RAMP-UP "DATA SCIENCE"
PART: CONTINUOUS OPTIMIZATION

CHRISTIAN KIRCHES

1. Unconstrained Optimization: Quick Questions

(1) Are the following sets convex?
- $A = \{x \in \mathbb{R}^2 \mid 0 \leqslant x_1 \leqslant 4,\, 2 \leqslant x_2 \leqslant 6\}$,
- $B = \{x \in \mathbb{R}^3 \mid x_1 + x_2 + 2x_3 \leqslant 24\}$,
- $C = \{x \in \mathbb{R}^n \mid x = \alpha y, \alpha \in \mathbb{R},\, y \in X \subset \mathbb{R}^n \text{ convex}\}$.

(2) Are the following functions convex?
- $f\colon \mathbb{R}^2 \to \mathbb{R}$, $f(x) = 3x_1^2 - 2x_2^2$,
- $f\colon \mathbb{R}^2 \to \mathbb{R}$, $f(x) = (x_1-3)^3 + (x_2+1)^2$,
- $f\colon \mathbb{R}^3 \to \mathbb{R}$, $f(x) = e^{x_1} + x_2^2 + x_3^2$,
- $f\colon \mathbb{R}^3 \to \mathbb{R}$, $f(x) = e^{2x_1} + x_2^2 x_3$,
- $f\colon \mathbb{R}^2 \to \mathbb{R}$, $f(x) = x_1 x_2$.

(3) Define 'global convergence' and 'local convergence'

(4) Define 'global', 'local', and 'strict local minima' of a nonlinear optimization problem

(5) Give the necessary and sufficient optimality conditions of first and second order for unconstrained optimization

(6) Define the direction of steepest descent for a function f in a point $x^{(k)}$

(7) Define linear, superlinear, and quadratic convergence

(8) What are the convergence properties of gradient descent in unconstrained optimization?

(9) What is the Armijo condition in line search? How does it work?

(10) When is the Newton direction also a direction of descent?

(11) Which of the following statements are true, which ones are false?
The function $f\colon \mathbb{R}^3 \to \mathbb{R}$ with $f(x) = (x_1-2)^2 + (x_2-3)^2 + (x_3-1)^2$ has
- a global minimum,
- a global maximum,
- a saddle point

in $x^* = (2,3,1)^T$,

(12) Consider the unconstrained minimization problem

$$\min_{x\in\mathbb{R}^n} \frac{1}{2} x^T H x + c^T x$$

with indefinite Hessian H. Describe the set of all local minima.

2. Taylor's Theorem

Consider

$$f\colon \mathbb{R}^2 \to \mathbb{R}, \quad f(x) = e^{-(x_1^2+x_2^2)}.$$

Find a quadratic approximation of f in the location $(\bar{x}_1, \bar{x}_2) = (1, -1)$ using Taylor's theorem.
Now consider

$$f\colon \mathbb{R}^2 \to \mathbb{R}, \quad f(x) = \frac{x_1 - x_2}{x_1 + x_2}.$$

Again find a quadratic approximation of f in the location $x_0 = (\bar{x}_1, \bar{x}_2) = (1, 1)$ using Taylor's theorem.

Date: February 2, 2023.

3. Stationarity A

Find all stationary points of the following functions

- $f\colon \mathbb{R}^2 \to \mathbb{R}$, $f(x) = 2x_1^2 + x_1x_2 + 2x_2 - 4x_1 - x_2$,
- $f\colon \mathbb{R}^2 \to \mathbb{R}$, $f(x) = x_1 \ln(x_2 + 1)$,
- $f\colon \mathbb{R}^2 \to \mathbb{R}$, $f(x) = x_1^2 - x_1x_2^2 + x_2^4 - 3x_2x_3 + x_3^3$.

4. Stationarity B

Consider $f\colon \mathbb{R}^2 \to \mathbb{R}$ with $f(x,y) = x^2 + 2y^4 - 3xy^2$. For every point $(x_0, y_0) \in \mathbb{R}^2 \backslash \{(0,0)\}$ let

$$\varphi\colon \mathbb{R} \to \mathbb{R},\ \varphi(t) = f(tx_0, ty_0)$$

be the restriction of f onto the line through both the origin and the point (x_0, y_0).

(1) Show that φ has a local minimum in $t = 0$ for all $(x_0, y_0) \in \mathbb{R}^2 \backslash \{(0,0)\}$.
(2) Show or give a counterexample: $(0,0)$ is a local minimum of f.

