ECE 6553: Homework #1

Klaus Okkelberg

January 26, 2017

1. (a) Consider vectors $\epsilon v \in \mathbb{R}^m$, $\epsilon \in \mathbb{R}$, $\|v\| = 1$. Notice that ϵ is the magnitude and v is the direction of the vector. The size of the growth of $g(u)$ in the direction of ϵv is given by

$$\begin{aligned} g(u+\epsilon v) - g(u) &= g(u) + \epsilon \frac{\partial g}{\partial u}(u) v + o(\epsilon) - g(u) \\ &= \epsilon \frac{\partial g}{\partial u}(u) v + o(\epsilon) \\ &= \epsilon \frac{\partial g^{\mathrm{T}}}{\partial u}(u) \cdot v + o(\epsilon) \\ &= \epsilon \nabla g(u) \cdot v + o(\epsilon) \\ &= \epsilon \|\nabla g(u)\| \, \|v\| \cos\theta + o(\epsilon) \\ &= \epsilon \|\nabla g(u)\| \cos\theta + o(\epsilon), \end{aligned}$$

where θ is the angle between $\nabla g(u)$ and v. For a given ϵ, we can see that the amount of increase is maximized when $\theta = 0$, i.e. v points along $\nabla g(u)$. Therefore, g grows the most in the direction of $\nabla g(u)$. $\square$

(b) Let $u = r(t)$, $t \in \mathbb{R}$, be the curve that satisfies the constraint $h(u) = 0$, i.e. $h(r(t)) = 0$ $\forall t$. Since the constraint condition is constant, taking the derivative of the constraint gives

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} h(r(t)) &= \frac{\partial h(r(t))}{\partial r(t)} \frac{\mathrm{d}r(t)}{\mathrm{d}t} \\ &= \frac{\partial h(u)}{\partial u} r'(t) \\ &= \nabla h(u) \cdot r'(t) = 0 \end{aligned}$$

Notice that the derivative of the curve $r'(t)$ is the tangent plane to the constraint set at u. Since the dot product of the gradient and $r'(t)$ is 0, the gradient $\nabla g(u)$ is orthogonal to the tangent plane to the constraint set at u. $\square$

2. The Lagrangian is

$$L = (u_1 - 2)^2 + 2(u_2 - 1)^2 + \lambda_1(u_1 + 4u_2 - 3) + \lambda_2(u_2 - u_1)$$

The FONCs are

$$\begin{aligned} \frac{\partial L}{\partial u_1} &= 2(u_1 - 2) + \lambda_1 - \lambda_2 = 0 \\ \frac{\partial L}{\partial u_2} &= 4(u_2 - 1) + 4\lambda_1 + \lambda_2 = 0 \\ \lambda_1(u_1 + 4u_2 - 3) &= 0 \\ \lambda_2(u_2 - u_1) &= 0 \\ u_1 + 4u_2 - 3 &\leq 0 \\ u_2 - u_1 &\leq 0 \\ \lambda_1 &\geq 0 \\ \lambda_2 &\geq 0 \end{aligned}$$

Consider the following combinations in which the constraints can be active/inactive:

(a) $\lambda_1 = \lambda_2 = 0$

$$\begin{cases} 2(u_1 - 2) = 0 \\ 4(u_2 - 1) = 0 \end{cases} \implies \begin{cases} u_1 = 2 \\ u_2 = 1 \end{cases}$$

Since $u_1 + 4u_2 \not\leq 3$, this solution is not feasible.

(b) $u_1 + 4u_2 - 3 = 0$, $\lambda_2 = 0$

$$\begin{cases} 2(u_1 - 2) + \lambda_1 = 0 \\ 4(u_2 - 1) + 4\lambda_1 = 0 \\ u_1 + 4u_2 = 3 \end{cases} \implies \begin{bmatrix} u_1 \\ u_2 \\ \lambda_1 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 4 & 4 \\ 1 & 4 & 0 \end{bmatrix}^{-1} \begin{bmatrix} 4 \\ 4 \\ 3 \end{bmatrix} = \begin{bmatrix} 5/3 \\ 1/3 \\ 2/3 \end{bmatrix}$$

We can verify that the constraints hold:

$$\begin{aligned} u_1 + 4u_2 &\leq 3 \\ u_1 &\geq u_2 \end{aligned}$$

Therefore, $u_1 = 5/3$, $u_2 = 1/3$ is a local minimizer.

