1. Prove that if P is positive definite, then $\log(\det(P^{-1}))$ is a convex function of P.

(a) $\log(\det(P^{-1})) = \log(-\det(P)) = -\log(\det(P))$

(4 Marks) Since $\log(\det(P))$ is concave, $-\log(\det(P))$ is convex!

Solution to the non-inverse case is already there at http://www.cse.iitb.ac.in/~CS709/notes/eNotes/basicsOfUnivariateOptAndItsGeneralisation-withHighlights.pdf. Only difference is that I have introduced the inverse of the matrix.

(b) You can consider instead: λ_i are eigenvalues of $P^{1/2} V P^{1/2}$

2. Consider the following L4 norm approximation problem

$$\underset{\mathbf{x}}{\text{minimize}} ||A\mathbf{x} - \mathbf{b}||_4$$

where

$$||A\mathbf{x} - \mathbf{b}||_4 = \left(\sum_{i=1}^{m} (\mathbf{a}_i^T \mathbf{x} - b_i)^4 \right)^{\frac{1}{4}}$$

The matrix $A \in \Re^{m \times n}$ (with rows $\mathbf{a}_i^T$) and the vector $\mathbf{b} \in \Re^m$ are given. Express this minimization problem as a constrained convex optimisation problem with:

(a) convex quadratic objective

(b) convex quadratic inequality constraints

(c) linear equality constraints

(4 Marks)

Ans:

$$\text{minimize} \quad \sum_{i=1}^{m} z_i^2$$

$$\text{s.t.} \quad a_i^T x - b_i = y_i \qquad i = 1 \ldots m$$

$$y_i^2 \leq z_i \qquad i = 1 \ldots m$$

3. In a quasi-Newton algorithm, $B^{(k+1)}$ (approximation to the Hessian) is obtained from a positive denite matrix $B^{(k)}$ from the previous iteration on k, by using the Davidon-Fletcher-Powell (DFP) updating formula, which is specified below:

$$B^{(k+1)} = B^{(k)} + \frac{\Delta\mathbf{x}^{(k)} \left(\Delta\mathbf{x}^{(k)}\right)^T}{\left(\Delta\mathbf{x}^{(k)}\right)^T \Delta\mathbf{g}^{(k)}} - \frac{B^{(k)}\Delta\mathbf{g}^{(k)} \left(\Delta\mathbf{g}^{(k)}\right)^T B^{(k)}}{\left(\Delta\mathbf{g}^{(k)}\right)^T B^{(k)}\Delta\mathbf{g}^{(k)}}$$

Show that the condition

$$\left(\Delta\mathbf{x}^{(k)}\right)^T \Delta\mathbf{g}^{(k)} > 0$$

will ensure that $B^{(k+1)}$ is positive denite.

You can assume that values of $\Delta\mathbf{x}^{(k)}$, $B^{(k)}$ and $\Delta\mathbf{g}^{(k)}$ are known from the previous iteration (on k). As stated in the class (in the context of the BFGS algorithm),

(a) $\Delta\mathbf{x}^{(k)} = -B^{(k)}\nabla f(\mathbf{x}^{(k)})$.

(b) $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + t^{(k)}\Delta\mathbf{x}^{(k)}$, where $t^{(k)}$ can be obtained using any method such as line search, *etc.*

(c) $\Delta\mathbf{g}^{(k)} = \nabla f(\mathbf{x}^{(k+1)}) - \nabla f(\mathbf{x}^{(k)})$.