5. Stationarity C

Find all stationary points of the unconstrained minimization problem.

$$\min_{x \in \mathbb{R}^2} 100(x_2 - x_1^2)^2 + (1 - x_1)^2.$$

Which ones are local, which ones are global minima?

6. Stationarity D

Show that the unconstrained minimization problem

$$\min_{x \in \mathbb{R}^2} x_1^2 - 2x_2^2 + 8x_1 + 12x_2$$

has exactly one stationary point, and show it to be a saddle point.

7. Stationarity E

Let $\varepsilon > 0$, let $a < b$, and let $\gamma \in \mathcal{C}^1([a,b], \mathbb{R}^n)$ be a $\mathcal{C}^1$-path through $y \in \mathbb{R}^n$ with $\gamma(0) = y$. Let $(-\varepsilon, \varepsilon) \subset [a,b]$, let $f \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R})$, and let y be chosen such that $\nabla f(y) \neq 0$ holds. Moreover, let

$$f(\gamma(t)) = f(y) \quad \text{for all } t \in (-\varepsilon, \varepsilon).$$

Now show that $\nabla f(y)$ is a vector orthogonal to the tangent of γ in the location $t = 0$.
Draw a sketch of the situation before trying a proof.

8. Rates of Convergence

Consider the sequences $\{x^{(k)}\}$ defined by

$$x^{(k)} = \begin{cases} \left(\frac{1}{4}\right)^{2^k}, & k \text{ even} \\ \frac{x^{(k-1)}}{k}, & k \text{ odd} \end{cases}.$$

Is this sequence linearly convergenct?
Is it superlinearly convergent?
Is it quadratically convergent?

9. Newton's Method for Root Finding

Consider the nonlinear root finding problem

$$\begin{aligned} x_1 x_2 + x_1 - x_2 &= 1 \\ x_1 x_2 &= 0 \end{aligned}$$

First, find the exact solutions of the system of equations
Afterwards, solve the problem using Newton's method for the initial guesses

(1) $x^{(0)} = \begin{pmatrix} 2 \\ -1 \end{pmatrix}$
(2) $x^{(0)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

10. Steepest Descent and Newton Directions

Let $F \in \mathcal{C}(\mathbb{R}^n, \mathbb{R}^n)$ such that ∇F is an invertible matrix. Let $x \in \mathbb{R}^n$ and let $d = -\nabla F(x)^{-T} F(x)$ be the Newton direction for F in x. Furthermore, let $A \in \mathbb{R}^{n \times n}$ be an invertible matrix and let $f\colon \mathbb{R}^n \to \mathbb{R}$ with $f(x) = \|AF(x)\|^2$.

(1) Compute the direction of steepest descent for f in x.
(2) Compute the Newton direction $-\left(\nabla^2 f(x)\right)^{-1} \nabla f(x)$ if $F(x) = 0$.
(3) Show that d is a descent direction for f in x.

11. Newton's Method for Root Finding

Apply Newton's method to the following nonlinear root finding methods. Use the initial guess $x^{(0)} = 1$.

(1) $f\colon \mathbb{R} \to \mathbb{R}$ with $f(x) = |x|$,
(2) $f\colon \mathbb{R} \to \mathbb{R}$ with $f(x) = |x|^2$,
(3) $f\colon \mathbb{R} \to \mathbb{R}$ with $f(x) = |x-1|^2$

To do so, write Newton's iteration for the functions f explicitly and determine the sequence of iterates.

12. Newton's Method for Quadratic Minimization Problems

Consider $f\colon \mathbb{R}^n \to \mathbb{R}$ quadratic,

$$f(x) = \frac{1}{2} x^T Q x + c^T x$$

with Q positive definite. Show that Newton's method using the step length $\alpha = 1$ applied to the minimization problem $\min_{x \in \mathbb{R}^n} f(x)$ finds the minimum in a single iteration.

Solutions to Computational Exercises

1. Unconstrained Optimization: Quick Questions

Def. A set M is convex iff for all pairs $x, y \in M$ it holds that $tx + (1-t)y \in M$ for all $t \in [0,1]$.

