + (n-2)(-1)) = 0$$

The ii entry can be expressed as

$$MB_{ii} = \frac{1}{n+1} (2n + (n-1)(-1)) = 1$$

This shows that $MB = I$. Now since both M and B are symmetric their product is also symmetric. This implies that

$$I = MB = (MB)^T = B^T M^T = BM$$

Therefore $BM = I$ and $M^{-1} = B$.

- (e) Let x^* be an optimal solutions such that $x_i^* \neq 0$ for all $i \leq p$, and that $x_i^* = 0$ for $p < i \leq n$. This implies that $\lambda_i = 0$ for $i \leq p$. According to the KKT conditions x^* and λ will satisfy

$$\begin{aligned} Mx^* + r - ae - \lambda &= 0 \\ \lambda_i x_i^* &= 0 \quad \forall i. \end{aligned}$$

The first equation can be rewritten as

$$x^* = M^{-1}(ae - r + \lambda).$$

First consider only the $p+1$ to n entries of x^* . I will use the following notation, $x^*[(p+1) : n]$ will represent the vector made up of the $p+1$ to n entries and $M^{-1}[(p+1) : n, :]$ will represent the submatrix consisting of the rows $p+1$ to n and all of the columns. This notation can now be used to solve for λ as follows.

$$\begin{aligned} x^*[(p+1) : n] &= M^{-1}[(p+1) : n, :](ae - r + \lambda) \\ 0 &= M^{-1}[(p+1) : n, :](ae - r) + M^{-1}[(p+1) : n, :]\lambda \\ M^{-1}[(p+1) : n, :](r - ae) &= M^{-1}[(p+1) : n, :]\lambda \end{aligned}$$

Since $\lambda_i = 0$ for $i \leq p$, then

$$M^{-1}[(p+1) : n, :](r - ae) = M^{-1}[(p+1) : n, (p+1) : n]\lambda[(p+1) : n]$$

$M^{-1}[(p+1) : n, (p+1) : n]$ is a square invertible matrix whose inverse is $M[(p+1) : n, (p+1) : n]$, so

$$\lambda[(p+1) : n] = M[(p+1) : n, (p+1) : n]M^{-1}[(p+1) : n, :](r - ae)$$

This product is well defined as $M[(p+1) : n, (p+1) : n] \in \mathbb{R}^{n-p \times n-p}$ and $M^{-1}[(p+1) : n, :] \in \mathbb{R}^{n-p \times n}$ and $r - ae \in \mathbb{R}^n$.

So the full vector λ is the concatenation of the zero vector in \mathbb{R}^p the vector found previously $\lambda[(p+1) : n]$.

Now that we have found the value of λ the solution is exactly

$$x^* = M^{-1}(ae - r + \lambda).$$

4. (a) This problem can be written as a linear programming problem in the following way.

$$\begin{aligned} \max \quad & 3P + 4Q - \theta R \\ \text{subject to} \quad & 2P + 2Q - R = 0 \\ & 2P + Q \leq 10 \\ & P \leq 8 \\ & P, Q, R \geq 0 \end{aligned}$$

The objective function describes the sale price of the products less the cost of the materials. The first constraint enforces the amount of materials required for the products. The second constraint enforces the labor restriction. The third constraint enforces the equipment restriction, and the final constraints enforce nonnegativity. In standard form, linear programs can be written as

$$\begin{aligned} \max \quad & \underline{c}^T \underline{x} \\ \text{subject to} \quad & \underline{Ax} = \underline{b} \\ & \underline{x} \geq 0 \end{aligned}$$

