Math 450/558
Fall 2015
Midterm 1
Solution Sheet

Question 1 This question concerns a quasi-Newton method based on rank two symmetric inverse Hessian updates of the form

$$H_{k+1} = H_k + u s_k^T + s_k u^T. \tag{1}$$

Above, H_{k+1} is an approximation for the inverse Hessian $\left[\nabla^2 f(x^{(k+1)})\right]^{-1}$ and $s_k := x^{(k+1)} - x^{(k)}$.

(a) (10 points) Derive an update rule of the form (1) such that H_{k+1} satisfies the secant equation.

Right-Multiply both sides by y_k

$$\underbrace{H_{k+1} y_k}_{s_k} = H_k y_k + u s_k^T y_k + \underbrace{s_k^{\cancel{T}} u^T y_k}_{s_k y_k^T u}$$

$$u = \left[s_k^T y_k I + s_k y_k^T \right]^{-1} \left[s_k - H_k y_k \right]$$

Note that

$$\left[s_k^T y_k I + s_k y_k^T \right]$$

is invertible, unless $s_k^T y_k = 0$ (which can be avoided for instance by imposing Wolfe conditions), i.e.,

$$\left[s_k^T y_k I + s_k y_k^T \right] x = 0$$

$$\Rightarrow (x s_k^T) y_k = -(s_k y_k^T) x$$

$$\Rightarrow x = \alpha s_k$$

$$\Rightarrow \alpha (s_k s_k^T) y_k = -\alpha (s_k s_k^T) y_k$$

$$\Rightarrow \underset{(\text{that is } x=0)}{\alpha = 0} \quad \text{or} \quad s_k^T y_k = 0$$

Update rule (assumes $s_k^T y_k \neq 0$)

$$H_{k+1} = H_k + \left[s_k^T y_k I + s_k y_k^T \right]^{-1} \left[s_k - H_k y_k \right] s_k^T + s_k \left[s_k - H_k y_k \right]^T \left[s_k^T y_k I + s_k y_k^T \right]^{-1}$$

(b) (10 points) Consider a line search method outlined below. Does this

---

Choose an estimate $x^{(0)}$ for a local minimizer.
Calculate $\nabla f(x^{(0)})$, and choose a positive definite H_0.
for $k = 0, 1, 2, 3, \dots$ **do**
  (1) $p_k \leftarrow -H_k \cdot \nabla f(x^{(k)})$
  (2) Perform a line search in the direction p_k to determine a step-length α_k such that $x^{(k)} + \alpha_k p_k$ satisfies the Wolfe conditions.
  (3) $x^{(k+1)} \leftarrow x^{(k)} + \alpha_k p_k$, and calculate $\nabla f(x^{(k+1)})$.
  (4) Form H_{k+1} from H_k, $s_k = x^{(k+1)} - x^{(k)}$ and $y_k = \nabla f(x^{(k+1)}) - \nabla f(x^{(k)})$ using the update rule (1).
end for

---

algorithm guarantee the generation of descent search directions p_k, i.e., does p_k necessarily satisfy $f(x^{(k)} + \alpha p_k) < f(x^{(k)})$ for all $\alpha > 0$ small enough? Explain.

H_{k+1} does not have to be positive definite, even if H_k is positive definite. Consequently,

$$\nabla f(x^{(k+1)})^T p_{k+1} = -\nabla f(x^{(k+1)})^T H_{k+1} \nabla f(x^{(k+1)}) \geq 0$$

is possible. This means p_{k+1} does not have to be descent.