(1) (a) A: Yes. A square. To see this, verify for all $x, y \in A$ and all $t \in [0,1]$ that

$$0 \leqslant tx_1 + (1-t)y_1 \leqslant 4$$

and

$$2 \leqslant tx_2 + (1-t)y_2 \leqslant 6.$$

(b) B. Yes.
A single linear inequality partions $\mathbb{R}^3$ into two convex halves.
To see this, verify that for all $x, y \in B$ and all $t \in [0,1]$ we have

$$\begin{aligned} & tx_1 + (1-ty_1) + tx_2 + (1-t)y_2 + 2(tx_3 + (1-t)y_3) \\ = {} & t(x_1 + x_2 + 2x_3) + (1-t)(y_1 + y_2 + 2t + 3) \\ \leqslant {} & t24 + (1-t)24 = 24. \ q.e.d. \end{aligned}$$

(c) C. No. (I made a mistake in class here)
Pick $y_1 \neq y_2 \in X$ and consider the points $x_1 = \alpha_1 y_1 \in C$ with $\alpha_1 = 1$ and $x_2 = \alpha_2 y_2 \in C$ with $\alpha_2 = -1$.
If C is convex, then $\frac{1}{2}x_1 + \frac{1}{2}x_2 \in C$ must hold (for any pair x_1, x_2 and any t, but in particular for the pair chosen here and $t = \frac{1}{2}$). Now

$$\tfrac{1}{2}x_1 + \tfrac{1}{2}x_2 = \tfrac{1}{2}y_1 - \tfrac{1}{2}y_2 = \tfrac{1}{2}(y_1 - y_2).$$

So $\alpha = \frac{1}{2}$ and $y_1 - y_2 \in X$ must hold by definition of C. But convexity of X is not a guarantee that this is the case.

(2) (a) No. The Hessian is $\begin{pmatrix} 6 & 0 \\ 0 & -4 \end{pmatrix}$ and indefinite.

(b) No, not everywhere. The Hessian is $\begin{pmatrix} 6(x_1-3) & 0 \\ 0 & 2 \end{pmatrix}$ and can be made indefinite if $x_1 < 3$.

(c) Yes. The Hessian is $\begin{pmatrix} e^{x_1} & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$ and since $e^{x_1} > 0$ for all $x_1 \in \mathbb{R}$, this matrix is positive definite everywhere on $\mathbb{R}^3$.

(d) No. The Hessian is $\begin{pmatrix} 4e^{2x_1} & 0 & 0 \\ 0 & 2x_3 & 2x_2 \\ 0 & 2x_2 & 0 \end{pmatrix}$.
Its eigenvalues are $4e^{2x_1} > 0$ and $x_3 \pm \sqrt{x_3^2 + 4x_2^2}$, which can be made negative.

(e) No. The Hessian is $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ with eigenvalues $+1$ and -1, so it is indefinite.

(3) $\bar{x} \in \mathbb{R}^n$ is called a feasible point of the NLP $\min f(x)$ s.t. $g(x) = 0$, $h(x) \geqslant 0$ iff $\bar{x} \in \mathcal{F} := \{x \in \mathbb{R}^n \mid g(x) = 0, h(x) \geqslant 0\}$ holds, i.e., $\bar{x}$ lies in the feasible set $\mathcal{F}$.

(4) $\bar{x} \in \mathbb{R}^n$ is called a local minimum of the NLP $\min f(x)$ s.t. $g(x) = 0$, $h(x) \geqslant 0$ iff there exists $\varepsilon > 0$ such that $f(x) \geqslant f(\bar{x})$ for all $x \in \mathcal{B}_\varepsilon(\bar{x})$ holds. It is strict local minimum if the inequality holds in its strict form, $f(x) > f(\bar{x})$. It is a global minimum if $f(x) \geqslant f(\bar{x})$ for all $x \in \mathcal{F}$ holds.

(5) Unconstrained NOC 1.O.: Let $\bar{x} \in \mathbb{R}^n$ be a minimum of the unconstrained problem $\min f(x)$. Then $\nabla f(\bar{x}) = 0$ holds.
Unconstrained NOC 2.O.: Let $\bar{x} \in \mathbb{R}^n$ be a minimum of the unconstrained problem $\min f(x)$. Then $\nabla f(\bar{x}) = 0$ and $\nabla^2 f(\bar{x})$ positive semidefinite hold.

Unconstrained SOC 2.O.: If $\bar{x} \in \mathbb{R}^n$ satisfies $\nabla f(\bar{x}) = 0$ and $\nabla^2 f(\bar{x})$ positive definite, then $\bar{x}$ is a local minimum of the unconstrained problem $\min f(x)$.

