Homework 1: Dimitri Lezcano

Thursday, September 17, 2020 2:48 PM

1) a) $L(x) = (1-x_1)^2 + 200(x_2 - x_1^2)^2$

$$\nabla L = \begin{pmatrix} -2(1-x_1) + 400(-2x_1)(x_2 - x_1^2) \\ 400(x_2 - x_1^2) \end{pmatrix}$$

$$= \begin{pmatrix} -2 + 2x_1 - 800(x_1 x_2 - x_1^3) \\ 400(x_2 - x_1^2) \end{pmatrix}$$

$\nabla L = 0$ when

$$400(x_2 - x_1^2) = 0$$

$$\Rightarrow x_2 = x_1^2$$

and

$$-2 + 2x_1 - 800(x_1 \cdot x_2^2 - x_1^3) = 0$$

$$\Rightarrow -2 + 2x_1 = 0$$

$$\Rightarrow x_1 = 1 \Rightarrow x_2 = 1^2 = 1$$

so $x^* = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is the critical point

So look at the Hessian:

$$H_L(x) = \begin{pmatrix} \partial^2 L/\partial x_1^2 & \partial^2 L/\partial x_1 \partial x_2 \\ \partial^2 L/\partial x_2 \partial x_1 & \partial^2 L/\partial x_2^2 \end{pmatrix}$$

$$\frac{\partial^2 L}{\partial x^2} = 2 - 800(x_2 - 3x_1^2)$$

$$\frac{\partial^2 L}{\partial x_1 \partial x_2} = \frac{\partial^2 L}{\partial x_2 \partial x_1} = -800 x_1$$

$\partial^2 L$ 400

$$\frac{\partial^2 L}{\partial x_2^2} = 400$$

$$H_L(x^*) = \begin{pmatrix} 2-800(1-3) & -800 \\ -800 & 400 \end{pmatrix}$$

$$H_L(x^*) = \begin{pmatrix} 3202 & -800 \\ -800 & 400 \end{pmatrix}$$

where $\lambda(H_L(x^*)) \hat{=} 3415, 188$

$\therefore \quad H_L \geq 0 \quad \therefore$

x^* is a minima.

b) $L(u) = (u-1)(u+2)(u-3)$

$$\frac{dL}{du} = (u+2)(u-3) + (u-1)(u-3) + (u-1)(u+2)$$

$$= (u^2 - \cancel{u} - 6) + (u^2 - 4u + 3) + (u^2 + \cancel{u} - 2)$$

$dL/du = 3u^2 - 4u - 5 = 0$

when $u = \dfrac{4 \pm \sqrt{16 + 4 \cdot 3 \cdot 5}}{2 \cdot 3}$

$$u = \frac{4 \pm \sqrt{76}}{6}$$

$$= \frac{2}{3} \pm \frac{\sqrt{19}}{3} = u_1^*, u_2^*$$

$$\frac{d^2L}{du^2} = [(u+2) + (u-3)] + [(u-1) + (u-3)] + [(u-1) + (u+2)]$$

$$\frac{d^2L}{du^2} = 6u - 4$$

$$\frac{d^2L}{du^2} = 6u - 4$$

$$\left.\frac{d^2L}{du^2}\right|_{u=u^*} = 6u^* - 4 = 4 \pm 2\sqrt{19} - 4$$
$$= \pm 2\sqrt{19}$$

So $\underline{u_1^* = \frac{2}{3} + \frac{\sqrt{19}}{3} \text{ is a minimum}}$

$\underline{u_2^* = \frac{2}{3} - \frac{\sqrt{19}}{3} \text{ is a maximum}}$

c) $L(u) = (u_1^2 + 3u_1 - 4)(u_2^2 - u_2 + 3)$

$$\nabla L = \begin{pmatrix} (2u_1+3)(u_2^2 - u_2 + 3) \\ (u_1^2 + 3u_1 - 4)(2u_2 - 1) \end{pmatrix} = 0$$