To see H_{k+1} does not have to be PD, note that $s_k^T y_k \approx 0$ (even though $s_k^T y_k \neq 0$) implies $[s_k^T y_k I + s_k y_k^T]$ is nearly singular. In this case u can have very large norm, and for instance u can also be nearly orthogonal to s_k. Letting $\hat{u} := u/\|u\|$

$$(\hat{u} - s_k)^T H_{k+1} (\hat{u} - s_k) = \underbrace{(\hat{u} - s_k)^T H_k (\hat{u} - s_k)}_{\leq \|H_k\|_2 \|\hat{u} - s_k\|^2} + \underbrace{(\hat{u} - s_k)^T u s_k^T (\hat{u} - s_k)}_{\approx -\|u\| \|s_k\|^2} + \underbrace{(\hat{u} - s_k)^T s_k u^T (\hat{u} - s_k)}_{\approx -\|s_k\|^2 \|u\|}$$

$$\lesssim \|H_k\|_2 (1 + \|s_k\|^2) - 2\|u\| \|s_k\|^2 < 0.$$

(To see u can nearly be orthogonal to s_k, $w := s_k - H_k y_k$ can nearly be orthogonal to s_k and y_k. But then

$$[s_k^T y_k I + s_k y_k^T] u = w \implies u \approx w / s_k^T y_k \text{)}$$

Question 2 (25 points) Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is a twice continuously differentiable function, and x_* is a local minimizer of f such that $\nabla^2 f(x_*)$ is *not invertible*. Furthermore, suppose that there exist a ball $B(x_*, \delta)$ with positive radius $\delta > 0$ such that $\nabla^2 f(x)$ is invertible at all $x \in B(x_*, \delta)\backslash\{x_*\}$, and that the limit

$$\lim_{x \to x_*} \left\| \left[\nabla^2 f(x)\right]^{-1} \right\|_2 \|x - x_*\|$$

exists and is finite.

Prove that there exists a ball $B(x_*, \eta)$ (with positive radius $\eta > 0$) such that, for all $x^{(0)} \in B(x_*, \eta)$, pure Newton's method for unconstrained optimization generates a sequence $\{x^{(k)}\}$ satisfying

$$\lim_{k \to \infty} x^{(k)} = x_*.$$

Let

$$L := \lim_{x \to x_*} \left\| \left[\nabla^2 f(x)\right]^{-1} \right\|_2 \|x - x_*\|.$$

Convergence can be deduced under the assumption that $\nabla^2 f(x)$ is Lipschitz continuous, that is

$$\|\nabla^2 f(x) - \nabla^2 f(y)\|_2 \leq \gamma \|x - y\| \quad \forall x, y$$

with Lipschitz constant $\gamma < \frac{2}{L}$ (γ is any positive real number if $L = 0$).

Since $L < \frac{2}{\gamma}$, there exists an ϵ and a ball $B(x_*, \mu)$ such that

$$\left\| \left[\nabla^2 f(x)\right]^{-1} \right\|_2 \|x - x_*\| \leq \frac{2}{\gamma} - \epsilon \quad \forall x \in B(x_*, \mu)$$

where $\mu < \delta$. By Taylor's thm with integral remainder

$$\nabla f(x_*) = \nabla f(x^{(k)}) + \int_0^1 \nabla^2 f(x^{(k)} + t(x_* - x^{(k)})) \times (x_* - x^{(k)})\, dt$$

$$0 = \underbrace{\left[\nabla^2 f(x^{(k)})\right]^{-1} \nabla f(x^{(k)})}_{-p_k = x^{(k)} - x^{(k+1)}} + (x_* - x^{(k)}) + \left[\nabla^2 f(x^{(k)})\right]^{-1} \int_0^1 \left[\nabla^2 f(x^{(k)} + t(x_* - x^{(k)})) - \nabla^2 f(x^{(k)})\right](x_* - x^{(k)})\, dt$$

$$\Longrightarrow (x^{(k+1)} - x_*) = \left[\nabla^2 f(x^{(k)})\right]^{-1} \int_0^1 \left[\nabla^2 f(x^{(k)} + t(x_* - x^{(k)})) - \nabla^2 f(x^{(k)})\right](x_* - x^{(k)})\, dt$$

$$\Longrightarrow \|x^{(k+1)} - x_*\| \leqslant \|[\nabla^2 f(x^{(k)})]^{-1}\|_2 \, \|x^{(k)} - x_*\| \int_0^1 \gamma t \, \|x^{(k)} - x_*\| \, dt$$

$$(*) \qquad = \left(\frac{\gamma}{2} \|[\nabla^2 f(x^{(k)})]^{-1}\| \, \|x^{(k)} - x_*\| \right) \|x^{(k)} - x_*\|$$