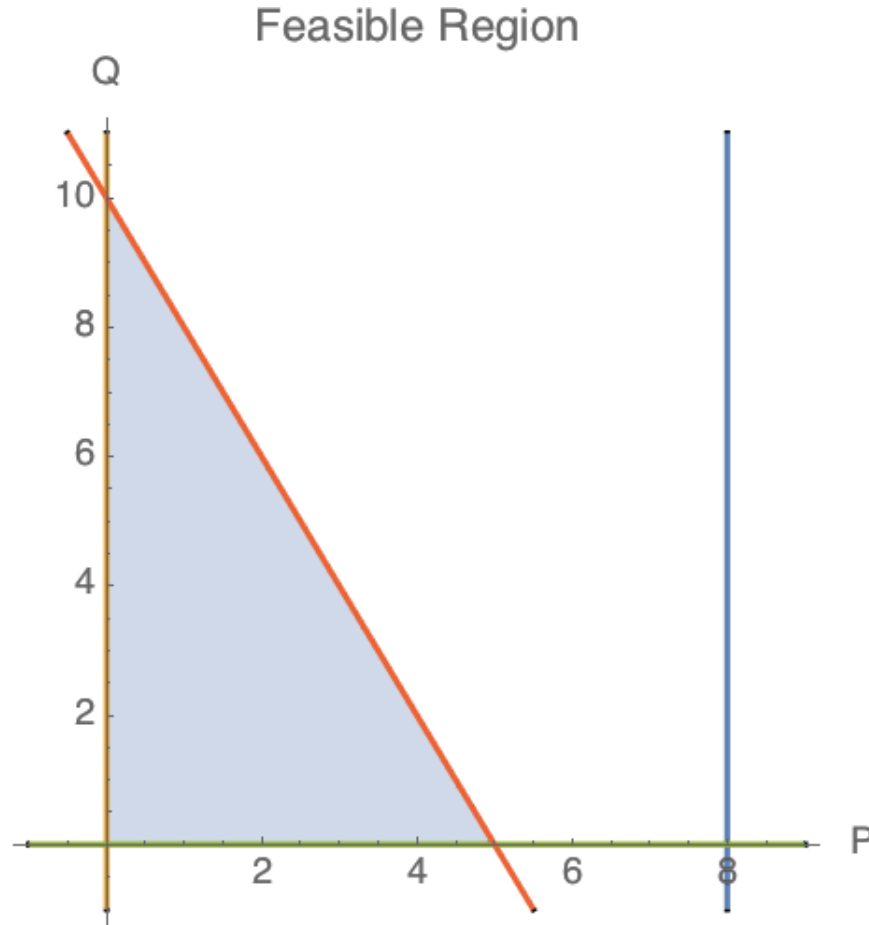
In order to put our linear program in this form I will introduce slack variables, s_1 and s_2 . In this case the linear program can be expressed as

$$\begin{aligned} \max \underline{c}^T \underline{x} \\ \underline{Ax} &= \underline{b} \\ \underline{x} &\geq 0 \end{aligned}$$

where

$$\underline{x} = \begin{bmatrix} P \\ Q \\ R \\ s_1 \\ s_2 \end{bmatrix} \quad \underline{c} = \begin{bmatrix} 3 \\ 4 \\ -\theta \\ 0 \\ 0 \end{bmatrix} \quad \underline{A} = \begin{bmatrix} 2 & 2 & -1 & 0 & 0 \\ 2 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix} \quad \underline{b} = \begin{bmatrix} 0 \\ 8 \\ 10 \end{bmatrix}$$

- (b) This linear program is simple enough to be solved graphically. The following image shows the feasible region for P and Q .



Note that this problem can be expressed in two dimensions as R solely depends on P and Q and doesn't have any restrictions. Once P and Q are determined R can be determined and the value of the objective function can be found. Since the objective function is linear, the optimal solution to this problem must lie on one of the corner points of the feasible region or there will be infinitely many optimal solutions along the border. Therefore I will first consider the 3 corner points of the feasible region. The first corner point $(0,0)$ clearly has zero profit, as $P = 0$ and $Q = 0$ so therefore $R = 0$ and $\underline{c}^T \underline{x} = 0$. At the second corner point $P = 5$ and

$Q = 0$, so $R = 10$. In this case the profit will be $15 - 10\theta$. At the last corner point $Q = 10$ and $P = 0$, thus $R = 20$. In this case the profit will be $40 - 20\theta$. These two profits are linear with respect to θ and they are not parallel, so therefore there must be values of θ such that $40 - 20\theta > 15 - 10\theta$. This can be solved to give an interval on which the profit at $(0, 10)$ is greater.

$$40 - 20\theta > 15 - 10\theta$$

$$35 > 10\theta$$

$$3.5 > \theta$$

When $\theta < 3.5$, then the profit at $Q = 10$ and $P = 0$ is greater. However note that for $\theta > 2$ the profit at $Q = 10$ and $P = 0$ is negative as $40 - 20\theta = 0$ at $\theta = 2$. In this case when $\theta > 2$, then the optimal profit is at $P = 0$ and $Q = 0$. This represents when the raw materials cost more than the value of the finished product. In this case it is best not to make any of the final product. So for $\theta < 2$ the optimal solution is $P = 0$, $Q = 10$, and $R = 20$. For $\theta > 2$ the optimal solution is $P = 0$, $Q = 0$, and $R = 0$.