(8 Marks)

<u>Ans:</u>

$$x^T B^{(k+1)} x = x^T B^{(k)} x - \frac{x^T B^{(k)} \Delta g^{(k)} \left(\Delta g^{(k)}\right)^T B^{(k)} x}{\left(\Delta g^{(k)}\right)^T B^{(k)} \Delta g^{(k)}} + \frac{x^T \Delta x^{(k)} \left(\Delta x^{(k)}\right)^T x}{\left(\Delta x^{(k)}\right)^T \Delta g^{(k)}}$$

$$= \frac{\|u\|^2\|v\|^2 - (u^Tv)^2}{\|v\|^2} + \frac{(x^T\Delta x^{(k)})^2}{(\Delta g^{(k)})^T(\Delta x^{(k)})} \quad (1)$$

where $u = (B^k)^{1/2}x$ & $v = (B^k)^{1/2}\Delta g^{(k)}$

Using the Cauchy Schwartz inequality . . .

$$(u^Tv)^2 \leq (\|u\|\|v\|)^2$$

Equality holds iff $u = \theta v$ for some $\theta \neq 0$

Case (a) $u = \theta v \Rightarrow [u^Tv]^2 = \|u\|^2\|v\|^2$

$\Rightarrow$ (1) becomes

$$x^TB^{(k+1)}x = \frac{(x^T\Delta x^{(k)})^2}{(\Delta x^{(k)})^T\Delta g^{(k)}}$$

Since $u = \theta v$, $(B^{(k)})^{1/2}x = \theta(B^k)^{1/2}\Delta g^{(k)}$

$\Rightarrow x = \theta\Delta g^{(k)}$ [$\det(B) = \det(B^{1/2}B^{1/2}) \neq 0$
$\Rightarrow \det(B^{1/2}) \neq 0 \Rightarrow B^{1/2}$ has independent rows/columns]

$$\therefore\ x^T B^{(k+1)} x = \theta^2 \frac{\left[\left(\Delta g^{(k)}\right)^T \Delta x^{(k)}\right]^2}{\left[\left(\Delta x^{(k)}\right)^T \Delta g^{(k)}\right]}$$

$$= \theta^2 \underbrace{\left(\Delta g^{(k)}\right)^T \Delta x^{(k)}}_{\text{given to be } > 0} > 0$$

Case (b) If $u \neq \theta v$

$$\underbrace{\frac{\|u\|^2 \|v\|^2 - \left(u^T v\right)^2}{\|v\|^2}}_{\text{first term in (1)}} > 0$$

Since $\left(\Delta g^{(k)}\right)^T \Delta x^{(k)} > 0$

$$\underbrace{\frac{\left(x^T \Delta x^{(k)}\right)^2}{\left(\Delta g^{(k)}\right)^T \Delta x^{(k)}}}_{\text{Second term in (1)}} > 0$$

Together imply

$$x^T B^{k+1} x > 0$$

$$\Downarrow$$

$B^{(k+1)}$ is positive definite

4. Consider the overdetermined system of nonlinear equations

$$x_1^2 - x_2^2 - x_1 - 3x_2 = 2$$

$$x_1^3 - x_2^4 = -2$$

$$x_1^2 + x_2^3 + 2x_1 - x_2 = -1.1$$

Suppose we decide to find a solution for the above equations by minimizig

$$F(\mathbf{x}) = \sum_{i=1}^{3} f_i^2(\mathbf{x})$$

where

$$f_1(\mathbf{x}) = x_1^2 - x_2^2 - x_1 - 3x_2 - 2$$

$$f_2(\mathbf{x}) = x_1^3 - x_2^4 + 2$$

$$f_3(\mathbf{x}) = x_1^2 + x_2^3 + 2x_1 - x_2 + 1.1$$

For such an objective as $F(\mathbf{x})$ (decomposable as a sum of squares of objectives), the Gauss-Newton algorithm is applicable. Write down the steps of the Gauss Newton algorithm with the specific expressions for the relevant Jacobians and Hessian matrices, *etc.* for the specific objectives discussed here. As hint, to begin with, you can first identify what l and m (refer class notes) are.