Now, suppose $x^{(0)} \in B(x_*, \mu)$. We first prove $x^{(k)} \in B(x_*, \mu)$ for $k = 1, 2, 3, \ldots$ by induction. Suppose $x^{(k)} \in B(x_*, \mu)$ as the inductive hypothesis. It follows from (*) that

$$\|x^{(k+1)} - x_*\| \leqslant \underbrace{\frac{\gamma}{2} \left(\frac{2}{\gamma} - \epsilon \right)}_{\rho < 1} \|x^{(k)} - x_*\|,$$

so $x^{(k+1)} \in B(x_*, \mu)$.

Consequently

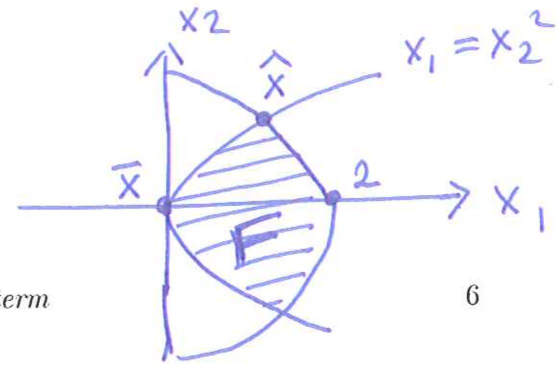
$$\|x^{(k)} - x_*\| \leqslant \rho^k \|x^{(0)} - x_*\|,$$

taking the ~~limist~~ limi as $k \to \infty$ yields

$$\lim_{k \to \infty} \|x^{(k)} - x_*\| = 0$$

$$\Longrightarrow \lim_{k \to \infty} x^{(k)} = x_*.$$

$\square$



Question 3 Consider the inequality constrained optimization problem

$$\begin{aligned} \text{minimize}_{x\in\mathbb{R}^2} \quad & \tfrac{1}{2}(x_1-1)^2 - x_1 - x_2 \\ \text{subject} \quad & x_1 - x_2^2 \geq 0 \\ & 2 - x_1^2 - x_2^2 \geq 0, \end{aligned} \qquad (2)$$

and the points $\bar{x} = (0,0)$ and $\hat{x} = (1,1)$.

(a) (5 points) Write down a feasible sequence $\{z^{(k)}\}$ leading to $\hat{x}$. Write down also a limiting direction associated with this feasible sequence.

The sequence $\{(1, 1-1/k)\}$ is feasible and leads to $(1,1)$, i.e.,

(i) $(1, 1-1/k) \in F$ for $k = 1,2,3,\dots$

$$\underbrace{1}_{x_1} - \underbrace{\left(1-\tfrac{1}{k}\right)^2}_{x_2^2} = \frac{2}{k} - \frac{1}{k^2} \geq 0$$

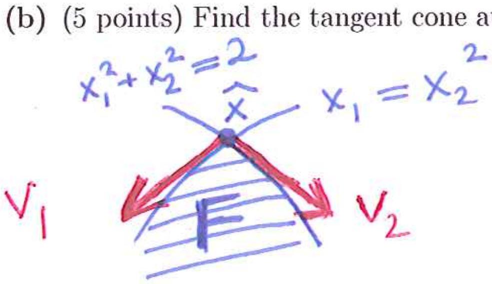
$$2 - 1^2 - \left(1-\tfrac{1}{k}\right)^2 = \frac{2}{k} - \frac{1}{k^2} \geq 0$$

(ii) $(1, 1-1/k) \neq (1,1)$ for $k = 1,2,3,\dots$

(iii) $\lim_{k\to\infty} (1, 1-1/k) = (1,1)$.