(c) $\lambda_1 = 0$, $u_1 = u_2$

$$\begin{cases} 2(u_1 - 2) - \lambda_2 = 0 \\ 4(u_2 - 1) + \lambda_2 = 0 \\ u_1 = u_2 \end{cases} \implies \begin{bmatrix} u_1 \\ u_2 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} 2 & 0 & -1 \\ 0 & 4 & 1 \\ 1 & -1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} 4 \\ 4 \\ 0 \end{bmatrix} = \begin{bmatrix} 4/3 \\ 4/3 \\ -4/3 \end{bmatrix}$$

Since $\lambda_2 < 0$, this solution is not feasible.

(d) $u_1 + 4u_2 - 3 = 0$, $u_1 = u_2$

$$\begin{cases} 2(u_1 - 2) + \lambda_1 - \lambda_2 = 0 \\ 4(u_2 - 1) + 4\lambda_1 + \lambda_2 = 0 \\ u_1 + 4u_2 = 3 \\ u_1 = u_2 \end{cases} \implies \begin{bmatrix} u_1 \\ u_2 \\ \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 1 & -1 \\ 0 & 4 & 4 & 1 \\ 1 & 4 & 0 & 0 \\ 1 & -1 & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} 4 \\ 4 \\ 3 \\ 0 \end{bmatrix} = \begin{bmatrix} 3/5 \\ 3/5 \\ 22/25 \\ -48/25 \end{bmatrix}$$

Since $\lambda_2 < 0$, this solution is not feasible.

Since the optimization problem only has one feasible solution, the global minimum is given by

$$\begin{aligned} u_1^* &= 5/3 \\ u_2^* &= 1/3 \end{aligned}$$

3. (a) The minimization problem is

$$\begin{aligned} &\min \alpha u_1^2 + \beta u_2^2 \\ &\text{s.t. } u_1 + u_2 = q \end{aligned}$$

The Lagrangian is

$$L = \alpha u_1^2 + \beta u_2^2 + \lambda(u_1 + u_2 - q)$$

The FONCs are

$$\begin{aligned}
\frac{\partial L}{\partial u_1} &= 2\alpha u_1 + \lambda = 0 \\
\frac{\partial L}{\partial u_2} &= 2\beta u_2 + \lambda = 0 \\
\frac{\partial L}{\partial \lambda} &= u_1 + u_2 - q = 0
\end{aligned}$$

From the first two conditions, $\alpha u_1 = \beta u_2$. Substituting this into the third condition gives

$$\begin{aligned}
u_1 &= \frac{\beta}{\alpha + \beta} q \\
u_2 &= \frac{\alpha}{\alpha + \beta} q
\end{aligned}$$

(b) It is clear that the highest cost is achieved by only shipping through the more expensive option, i.e.

$$\begin{cases} \max \alpha u_1^2 + \beta u_2^2 \\ \text{s.t. } u_1 + u_2 = q \end{cases} \implies \begin{cases} u_i^* = q, & i = \arg\max(\alpha, \beta) \\ u_j^* = 0, & j = \arg\min(\alpha, \beta) \end{cases}$$

This gives a total cost of $\max(\alpha, \beta) \cdot q^2$. The cost of the minimizers from the previous part is

$$\alpha \left(\frac{\beta}{\alpha + \beta} q \right)^2 + \beta \left(\frac{\alpha}{\alpha + \beta} q \right)^2 = \frac{\alpha\beta}{\alpha + \beta} q^2 < \min(\alpha, \beta) \cdot q^2 \leq \max(\alpha, \beta) \cdot q^2$$

Therefore, the combination found in the previous part is not the worst combination and must be the best combination instead.

4. By convexity of g, for all $u \in \mathbb{R}^m$ and $\alpha \in [0, 1]$,

$$\begin{aligned}
g(\alpha u + (1 - \alpha) u^*) &\leq \alpha g(u) + (1 - \alpha) g(u^*) \\
g(\alpha u + (1 - \alpha) u^*) &\leq g(u^*) + \alpha(g(u) - g(u^*)) \\
g(u) - g(u^*) &\geq \frac{g(\alpha u + (1 - \alpha) u^*) - g(u^*)}{\alpha}
\end{aligned}$$

Using L'Hôpital's rule to take the limit as $\alpha \to 0$,

$$\begin{aligned}
g(u) - g(u^*) &\geq \lim_{\alpha \to 0} \frac{g(\alpha u + (1 - \alpha) u^*) - g(u^*)}{\alpha} \\
&= \left[\frac{\mathrm{d}}{\mathrm{d}\alpha} \left\{ g(\alpha u + (1 - \alpha) u^*) - g(u^*) \right\} \right]_{\alpha = 0} \\
&= \frac{\partial g}{\partial u}(u^*) \left[\frac{\mathrm{d}}{\mathrm{d}\alpha} (\alpha u + (1 - \alpha) u^*) \right]_{\alpha = 0} \\
&= \frac{\partial g}{\partial u}(u^*)(u - u^*) = 0 \cdot (u - u^*) = 0
\end{aligned}$$