(6) $d \in \mathbb{R}^n$ is called a direction of steepest descent at $\bar{x} \in \mathbb{R}^n$ for the problem $\min f(x)$, iff $d = -\lambda \nabla f(\bar{x})$ for some $\lambda > 0$. Equivalently, d solves the problem $\min \nabla f(\bar{x})^T d$ s.t. $||d||_2 = 1$.

(7) See the essentials' sheet for the definitions.

(8) Gradient descent has local q-linear convergence. If Armijo line search is used for selecting step sizes, it is globally convergent. Descent directions are perpendicular to tangents onto the level sets at the iterates. This leads to zig-zagging of the descent path, which becomes worse if the eigenvalues of the Hessian are spread out.

(9) The Armijo condition asks for a step size α to satisfy

$$f(x^{(k)} + \alpha d) \leqslant f(x^{(k)}) + \gamma \alpha \nabla f(x^{(k)})^T d$$

wherein $x^{(k)}$ is the current iterate of a descent method and d is a descent direction. $\gamma \in (0, 1]$ is a tuning factor. The condition asks for the descent in terms of the function values to be proportional to the (negative) slope of the function into direction d.

(10) The Newton direction $d_N = -\nabla^2 f(x^{(k)})^{-1} \nabla f(x^{(k)}$ is a descent direction if its angle with the steepest descent direction $-\nabla f(x^{(k)}$ is acute:

$$-d_N^T \nabla f(\bar{x}) > 0.$$

One usually checks ... $\geqslant \eta > 0$ for some $\eta \in (0, 1)$.

(11) f is a convex quadratic, the Hessian is two times the identity. The point x^* is stationary. Hence
- true
- false
- false

(12) If H is indefinite, there is at least a negative and a positive eigenvalue. Moving along the eigenvector associated with the negative eigenvalue is a second order descent direction into which the objective is unbounded from below. Hence, the set of local minima is empty.
Remark: Saddle points x^* satisfy $Hx^* = -c$ and if H is regular, there is a unique saddle point.

2. Taylor's Theorem

This is a function in $\mathbb{R}^2$ so we need to use the vector calculus variant of Taylor's theorem. Compute the first and second order derivatives, and insert the point given. This tests the rule for derivatives of the exponential.

$$\begin{aligned} f(\bar{x} + d) &\approx f(\bar{x}) + \nabla f(\bar{x})^T d + \tfrac{1}{2} d^T \nabla^2 f(\bar{x}) d \\ &= e^{-2} + \left(-2e^{-2}, 2e^{-2}\right) d + \tfrac{1}{2} d^T \begin{pmatrix} 2e^{-2} & -4e^{-2} \\ -4e^{-2} & 2e^{-2} \end{pmatrix} d \end{aligned}$$

For the second function, the same. This tests the rule for derivatives of quotients.

$$\begin{aligned} f(\bar{x} + d) &\approx f(\bar{x}) + \nabla f(\bar{x})^T d + \tfrac{1}{2} d^T \nabla^2 f(\bar{x}) d \\ &= 0 + \left(1/2, -1/2\right) d + \tfrac{1}{2} d^T \begin{pmatrix} -1/2 & 0 \\ 0 & 1/2 \end{pmatrix} d \end{aligned}$$

3. Stationarity A

A stationary point x^* is defined by $\nabla f(x^*) = 0$.
- $0 = \nabla f(x) = \begin{pmatrix} 4x_1 + x_2 - 4 \\ x_1 + 1 \end{pmatrix}$, so $x^* = (-1, 8)$.
- $0 = \nabla f(x) = \begin{pmatrix} \ln(x_2 + 1) \\ \frac{x_1}{x_2 + 1} \end{pmatrix}$, so $x^* = (0, 0)$.

- $0 = \nabla f(x) = \begin{pmatrix} 2x_1 - x_2^2 \\ -2x_1x_2 + 4x_2^3 - 3x_3 \\ 3x_2 + 3x_3^2 \end{pmatrix}$.
From the first entry, $x_1 = \frac{1}{2}x_2^2$. From the third entry, $x_2 = -x_3^2$. Substituting for x_1 and x_2 in the second entry, cancelling terms and dividing by 3 yields $x_3^6 = -x_3$, which results in $x_3 \in \{0, -1\}$. Inserting into the first and third entry yields the two stationary points $x^* = (0,0,0)^T$ and $x^* = (1/2, -1, -1)^T$.