Limiting direction: $\lim_{k\to\infty} \dfrac{(1,1-1/k)-(1,1)}{\|(1,1-1/k)-(1,1)\|} = (0,-1)$

(b) (5 points) Find the tangent cone at $\hat{x}$.



$$V_1 = \left.\frac{d(x_2^2, -x_2)}{dx_2}\right|_{x_2=1} = (-2,-1)$$

$$V_2 = \left.\frac{d(x_1, \sqrt{2-x_1^2})}{dx_1}\right|_{x_1=1} = (1,-1)$$

$$T(\hat{x}) = \{\alpha_1 V_1 + \alpha_2 V_2 \mid \alpha_1, \alpha_2 \geq 0\}$$

$$= \left\{ \alpha_1 \begin{bmatrix} -2 \\ -1 \end{bmatrix} + \alpha_2 \begin{bmatrix} 1 \\ -1 \end{bmatrix} \,\middle|\, \alpha_1, \alpha_2 \geq 0 \right\}$$

(c) (20 points) Show, for each one of $\bar{x}$ and $\hat{x}$, either the point is a local minimizer of (2) or not a local minimizer of (2).

$$\nabla f(x) = \begin{bmatrix} x_1 - 2 \\ -1 \end{bmatrix} \quad \nabla c_1(x) = \begin{bmatrix} 1 \\ -2x_2 \end{bmatrix} \quad \nabla c_2(x) = \begin{bmatrix} -2x_1 \\ -2x_2 \end{bmatrix}$$

At $\bar{x}$,

$$A(\bar{x}) = \{1\}$$

$$\nabla f(\bar{x}) = \begin{bmatrix} -2 \\ -1 \end{bmatrix} \quad \nabla c_1(\bar{x}) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \left(\begin{array}{l} \text{LICQ holds,} \\ \text{i.e.,} \\ \nabla c_1(\bar{x}) \neq 0 \end{array} \right)$$

since $\nabla f(\bar{x}) \neq \lambda \nabla c_1(\bar{x}) \quad \forall \lambda$

by first order necessary conditions $\bar{x}$ is not a local minimizer.

At $\hat{x}$

$$A(\hat{x}) = \{1, 2\}$$

$$\nabla f(\hat{x}) = \begin{bmatrix} -1 \\ -1 \end{bmatrix} \quad \nabla c_1(\hat{x}) = \begin{bmatrix} 1 \\ -2 \end{bmatrix} \quad \nabla c_2(\hat{x}) = \begin{bmatrix} -2 \\ -2 \end{bmatrix}$$

$$\nabla f(\hat{x}) = \lambda_1 \nabla c_1(\hat{x}) + \lambda_2 \nabla c_2(\hat{x}) \implies \lambda_1 = 0, \quad \lambda_2 = 1/2.$$

Moreover,

$$\nabla_x \mathcal{L}(\hat{x}, \lambda) = \begin{bmatrix} x_1 - 2 \\ -1 \end{bmatrix} + \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad \nabla^2_{xx} \mathcal{L}(\hat{x}, \lambda) = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$$

Since, $\nabla^2_{xx} \mathcal{L}(\hat{x}, \lambda)$ is PD,

$$d^T \nabla^2_{xx} \mathcal{L}(\hat{x}, \lambda) d > 0 \quad \forall d \in S_2 \setminus \{0\}.$$

By second order sufficient conditions $\hat{x}$ is a local minimizer.

Question 4 Let us focus on the following linearly constrained nonlinear program.

$$\begin{array}{ll} \text{minimize}_{x\in\mathbb{R}^n} & f(x) \\ \text{subject} & Ax = b \\ & x \geq 0 \end{array} \qquad (3)$$

Above, $A \in \mathbb{R}^{m\times n}$ is a given matrix, $b \in \mathbb{R}^m$ is a given vector, and the objective function $f(x)$ is twice continuously differentiable.