Since $g(u) - g(u^*) \geq 0$, or $g(u^*) \leq g(u)$ $\forall u \in \mathbb{R}^m$, u^* is the global minimum to g. $\square$

5. The minimization problem is

$$\gamma_k^{\min} = \arg\min_{\gamma} g(u_k + \gamma d_k)$$

The first and second derivatives of $g(u)$ are

$$\begin{aligned}\frac{\partial g}{\partial u}(u) &= u^{\mathrm{T}} Q + b^{\mathrm{T}} \\ \frac{\partial^2 g}{\partial u^2}(u) &= Q\end{aligned}$$

From the Hessian, the cost is convex and a stationary point is a global minimum. Therefore, we solve the problem by taking the derivative of the cost at k:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\gamma} g(u_k + \gamma d_k) &= \frac{\partial g}{\partial u}(u_k + \gamma d_k) \frac{\mathrm{d}}{\mathrm{d}\gamma}(u_k + \gamma d_k) \\ &= \left[(u_k + \gamma d_k)^{\mathrm{T}} Q + b^{\mathrm{T}}\right] d_k \\ &= \left(u_k^{\mathrm{T}} Q + \gamma d_k^{\mathrm{T}} Q + b^{\mathrm{T}}\right) d_k = 0 \\ \gamma &= \frac{-(u_k^{\mathrm{T}} Q + b^{\mathrm{T}}) d_k}{d_k^{\mathrm{T}} Q d_k}\end{aligned}$$

Using $d_k = -(u_k^{\mathrm{T}} Q + b^{\mathrm{T}})^{\mathrm{T}} = -(Q u_k + b)$,

$$\begin{aligned}\gamma_k^{\min} &= \frac{d_k^{\mathrm{T}} d_k}{d_k^{\mathrm{T}} Q d_k} \qquad (\star) \\ &= \frac{(u_k^{\mathrm{T}} Q + b^{\mathrm{T}})(Q u_k + b)}{(u_k^{\mathrm{T}} Q + b^{\mathrm{T}}) Q (Q u_k + b)}\end{aligned}$$

The form in $(\star)$ was used in problem 6, since it was more convenient.

6. (a)

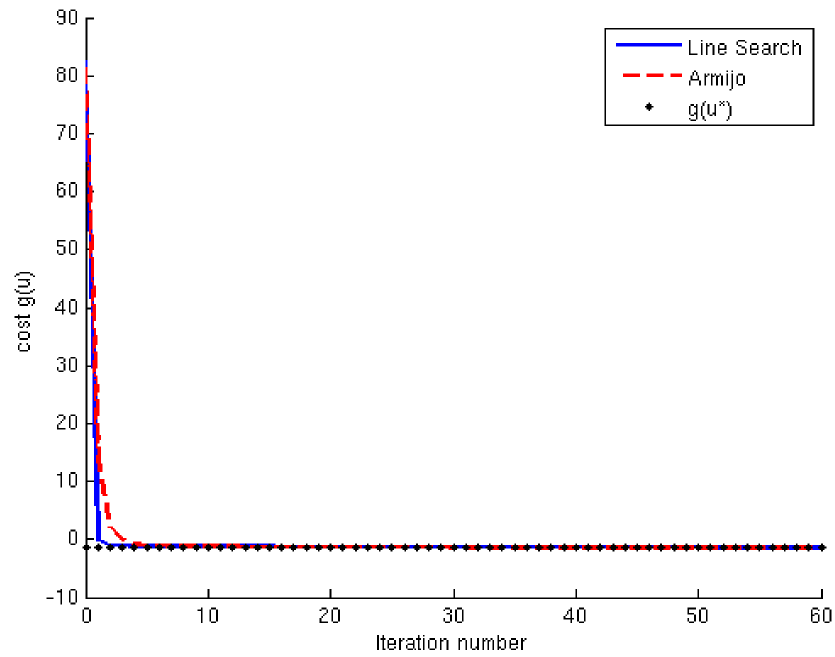

Figure 1: Cost of line search and Armijo vs iteration.