when

$(2u_1+3)(u_2^2 - u_2 + 3) = 0$ &

$(u_1^2 + 3u_1 - 4)(2u_2 - 1) = 0$

So if $u_1 = -3/2 \Rightarrow 2u_2 - 1 = 0 \Rightarrow u_2 = 1/2$

$$u_1^* = \begin{pmatrix} -3/2 \\ 1/2 \end{pmatrix}$$

If $(u_1^2 + 3u_1 - 4) = 0$

$\hookrightarrow (u_1 + 4)(u_1 - 1) = 0$

$\hookrightarrow u_1 = 1, -4.$

Then $(u_2^2 - u_2 + 3) = 0$

Then $(u_2^2 - u_2 + 3) = 0$

$$\Rightarrow u_2 = \frac{1 \pm \sqrt{1 - 4 \cdot 1 \cdot 3}}{2}$$

$$= \frac{1 \pm \sqrt{-11}}{2}$$

$$u_2 = \frac{1}{2} \pm i\frac{\sqrt{11}}{2} \notin \mathbb{R}$$

So only solution (in the reals) is

$$v_1^* = \begin{pmatrix} -3/2 \\ 1/2 \end{pmatrix}$$

$2(9/4 - \frac{9}{2} - 4) = (-\frac{9}{2} - 8)$

Now Hessian

$$\frac{\partial^2 L}{\partial u_1^2} = 2(u_2^2 - u_2 + 3)\Big|_{u=v^*} = 11/2$$

$$\frac{\partial^2 L}{\partial u_2^2} = 2(u_1^2 + 3u_1 - 4)\Big|_{u=v^*} = -25/2$$

$$\frac{\partial^2 L}{\partial u_1 \partial u_2} = (2u_1 + 3)(2u_2 - 1)\Big|_{u=v^*} = 0$$

So $H_L(v^*) = \begin{pmatrix} 11/2 & 0 \\ 0 & -25/2 \end{pmatrix}$

So $v^* = \begin{pmatrix} -3/2 \\ 1/2 \end{pmatrix}$ is a saddle point

2) $L(x) = \frac{1}{2} x^T x$

subject to $x_1 + x_2 + x_3 = 0$

or $e := (1 \ 1 \ 1)^T$

$e^T x = 0$

easy way: $c^T x = 0$ means

$$x \in \text{span}\left\{ \underbrace{\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}}_{v_1}, \underbrace{\begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}}_{v_2} \right\}$$

so $x = av_1 + bv_2$.

Then problem becomes

$L(a, b) = \frac{1}{2} (av_1 + bv_2)^T (av_1 + bv_2)$

w/ constraint satisfied.

$\therefore$

$$L(a, b) = \frac{1}{2} \left(a\|v_1\|^2 + b\|v_2\|^2 + 2ab\, v_1^T v_2 \right)$$
$$= \frac{1}{2} \left(a\sqrt{2} + b\sqrt{2} + 2ab \right)$$
$$= \frac{\sqrt{2}}{2} \left(a + b + \sqrt{2}\, ab \right)$$

So $\nabla_{(a,b)} L = \frac{\sqrt{2}}{2} \begin{pmatrix} 1 + \sqrt{2}b \\ 1 + \sqrt{2}a \end{pmatrix} = 0$

iff

$1 + \sqrt{2}b = 0 \Rightarrow b = -\sqrt{2}/2$

$1 + \sqrt{2}a = 0 \Rightarrow a = -\sqrt{2}/2$

$$H_L(a, b) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$\lambda(H_L) = \pm 1$ so

we have

$$x^* = -\frac{\sqrt{2}}{2} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} - \frac{\sqrt{2}}{2} \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$$ is a saddle point

point

$$x^* = -\frac{\sqrt{2}}{2}\begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}$$

b) $L(u) = (u_1^2 + 3u_1 - 4)(u_2^2 - u_2 + 3)$

s.t. $u_1 - 2u_2 = 0$

$\hookrightarrow u_1 = 2u_2$

$$L(u) = (4u_2^2 + 6u_2 - 4)(u_2^2 - u_2 + 3)$$
$$= 2(2u_2^2 + 3u_2 - 2)(u_2^2 - u_2 + 3)$$
$$= 2\left[2u_2^4 + (-2+3)u_2^3 + (6-2-3)u_2^2 + (9+2)u_2 - 6\right]$$
$$= 2(2u_2^4 + u_2^3 + u_2^2 + 11u_2 - 6)$$

$$\frac{1}{2}\frac{dL}{du} = 8u_2^3 + 3u_2^2 + 2u_2 + 11 = 0$$

when $u_2 \approx -1.168$, $\underline{0.397 \pm 1.01i}$

not real

so $u_2 = -1.168$ and

$u_1 = \frac{1}{2}u_2 = -0.584$

$$u^* = \begin{pmatrix} -0.584 \\ -1.168 \end{pmatrix}$$

$$\left.\frac{d^2L}{du_2^2}\right|_{u_2 = u_2^*} = \left. 24u_2^2 + 6u_2 + 2 \right|_{u_2 = u_2^*} = 27$$

so u^* is a minimum!

so u^* is a minimum!