4. Stationarity B

(1) To prove the first statement we show the NOC 1.O. $\phi'(0) = 0$. Remember to *first* differentiate $\phi(t)$ with respect to t, *then* insert.

$$\begin{aligned}
\phi'(t) &= \nabla f(tx_0, ty_0)^T \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} \\
&= \begin{pmatrix} 2(tx_0) - 3(ty_0)^2 \\ 8(ty_0)^3 - 6t^2x_0y_0 \end{pmatrix}^T \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} \\
&= 2tx_0^2 - 3t^2x_0y_0^2 + 8t^3y_0^4 - 6t^2x_0y_0^2
\end{aligned}$$

In $t = 0$ we now immediately see $\phi'(t) = 0$. Now show the SOC 2.O. (which doesn't have to hold in a minimum, but does here). We obtain

$$\begin{aligned}
\phi''(t) = (\phi'(t))' &= \begin{pmatrix} x_0 & y_0 \end{pmatrix} \nabla^2 f(tx_0, ty_0) \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} \\
&= \begin{pmatrix} x_0 & y_0 \end{pmatrix} \begin{pmatrix} 2 & -6ty_0 \\ -6ty_0 & 24t^2y_0^2 - 6tx_0 \end{pmatrix} \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}.
\end{aligned}$$

In $t = 0$ this evaluates to $2x_0^2 \geqslant 0$. This is NOC 2.O. only and doesn't yet prove our statement. If $y_0 = 0$ then $x_0 \neq 0$ since $(x_0, y_0) \neq (0,0)$, and $2x_0^2 > 0$ follows. This is SOC 2.O. and a local minimum is present.

If however $y_0 > 0$, then $x_0 = 0$ would be permitted and only $2x_0^2 \geqslant 0$ follows. But then, the function f reduces to $f = 2(ty_0)^4$ which is convex in t, and so the NOC 1.O. is also a sufficient condition for a local minimum.
Together, this proves that a local minimizer is present in $t = 0$ regardless of the choice of (x_0, y_0).
q.e.d.

(2) A counterexample is the curve $(x, y) = (\frac{3}{2}t^2, t)$ for $t \in \mathbb{R}$, which is not a ray covered by (1). Insert this curve into f to find the values f assumes along the curve:

$$\left(\frac{3}{2}t^2\right)^2 + 2t^4 - 3\left(\frac{3}{2}t^2\right)t^2 = \left(\frac{9}{4} + 2 - \frac{9}{2}\right)t^4 = -\frac{1}{4}t^4 \leqslant 0$$

for all $t \in \mathbb{R}$. Starting in $(0,0)^T$ on the curve, we may move into either direction along the curve and the function values decrease. Thus $(0,0)^T$ is *not* a local minimum of f, even though it is a local minimum along any ray through $(0,0)^T$.
Any number in the open interval $(1,2)$ produces a counterexample; I picked 3/2 here.

5. Stationarity C

$$\begin{aligned}
f(x) &= 100(x_2 - x_1^2)^2 + (1 - x_1)^2 \\
\nabla f(x) &= \begin{pmatrix} -400(x_2x_1 - x_1^3) + 2x_1 - 2 \\ 200(x_2 - x_1^2) \end{pmatrix}
\end{aligned}$$

Stationary points satisfy $\nabla_x f(x^*) = 0$, hence $x_2 = x_1^2$ from the second row and $x_2 = \frac{200x_1^3+x_1-1}{200x_1}$ from the first row. This yields the equation

$$\frac{200x_1^3 + x_1 - 1}{200x_1} = x_1^2$$

which simplifies to $x_1 - 1 = 0$, hence $x_1 = 1$ and $x_2 = 1$. The only stationary point is $x^* = (1,1)^T$. The objective is zero, and since the function is a sum of two quadratics, this is the global minimum.

6. Stationarity D

$$f(x) = x_1^2 - 2x_2^2 + 8x_1 + 12x_2$$
$$\nabla f(x) = \begin{pmatrix} 2x_1 + 8 \\ -4x_2 + 12 \end{pmatrix}$$

The only root clearly is $x^* = (-4, 3)$. The Hessian is

$$\nabla^2 f(x) = \begin{pmatrix} 2 & 0 \\ 0 & -4 \end{pmatrix}$$

(everywhere) and has eigenvalues $+2$ and -4, hence is indefinite. This makes x^* a saddle point.