(5 Marks)

Soln: Recall (from discussion of Gauss Newton algo at http://www.cse.iitb.ac.in/~CS709/notes/eNotes/unconstrained-optimisation.pdf)

F in our case

If $f = l \circ m$ → vector valued fn such as

eg: $l = m^T m$ such as here

$$m(x) = \begin{bmatrix} f_1(x) \\ f_2(x) \\ f_3(x) \end{bmatrix}$$

(Reproducing steps ditto)

$$\nabla^2 f(\mathbf{x}) = \underbrace{J_{\mathbf{m}}(\mathbf{x})^T \nabla^2 l(\mathbf{m}) J_{\mathbf{m}}(\mathbf{x})}_{G_f(\mathbf{x})} + \sum_{i=1}^{p} \nabla^2 m_i(\mathbf{x})(\nabla l(\mathbf{m}))_i$$

→ From chain rule

where $J_{\mathbf{m}}$ is the jacobian[28] of the vector valued function $\mathbf{m}$. It can be shown that if $\nabla^2 l(\mathbf{m}) \succeq 0$, then $G_f(\mathbf{x}) \succeq 0$. The term $G_f(\mathbf{x})$ is called the Gauss-Newton approximation of the Hessian $\nabla^2 f(\mathbf{x})$. In many situtations, $G_f(\mathbf{x})$ is the dominant part of $\nabla^2 f(\mathbf{x})$ and the approximation is therefore reasonable.

Q: If x is point of minimum, $\nabla^2 f(x) = \boxed{G_f(x)}$

The (approximate) Newton update rule will be:

$$\Delta \mathbf{x} = -(G_f(\mathbf{x}))^{-1} \nabla f(\mathbf{x}) = -(G_f(\mathbf{x}))^{-1} J_{\mathbf{m}}^T(\mathbf{x}) \nabla l(\mathbf{m})$$

where we use the fact that $(\nabla f(\mathbf{x}))_i = \sum_{k=1}^{p} \frac{\partial l}{\partial m_k} \frac{\partial m_k}{\partial x_i}$, since the gradient of a composite function is a product of the jacobians.

Only additional specification reqd here:

$$J_m(x) = \begin{bmatrix} \nabla^T f_1(x) \\ \nabla^T f_2(x) \\ \nabla^T f_3(x) \end{bmatrix} = \begin{bmatrix} 2x_1 - 1 & -2x_2 - 3 \\ 3x_1^2 & -4x_2^3 \\ 2x_1 + 2 & 3x_2^2 - 2 \end{bmatrix}$$

& $\nabla l(m) = 2m(x)$ $\quad \nabla^2 l(m) = 2I$

$\Rightarrow G_f(x) = 2J_m^T(x) J_m(x) = \boxed{}$

work it out

& $J_m^T(x)\nabla \ell(m) = 2J_m^T(x)\,m(x) =$

work it out

given a starting point $x \in \textbf{dom}\, f$, tolerance $\epsilon > 0$.
repeat

1. *Compute the Newton step and decrement.*
   $\Delta x_{\mathrm{nt}} := -\nabla^2 f(x)^{-1}\nabla f(x)$; $\quad \lambda^2 := \nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x)$.
2. *Stopping criterion.* **quit** if $\lambda^2/2 \le \epsilon$.
3. *Line search.* Choose step size t by backtracking line search.
4. *Update.* $x := x + t\Delta x_{\mathrm{nt}}$.

$\rightarrow$ steps

Replace by $\Delta x_{GN} = -(G_f(x))^{-1} J_m^T(x)\nabla \ell(m)$

$= \ldots\ldots\ldots$

5. Consider the optimisation problem

$$\begin{array}{lll} \text{minimize} & f(\mathbf{x}) = & x_1 + x_2 \\ \text{subject to} & c_1(\mathbf{x}) = & x_1^2 + x_2^2 - 2 = 0 \end{array} \qquad (1)$$

What is the solution to this problem?

Now consider optimising this constrained objective by optimising the quadratic penalty function

$$Q(\mathbf{x}, \mu) = x_1 + x_2 + \frac{\mu}{2}(x_1^2 + x_2^2 - 2)^2$$

Suppose the problem (1) has a minimum at $\mathbf{x}^*$ with Lagrange multiplier λ^*. Show that $Q(\mathbf{x}, \mu)$ does not have a local minimum at $\mathbf{x}^*$ unless $\mu > ||\lambda^*||_\infty$.

Hint: Consider directional derivative of Q at $\mathbf{x}^*$.