3) a)

$$\min L(x,u) = \tfrac{1}{2} x^T Q x + u^T R u + s^T x$$
$$\text{s.t. } f(x,u) = Ax + Bu + c = 0$$

$x \in \mathbb{R}^n$, $u \in \mathbb{R}^m$, $Q \geq 0$, $R > 0$,
$A \in M_n(\mathbb{R})$, $B \in M_{n,m}(\mathbb{R})$,
$s, c \in \mathbb{R}^n$,

i) Necessary and Sufficient conditions

$$H(x,u,\lambda) \doteq L(x,u) + \lambda^T f(x,u)$$

$$= \tfrac{1}{2} x^T Q x + u^T R u + s^T x + \lambda^T (Ax + Bu + c)$$

(1st order)

$\nabla_\lambda H = f(x,u) = 0$ (given by constraint)

$$\nabla_x H = \tfrac{1}{2}(Q + Q^T)x + s + A\lambda = 0$$

$$\nabla_u H = \tfrac{1}{2}(R + R^T)u + B^T\lambda = 0$$

$\because R, Q \geq 0$, $R = R^T$ & $Q = Q^T \therefore$

- $\nabla_x (c^T x) = c$
- $\nabla_x (x^T C x) = \tfrac{1}{2}(C + C^T)x$

$$\nabla_x H = Qx + s + A^T\lambda = 0$$

$$\nabla_u H = Ru + B^T\lambda = 0$$

$\hookrightarrow$ ① $\underline{u = -R^{-1}B^T\lambda}$ (R^{-1} exists b/c $|R| \neq 0$ b/c $R > 0$)

and we have

② $\underline{A^T\lambda = -Qx - s}$

$$L = H - \lambda^T f$$

$$dL = dH - \lambda^T df = 0 \text{ @ optimum}$$

Hessian (x,u)

$$dL \approx (\partial_x H, \partial_u H)\begin{pmatrix} dx \\ du \end{pmatrix} + \frac{1}{2}\begin{pmatrix} dx \\ du \end{pmatrix}^T H_H(x,u)\begin{pmatrix} dx \\ du \end{pmatrix} - \lambda^T df$$

(to 2nd order)

by constraint $f = 0 \Rightarrow df = 0$

$\Rightarrow df = \partial_x f dx + \partial_u f du = A dx + B du$

By 1st order condition

$$(\partial_x H, \partial_u H)\begin{pmatrix} dx \\ du \end{pmatrix} = 0$$

So

$$dL \approx \frac{1}{2}\begin{pmatrix} dx \\ du \end{pmatrix}^T \begin{pmatrix} H_{xx} & H_{xu} \\ H_{ux} & H_{uu} \end{pmatrix}\begin{pmatrix} dx \\ du \end{pmatrix}$$

$$H_{xx} = Q$$

$$H_{xu} = H_{ux} = 0$$

$$H_{uu} = R$$

$$dL = \frac{1}{2}\begin{pmatrix} dx \\ du \end{pmatrix}^T \begin{pmatrix} Q & 0 \\ 0 & R \end{pmatrix}\begin{pmatrix} dx \\ du \end{pmatrix}$$

③ $= \frac{1}{2} dx^T Q dx + \frac{1}{2} du^T R du \geq 0$

③ $= \frac{1}{2} dx^T Q dx + \frac{1}{2} du^T R du \geq 0$ which is true! $\because Q \geq 0$

s.t. $A dx = -B du$ (from $df = 0$) $R \geq 0$

①, ②, ③ are the necessary and sufficient conditions. (really only need ① + ②)

ii) Assume A full rank

① $u = -R^{-1} B^T \lambda$

② $A^T \lambda = -Qx - S$

③ $dx^T Q dx + du^T R du$ s.t.