7. Stationarity E

Compute using the chain rule

$$\frac{df}{dt} = \frac{df}{d\gamma}\frac{d\gamma}{dt} = \nabla f(\gamma(t))^T \dot{\gamma}(t).$$

Since f is constant along the curve $\gamma(t)$, the derivative must be zero. Hence

$$\nabla f(\gamma(t))^T \dot{\gamma}(t) = 0.$$

This means ∇f is perpendicular to the tangent of γ in the point $y = \gamma(t)$.

8. Rates of Convergence

Obviously the even indexed subsequence has zero as a limit. The odd indexed sequence members are based on the even ones, so the limit of the entire sequence is zero.
From the definition of linear convergence for even k

$$\frac{x^{(k+1)} - 0}{x^{(k)} - 0} = \frac{(\frac{1}{4})^{2^k}/k}{(\frac{1}{4})^{2^k}} = \frac{1}{k} < 1$$

for $k \geqslant 2$. This is linear convergence with rate $1/k$ in iteration $k \geqslant 2$.
Since the rate itself approaches zero as $k \to \infty$, this is even superlinear convergence.
From the definition of quadratic convergence for even k we need find C finite such that we have

$$\frac{x^{(k+1)} - 0}{(x^{(k)} - 0)^2} = \frac{(\frac{1}{4})^{2^k}/k}{((\frac{1}{4})^{2^k})^2} = \frac{1}{k(\frac{1}{4})^{2^k}} = \frac{4^{2^k}}{k} < C < \infty$$

for all $k \geqslant k_0$. This is not possible as 4^{2^k} diverges much faster than $1/k$ converges.

9. Newton's Method for Root Finding

As a root finding problem, the situation reads

$$F(x) = \begin{pmatrix} x_1x_2 + x_1 - x_2 - 1 \\ x_1x_2 \end{pmatrix} = 0.$$

The exact solutions are

$$x^* = (0, -1),\ x^* = (1, 0).$$

Newton's method reads $x^{(k+1)} \leftarrow x^{(k)} - \alpha^{(k)} d^{(k)}$, where $d^{(k)} = \nabla F(x^{(k)})^{-T} F(x^{(k)})$ and we use $\alpha^{(k)} = 1$ here. The gradient is

$$\nabla F(x)^T = \begin{pmatrix} x_2 + 1 & x_1 - 1 \\ x_2 & x_1 \end{pmatrix}.$$

To find its inverse, remember

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

With this, the inverse is

$$\nabla F(x)^{-T} = \frac{1}{(x_2+1)x_1 - (x_1-1)x_2} \begin{pmatrix} x_1 & 1 - x_1 \\ -x_2 & x_2 + 1 \end{pmatrix} = \frac{1}{x_1 + x_2} \begin{pmatrix} x_1 & 1 - x_1 \\ -x_2 & x_2 + 1 \end{pmatrix}.$$

First starting point:

$$x^{(0)} = (2, -1)^T \qquad d^{(0)} = \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ -2 \end{pmatrix} = (2, 0)^T$$

$$x^{(1)} = (0, -1)^T \qquad F(x^{(1)}) = (0, 0) \text{ and we stop successfully.}$$

Second starting point:

$$x^{(0)} = (1, 1)^T \qquad d^{(0)} = \begin{pmatrix} \frac{1}{2} & 0 \\ -\frac{1}{2} & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = (0, 1)^T$$

$$x^{(1)} = (1, 0)^T \qquad F(x^{(1)}) = (0, 0) \text{ and we stop successfully.}$$

10. Steepest Descent and Newton Directions

(1) Formally, the direction of steepest descent in a point $\bar{x} \in \mathbb{R}^n$ is given by the solution of

$$\min_d \nabla f(\bar{x})^T d \text{ s.t. } ||d||_2 = 1.$$

Here $f(x) = ||AF(x)||_2^2 = (AF(x))^T(AF(x)) = F(x)^T A^T AF(x)$ for some invertible matrix A. We compute the required gradient using the product and chain rules, keeping $A^T A$ as a constant factor, and taking care of proper transpositions,

$$\begin{aligned} \nabla f(\bar{x}) &= \nabla F(x) \cdot A^T A \cdot F(x) + [F(x)^T \cdot A^T A \cdot \nabla F(x)^T]^T \\ &= 2\nabla F(x) A^T AF(x) \end{aligned}$$