Aside (no need to prove): This claim for μ holds for any choice of non-linear f, c_1 and with $Q(\mathbf{x}, \mu) = f(\mathbf{x}) + \frac{\mu}{2}c_1(\mathbf{x})$.

(6 Marks)

Ans: Soln is $x^* = (-1,-1)$ $\lambda^* = \frac{1}{2}$. [obtained by solving KKT equations]

ie $x_1 + x_2 + \lambda(x_1^2 + x_2^2 - 2) = L(x_1, x_2, \lambda)$

$$\nabla L(x^*) = 0 \Rightarrow \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \lambda^* \begin{bmatrix} 2x_1^* \\ 2x_2^* \end{bmatrix} = 0$$

$$\Rightarrow x_1^* = -1/2\lambda^* \quad x_2^* = -1/2\lambda^*$$

Since $x_1^{*2} + x_2^{*2} = 2$, $\lambda^{*2} = 1/4$

It is obvious that $x_1^* + x_2^*$ is minimised if $\lambda^* > 0$ ie $\lambda^* = 1/2 \Rightarrow (x_1^*, x_2^*) = (-1,-1)$

The second part of the question was a trivial teaser:

$$\nabla Q(x, \mu) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \mu \begin{bmatrix} 2x_1(x_1^2 + x_2^2 - 2) \\ 2x_2(x_1^2 + x_2^2 - 2) \end{bmatrix}$$

$$\nabla Q(x^*, \mu) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Directional derivative along any direction $[p_1, p_2]$ at x^* is given as

$$\nabla^T Q(x^*)p = p_1 + p_2$$

i.e no matter what the value of μ, $\exists$ direction $[p_1, p_2]$ at x^* along which directional derivative is <0 (choose p_1 & p_2 s.t $p_1 + p_2 < 0$)

$\therefore$ x^* can never be a point of local min for $Q(x,\mu)$. This proves the claim in any case, that is

$Q(x,\mu)$ does not have local min at x^* if $\mu > \|\lambda^*\|_\infty$

In fact $Q(x,\mu)$ does not have a local min at x^* for any value of μ

In fact the claim holds more precisely if

$$Q(x,\mu) = x_1 + x_2 + \frac{\mu}{2}|x_1^2 + x_2^2 - 2|$$

$$Q(\mathbf{x}, \mu) = x_1 + x_2 + \frac{\mu}{2}|x_1^2 + x_2^2 - 2|$$

Suppose the problem (1) has a minimum at $\mathbf{x}^*$ with Lagrange multiplier λ^*. It can be shown that that $Q(\mathbf{x}, \mu)$ does not have a local minimum at $\mathbf{x}^*$ unless $\mu > ||\lambda^*||_\infty$.

6. Solve the constrained minimisation problem

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) = && x_1^2 + x_2^2 - 14x_1 - 6x_2 \\ \text{subject to} \quad & c_1(\mathbf{x}) = && 2 - x_1 - x_2 \geq 0 \\ & c_2(\mathbf{x}) = && 3 - x_1 - 2x_2 \geq 0 \end{aligned} \tag{2}$$

by applying KKT conditions.

(6 Marks)

Ans: KKT conditions

$$2x_1 - 14 + \lambda_1 + \lambda_2 = 0$$
$$2x_2 - 6 + \lambda_1 + 2\lambda_2 = 0$$
$$\lambda_1(2 - x_1 - x_2) = 0$$
$$\lambda_2(3 - x_1 - 2x_2) = 0$$
$$\lambda_1 \geq 0$$
$$\lambda_2 \geq 0$$

Case I: No active constraints:

ie $\lambda_1^* = \lambda_2^* = 0 \implies \cancel{x^* = \begin{bmatrix} 7 \\ 3 \end{bmatrix}}$ which violates both constraints & $\therefore$ Not solution

Case 2: One active constraint:

If first constraint is active:

$$\lambda_2^* = 0 \quad \text{and}$$

$$\left.\begin{array}{l} 2x_1 - 14 + \lambda_1 = 0 \\ 2x_2 - 6 + \lambda_1 = 0 \\ 2 - x_1 - x_2 = 0 \end{array}\right\} \text{Solving, } x^* = \begin{bmatrix} 3 \\ -1 \end{bmatrix} \ \& \ \lambda_1^* = 8$$

Since x^* also satisfies the second constraint, a soln to KKT is

$$\boxed{x^* = \begin{bmatrix} 3 \\ -1 \end{bmatrix} \qquad \lambda^* = \begin{bmatrix} 8 \\ 0 \end{bmatrix}}$$

If only second constraint is active:

$$\lambda_1^* = 0 \quad \text{and}$$

$$\left.\begin{array}{l} 2x_1 - 14 + \lambda_2 = 0 \\ 2x_2 - 6 + 2\lambda_2 = 0 \\ 3 - x_1 - x_2 = 0 \end{array}\right\} \text{Solving: } x^* = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \ \& \ \lambda^* = \begin{bmatrix} 20 \\ -8 \end{bmatrix}$$

Since $\lambda_2^* < 0$, this is not a soln to the optimisation problem

7. Let the feasible region $\mathcal{D}$ be given as

$$\mathcal{D}: \quad \begin{array}{ll} g_i(\mathbf{x}) \leq \mathbf{0} & \text{for } i = 1 \ldots m \\ h_j(\mathbf{x}) = \mathbf{0} & \text{for } j = 1 \ldots k \end{array}$$

At some feasible point $\mathbf{x}$, let $\mathcal{I}(\mathbf{x})$ be the active index set for the inequality constraints at $\mathbf{x}$, and define the sets $\mathcal{F}(\mathbf{x})$ and $F(\mathbf{x})$ as

$$\mathcal{F}(\mathbf{x}) = \left\{ \mathbf{s} : \begin{array}{ll} g_i(\mathbf{x}+\mathbf{s}) \leq \mathbf{0} & \text{for } i \in \mathcal{I}(\mathbf{x}) \\ h_j(\mathbf{x}+\mathbf{s}) = \mathbf{0} & \text{for } j = 1 \ldots \cancel{k}m \end{array} \right\}$$

and

$$F(\mathbf{x}) = \left\{ \mathbf{s} : \begin{array}{ll} \mathbf{s}^T \nabla g_i(\mathbf{x}) \leq \mathbf{0} & \text{for } i \in \mathcal{I}(\mathbf{x}) \\ \mathbf{s}^T \nabla h_j(\mathbf{x}) = \mathbf{0} & \text{for } j = 1 \ldots \cancel{k}m \end{array} \right\}$$

(a) We claim that $\mathcal{F}(\mathbf{x}) \subseteq F(\mathbf{x})$. Provide a sketch of the proof, stating what key properties you would use.

(3 Marks)

(b) Show that if the constraints that are active at $\mathbf{x}$ are all linear, then $F(\mathbf{x}) = \mathcal{F}(\mathbf{x})$.

(3 Marks)

(c) The condition that $F(\mathbf{x}) = \mathcal{F}(\mathbf{x})$ is called the *constraint qualificiation* of $\mathbf{x}$. Suppose the only constraints are given by

- $g_1(x_1, x_2) = x_2 - x_1^3$
- $g_2(x_1, x_2) = -x_2$.

Then, does the constraint qualification assumption hold at $\mathbf{x} = \mathbf{0}$? Prove your statement.

(3 Marks)

Ans: For (a), only a sketch is reqd. So if highlighted points are mentioned, it is ok

(a) Let $\{x^{(k)}, k = 1, 2, \ldots\}$ be a sequence of feasible points s.t $x^k \to x$ as $k \to \infty$.

Let $x^k - x = \alpha_k s^k \longrightarrow (2)$

where $\alpha_k > 0$ is a scalar and s^k is a vector s.t $\|s^k\| = 1$.