$A dx = -B du$

$A^{-T} \doteq (A^T)^{-1}$

② $\Rightarrow \lambda = -A^{-T} Q x - S$

① + ② $\Rightarrow u = -R^{-1} B^T \lambda$

$= -R^{-1} B^T (-A^{-T} Q x - A^{-T} S)$

④ $u = R^{-1} B^T A^{-T} (Qx + S)$

④ + constraint $\Rightarrow$

$A x + B u + C = 0$

$A x + B (R^{-1} B^T A^{-T})(Qx + S) + C = 0$

$(A + B R^{-1} B^T A^{-T} Q) x = -(B R^{-1} B^T A^{-T} S + C)$

$A(I + \underbrace{A^{-1} B R^{-1} B^T A^{-T} Q}) x = -(B R^{-1} B^T A^{-T} S + C)$

$$C = A^{-1} B R^{-1} B^T A^{-T} Q$$

Note that $(A^{-1} B R^{-1} B^T A^{-T})^T$

$$= A^{-1} B R^{-T} B^T A^{-T}$$

$$= A^{-1} B R^{-1} B^T A^{-T} \quad \because R^{-T} = R^{-1}$$

$$\Rightarrow \quad = U \text{ where } U = U^T \qquad \because R^{-1} \geq 0$$

$$C = U \cdot Q \qquad \therefore U \geq 0 \qquad \because R \geq 0$$

and since $U, Q \geq 0$, $C \geq 0$.

Thus $\exists\, V \in O(n)$, $D \geq 0$ diagonal s.t.

$C = V D V^T$ $\therefore$

$$(I + C) = (I + V D V^T) = V (I + D) V^T$$

and $(I + C)^{-1} = V (I + D)^{-1} V^T$

where $(I + D)^{-1} = \begin{pmatrix} \frac{1}{1+\lambda_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{1+\lambda_n} \end{pmatrix}$

$\lambda_i \in \sigma(C) \geq 0$.

Therefore:

$$A(I + C)x = -(B R^{-1} B^T A^{-T} s + c)$$

$$(I + C)x = -(U s + A^{-1} c)$$

$$\underline{x = -(I + C)^{-1} (U s + A^{-1} c)}$$

and finally

$$u = -R^{-1} B^T A^{-T} (Q x + s)$$

$$\underline{u = -R^{-1} B^T A^{-T} s + R^{-1} B^T A^{-T} (I + C)^{-1} (U s + A^{-1} c)}$$

b) $\min \left(L(y) = \frac{1}{2} y^T M y + k^T y \right)$

s.t.

s.t.

$$f(y) = Ay + c = 0$$

$y \in \mathbb{R}^n$, $M \succeq 0$, $A \in \mathbb{R}^{m \times n}$ for $m<n$ full rank,
$k \in \mathbb{R}^n$, $c \in \mathbb{R}^m$ $\Rightarrow$ $A^T \in \mathbb{R}^{n \times m}$ full rank $m<n$

$f \in \mathbb{R}^m$ $\quad rk(A) = rk(A^T) = m$

$c \in \mathbb{R}^n = \mathbb{R}^{(n-m)+m}$

$$\nabla_y L(y^*) = -\sum_{i=1}^{m} \lambda_i \nabla_y f_i(y^*) \in \mathbb{R}^n$$

$$\nabla_y L(y^*) = My + k \in \mathbb{R}^n$$

$$\left(\sum_i \lambda_i \nabla f_i\right)_j = \sum_{i=1}^{m} \lambda_i \frac{\partial f_i}{\partial x_j}$$

$$= \sum_{i=1}^{m} \frac{\partial f_i}{\partial x_j} \lambda_i$$

$$\frac{\partial f_i}{\partial x_j} = \frac{\partial}{\partial x_j} (Ax + c)_i$$

$$= \frac{\partial}{\partial x_j} (Ax)_i$$

$$= \frac{\partial}{\partial x_j} \left(\sum_k A_{ik} x_k\right)$$

$$= A_{ij}$$

$\Rightarrow$

$$\left(\sum_i \lambda_i \nabla f_i\right)_j = \sum_i \lambda_i \frac{\partial f_i}{\partial x_j}$$

$$= \sum_i A_{ij} \lambda_i$$

$$= \sum_j A_{ij} \lambda_i$$
$$= A^T \lambda$$

so $M y^* + K = -A^T \lambda$

$$y^* = -M^{-1} A^T \lambda - M^{-1} K$$

by constraint $A y^* + C = 0$

$$-A M^{-1} A^T \lambda - M^{-1} K + C = 0$$
$$A M^{-1} A^T \lambda = C - M^{-1} K$$

$\therefore A, M$ full rank, $A M^{-1} A^T$ is full rank so

$$\lambda = (A M^{-1} A^T)^{-1} (C - M^{-1} K)$$

so $\underline{y^* = -M^{-1} A^T (A M^{-1} A^T)^{-1} (C - M^{-1} K) - M^{-1} K}$

$H_L = M \succ 0$ $\therefore$ for all critical points y_c, y_c is a global minimum.