Hence solve (we may take the square in the norm constraint)

$$\min_d d^T [2\nabla F(x) A^T AF(x)] \text{ s.t. } d^T d = 1.$$

With the Lagrangian

$$\mathcal{L}(d, \lambda) = d^T [2\nabla F(x) A^T AF(x)] - \lambda(d^T d - 1)$$

the NOC 1.O. for the optimal (d^*, λ^*) read

(1) Stationarity: $\nabla_d \mathcal{L}(d^*, \lambda^*) = 0$
(2) Feasibility: ${d^*}^T d^* = 1$

We compute

$$0 = \nabla_d \mathcal{L}(d^*, \lambda^*) = 2\nabla F(x) A^T A F(x) - 2\lambda d^*$$

Hence, a steepest descent directions for f in x satisfy

$$d^* = \frac{1}{\lambda} \nabla F(x) A^T A F(x).$$

To satisfy $d^{*T} d^* = 1$, choose $\lambda = \pm ||\nabla F(x) A^T A F(x)||_2$. For the choice $-$, we obtain the minimum value, hence the steepest descent.

(2) Since $F(x) = 0$ leads to $\nabla f(x) = 0$ (see the expression we derived above), it is unnecessary to compute $\nabla^2 f(x)$ in the first place. The Newton direction will be $-(\nabla^2 f(x))^{-1} \nabla f(x) = 0$.

(3) There was a mixup in the original derivative notation this question used. $\nabla F(x)$ has as many columns as F has components, and as many rows as x has unknowns. Then $\nabla F(x)^T d = -F(x)$ is the root-finding Newton equation for F (Note that the minimizing Newton equation $\nabla^2 f(x) d = -\nabla f(x)$ is never affected by such issues since $\nabla^2 f(x)$ is symmetric). Then $d = -\nabla F(x)^{-T} F(x)$ (this is capital F) must be shown to satisfy

$$d^T \nabla f(x) < 0.$$

Hence compute

$$\begin{aligned} 0 > & - [\nabla F(x)^{-T} F(x)]^T [2\nabla F(x) A^T A F(x)] \\ = & - 2F(x)^T \nabla F(x)^{-1} \nabla F(x) A^T A F(x) \\ = & - 2||AF(x)||_2^2. \end{aligned}$$

This is obviously true since norms are positive definite.

11. Newton's Method for Root Finding

(1) $f(x) = |x|$.
$f'(x) = \text{sgn}(x)$, $x^{(k+1)} \leftarrow x^{(k)} - |x^{(k)}|/\text{sgn}(x^{(k)}) = x^{(k)} - x^{(k)} = 0$.
Start in $x^{(0)} = 1$ yields $x^{(1)} = 0$, which is the root. This is true for any initial value.

(2) $f(x) = |x|^2$. This is smooth, since $|x|^2 = x^2$.
$f'(x) = 2x$, $x^{(k+1)} \leftarrow x^{(k)} - \frac{x^{(k)2}}{2x^{(k)}} = x^{(k)}/2$.
Start in $x^{(0)} = 1$ yields $x^{(k)} = 1/2^k$. This converges to the root 0, but not finitely.

(3) $f(x) = |x - 1|^2$.
$f'(x) = 2|x - 1|\text{sgn}(x - 1)$, $x^{(k+1)} \leftarrow x^{(k)} - \frac{|x^{(k)}-1|^2}{2|x^{(k)}-1|\text{sgn}(x^{(k)}-1)} = x^{(k)} - (x^{(k)} - 1)/2 = (x^{(k)} + 1)/2$.
Start in $x^{(0)} = 1$ yields $x^{(1)} = 1$ and stays in the root.

12. Newton's Method for Quadratic Minimization Problems

We have $\nabla f(x) = Qx + c$ and $\nabla^2 f(x) = Q$. Newton's method applied to this problem reads

$$x^{(k+1)} \leftarrow x^{(k)} - Q^{-1}(Qx^{(k)} + c) = x^{(k)} - x^{(k)} - Q^{-1}c = -Q^{-1}c.$$

Hence, no matter how x^0 is chosen, $x^k = -Q^{-1}c$ for all $k \geqslant 1$ and the method stays there.
The NOC 1.O. for a minimum is $\nabla f(x^*) = 0$, hence, $Qx^* + c = 0$. But this is solved by $x^{(k)}$ for all $k \geqslant 0$. Since Q is positive definite, the NOC 1.O. is also sufficient.