(a) (15 points) Suppose that $\nabla^2 f(x)$ is *positive definite* at all x, and let x_* be a point such that the following hold for some λ_* and $s_* \geq 0$:

$$Ax_* = b, \quad x_* \geq 0, \quad x_*^T s_* = 0 \text{ and } \nabla f(x_*) = A^T\lambda_* + s_*.$$

Prove that x_* is a global minimizer of (3).

Defining

$$\mathcal{L}(x,\lambda,s) = f(x) - (-b+Ax)^T\lambda - x^T s$$

we have

$$\nabla_x \mathcal{L}(x_*,\lambda_*,s_*) = \nabla f(x_*) - A^T\lambda_* - s_* = 0.$$

For each z s.t. $Az=b$ and $z\geq 0$ we have

$$\mathcal{L}(z,\lambda_*,s_*) = f(z) - z^T s_* \leq f(z)$$

On the other hand, by a Taylor expansion of $\mathcal{L}(z,\lambda_*,s_*)$ about x_*

$$\mathcal{L}(z,\lambda_*,s_*) = \underbrace{\mathcal{L}(x_*,\lambda_*,s_*)}_{f(x_*)} + \underbrace{\nabla_x \mathcal{L}(x_*,\lambda_*,s_*)}_{0}(z-x_*) + \frac{1}{2}(z-x_*)^T \underbrace{\nabla^2_{xx}\mathcal{L}(x_*+t(z-x_*),\lambda_*,s_*)}_{\nabla^2 f(x_*+t(z-x_*))}(z-x_*)$$

$$> f(x_*)$$

Consequently, for each feasible z we have $f(z) > f(x_*)$. $\square$

(b) (10 points) Suppose that $\mathrm{Null}(A) \neq \{0\}$, and that Z is a matrix whose columns form an orthonormal basis for $\mathrm{Null}(A)$. Furthermore, suppose $Z^T \nabla^2 f(x) Z$ is *not positive semi-definite* at all x. Let x_* be a point such that the following hold for some λ_*:

$$Ax_* = b, \quad x_* > 0 \text{ and } \nabla f(x_*) = A^T \lambda_*.$$

Prove that x_* is not a local minimizer of (3).

It suffices to find a $d \in \mathbb{R}^n$ such that

$$\text{(i)} \quad A(x_* + \alpha d) = b$$

$$\text{(ii)} \quad (x_* + \alpha d) \geq 0$$

yet

$$f(x_* + \alpha d) < f(x_*)$$

for all $\alpha > 0$ small enough.

Since $Z^T \nabla^2 f(x_*) Z$ is not PSD, it has a negative eigenvalue. Let v be a corresponding eigenvector, define $d := Zv$. Condition (ii) trivially holds for all small α, because $x_* > 0$.

Furthermore,

$$A(x_* + \alpha d) = \underbrace{Ax_*}_{b} + \alpha \underbrace{AZv}_{0} = b.$$

Condition (i) also holds. Define

$$\mathcal{L}(x, \lambda_*) := f(x) - (Ax - b)^T \lambda_*$$

Note $\nabla_x \mathcal{L}(x_*, \lambda_*) = \nabla f(x_*) - A^T \lambda_* = 0$. Furthermore,

$$\mathcal{L}(x_* + \alpha d, \lambda_*) = f(x_* + \alpha d)$$

$$= \underbrace{\mathcal{L}(x_*, \lambda_*)}_{f(x_*)} + \underbrace{\nabla_x \mathcal{L}(x_*, \lambda_*)(\alpha d)}_{0} + \frac{1}{2}\alpha^2 \underbrace{d^T \nabla^2_{xx} \mathcal{L}(x_*, \lambda_*) d}_{<0} + o(\alpha^2)$$

Consequently, $f(x_* + \alpha d) < f(x_*)$ as desired. $\square$