$\therefore$ since y^* is a critical point

$\underline{y^* \text{ is a global minimum}}$

4) a) Changing the step-size causes for gradient descent to be unstable and overshoot and go to ∞. the smaller the step size, the more stable.

b) Newton's Method is <u>MUCH</u> more efficient than gradient Descent

c) Again Newton's method converges much faster than gradient descent.

d) $\min \ L(x,u) = x^2 + 4x + 8u^2$

s.t. $f(x,u) = x - 7u + 5 \le 0$

$y = \begin{pmatrix} x \\ u \end{pmatrix}$

unconstrained:

$$\nabla L = \begin{pmatrix} 2x+4 \\ 16u \end{pmatrix} = 0 \quad \text{if} \quad \begin{pmatrix} x \\ u \end{pmatrix} = \begin{pmatrix} -2 \\ 0 \end{pmatrix} = y^*$$

$$f(y^*) = -2 - 7 \cdot 0 + 5 = 3 \not\le 0 .$$

So

$$\nabla L = -\lambda \nabla f$$

$$\begin{pmatrix} 2x+4 \\ 16u \end{pmatrix} = -\lambda \begin{pmatrix} 1 \\ -7 \end{pmatrix}$$

$$16u = 7\lambda$$

$$\lambda = -2x - 4$$

$$16u = 7(-2x-4)$$

① $14x + 16u = -28$ and constraint

$$x - 7u + 5 = 0$$

② $x - 7u = -5$

① + ② $\Rightarrow$

$$\begin{pmatrix} 14 & 16 \\ 1 & -7 \end{pmatrix} \begin{pmatrix} x \\ u \end{pmatrix} = \begin{pmatrix} -28 \\ -5 \end{pmatrix}$$

$$y^* = \begin{pmatrix} x \\ u \end{pmatrix} \approx \begin{pmatrix} -2.4211 \\ 0.3684 \end{pmatrix}$$

5) $f^k(d) = f(x^k) + \nabla f(x^k)^T d + \frac{1}{2} d^T \nabla^2 f(x^k) d$

$$\min \; f^k(d) \qquad \text{(solution is } d^k\text{)}$$
$$\text{s.t.} \; \|d\| \le \gamma^k$$

for $\gamma^k > 0$

show above equivalent to solving

$$(\nabla^2 f(x^k) + \delta^k I) d^k = -\nabla f(x^k)$$

first $\|d\| \le \gamma_k \equiv \|d\|^2 \le \gamma_k^2$

$\equiv d^T d \le \gamma_k^2$

constraint $f(d) = d^T d - \gamma_k^2 \le 0$

If $f(d^*) \le 0$, then

$\nabla_d f^k(d) = \nabla f(x^k) + \nabla^2 f(x^k) d = 0$

if $\underline{(\nabla^2 f + 0 I) d = -\nabla f}$

If $f(d^*) > 0$,

$H(d, \lambda) = f^k(d) + \lambda f(d)$

$= f(x^k) + \nabla f(x^k)^T d + \frac{1}{2} d^T \nabla^2 f(x^k) d + \lambda(d^T d - \gamma_k^2)$

$\nabla_d H = \nabla f + \nabla^2 f d + \lambda d = 0$

$\Rightarrow (\nabla^2 f + \lambda I) d = -\nabla f$

$$\Rightarrow (\nabla^2 f + tI) d = -\nabla f$$

Solution $d = d^k$ of problem, in otherwords d^k is

$$(\nabla^2 f + \delta^k I) d^k = -\nabla f$$

where we can think of δ^k as either 0 for a direct minimization of the problem or a small perturbation to the direct minimization to ensure that we satisfy the constraints, since we know $\nabla^2 f$ is symmetric, it perturbs the eigen values of $\nabla^2 f$ to conserve the constraint.

$\because$ $\nabla^2 f$ symmetric $\exists\, U \in O(n)$, D diagonal s.t. $\nabla^2 f = U D U^T$, Then

$$(\nabla^2 f + \delta^k I) = (U D U^T + \delta^k I)$$

$$= U \underbrace{(D + \delta^k I)} U^T$$

Small perturbations of eigenvalues of Hessian in order to maintain constraint.