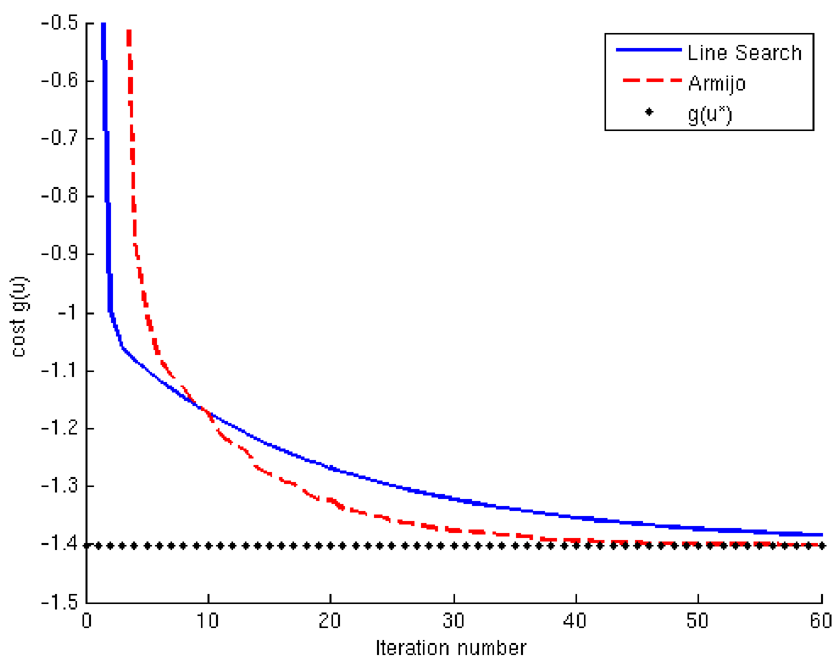

Figure 2: Detailed view of cost of line search and Armijo vs iteration.

```
% Klaus Okkelberg
% ECE 6553
% HW1 P6a

%% Setup

% cost function parameters
Q = [5 0 8 -1 -3;
    0 10 9 7 11;
    8 9 25 0 6;
    -1 7 0 19 5;
    -3 11 6 5 18];
b = [-2 1 -1 3 1]';
% cost function
g = @(u) 0.5*u'*Q*u + b'*u;
d = @(u) -(Q*u + b);

% line search step-size (from P5)
gamma = @(u) d(u)'*d(u) / (d(u)'*Q*d(u));

% common numerical optimization paramters
u0 = [1 1 1 1 1]'; % initial point
N = 60; % number of iterations

% Armijo parameters
alpha = 0.5;
beta = 0.5;

%% Numerical optimization

gk = zeros(N+1,2); % gk(k,:) = cost at iteration k-1
gk(1,:) = g(u0);
% Line search
for k = 1:N
    if k == 1
        u = u0;
    end
    % update
    u = u + gamma(u)*d(u);
    gk(k+1,1) = g(u);
end
% Armijo
for k = 1:N
    if k == 1
        u = u0;
    end
    % find i for step-size
    i = 1;
    while g(u+beta^i*d(u)) - g(u) >= -alpha*beta^i*norm(d(u))^2
        i = i + 1;
    end
    % update
    u = u + beta^i*d(u);
```

```
    gk(k+1,2) = g(u);
end

% global minimizer
ustar = -Q\b;
g_ustar = g(ustar);

figure(1)
set(gcf,'DefaultLineLineWidth',2)
hold on
plot(0:N,gk(:,1),'b')
plot(0:N,gk(:,2),'r--')
plot(0:N,g_ustar,'k.','LineWidth',4)
xlabel('Iteration number')
ylabel('cost g(u)')
legend('Line Search','Armijo','g(u*)')

figure(2)
set(gcf,'DefaultLineLineWidth',2)
hold on
plot(0:N,gk(:,1),'b')
plot(0:N,gk(:,2),'r--')
plot(0:N,g_ustar,'k.','LineWidth',4)
xlabel('Iteration number')
ylabel('cost g(u)')
legend('Line Search','Armijo','g(u*)')
ylim([-1.5 -0.5])
```

(b) While line search is initially faster, Armijo is faster overall.

This is not always the case. For example, for u_0 such that $-\nabla g(u_0)$ is in the direction of the minimizer u^*, line search would reach u^* in one iteration, whereas the Armijo method would need more than one iteration with most α and β.

Even if one method is always better than the other (in terms of iteration count), it likely is more expensive for other costs like computation or memory. It is true that Armijo converges faster, but each iteration of the method is much more costly computation-wise than the line search method, since another minimization problem needs to be solved to determine the step-size β^i.