Then if $x^k \to x$, $\alpha_k \to 0$

Vector s is said to be feasible at point x if there is a sequence of feasible pts $\{x^k\}$ as described above st $s^k \to s$

Let $s \in \mathcal{F}(x)$. We will try and show that $s \in F(x)$

Let $\{x^k\}$ be a sequence s.t x^k in (2) are feasible pts with $s^k \to s$

Let us write the Taylor expansions of

$h_j(x)$ & $g_i(x)$ at $x^k = x + \alpha_k s^k$

$$h_j(x^k) = h_j(x) + \alpha_k \nabla^T h_j(x) s^k + o(\alpha^k)$$

for $j = 1 \ldots m$

$$g_i(x^k) = g_i(x) + \alpha_k \nabla^T g_i(x) s^k + o(\alpha^k)$$

for $i = I(x)$

Since $h_j(x) = 0$, $h_j(x^k) = 0$, $g_i(x) = 0$ & $g_i(x^k) \geq 0$

we get

$$\nabla^T h_j(x) s^k + o(1) = 0 \text{ for } j = 1 \ldots m$$

$$\nabla^T g_i(x) s^k + O(1) = 0 \text{ for } i \in I(x)$$

Letting $k \to \infty$, we obtain

$$\nabla^T h_j(x) s = 0 \ \& \ \nabla^T g_i(x) s \geq 0$$

which implies that $s \in F(x)$

$$\therefore \mathcal{F}(x) \subseteq F(x)$$

(b) We know that $\mathcal{F}(x) \subseteq F(x)$

Let $s \in F(x)$ i.e

$$\nabla^T h_j(x) s = 0 \quad \text{for } j = 1 \cdots m$$

$$\nabla^T g_i(x) s = 0 \quad \text{for } i \in \mathcal{I}(x)$$

Since $h_j(x)$ & $g_i(x)$ are linear and since x is feasible, we have

$$h_j(x+s) = h_j(x) + \nabla^T h_j(x) s = 0 \quad j = 1 \cdots m$$

$$g_i(x+s) = g_i(x) + \nabla^T g_i(x) s = \nabla^T g_i(x) s \geq 0 \quad i \in \mathcal{I}(x)$$

That is, s is feasible & $\therefore s \in \mathcal{F}(x)$

$$\therefore F(x) \subseteq \mathcal{F}(x)$$

We already know that in general

$$\mathcal{F}(x) \subseteq F(x)$$

$$\Rightarrow \boxed{F(x) = \mathcal{F}(x)}$$

(c) The feasible region is given by

$$\mathcal{F}(x) = \left\{ x \;\middle|\; \begin{array}{l} g_1(x) = -x_1^3 + x_2 \le 0 \\ g_2(x) = -x_2 \le 0 \end{array} \right\}$$

At $x=0$, both constraints are active.
At $x=0$ their gradients are

$$\nabla g_1(0) = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \;\&\; \nabla g_2(0) = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$$

Consider $s = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$. You can easily see that

$$s \notin \tilde{F}(x)$$

However

$$\nabla^T g_1(x) s = 0 \quad \& \quad \nabla^T g_2(x) s = 0$$

$$\Rightarrow s \in F(x)$$

$$\therefore \tilde{F}(x) \not\supseteq F(x)$$

i.e $\tilde{F}(x) \neq F(x)$

i.e constraint qualification assumption does not hold at $x = 0$

8. Consider the problem

$$\text{minimize} \quad \tfrac{1}{2}\mathbf{x}^T A\mathbf{x} - \mathbf{b}^T\mathbf{x} \tag{3}$$

where A is a symmetric positive definite matrix. Let $\{\mathbf{d}^0, \mathbf{d}^1, \ldots, \mathbf{d}^{(n-1)}\}$ be a set of nonzero vectors that are mutually conjugate with respect to A.

The algorithm is iterative (like the conjugate gradient method outlined in notes). The k^{th} iteration consists of the following step:

- $\mathbf{x}^{(k+1)} = \mathbf{x}^k + \alpha_k\mathbf{d}^k$ where α_k is the one dimensional minimizer of $\phi(\alpha) = f(\mathbf{x}^k + \alpha\mathbf{d}^k)$ and is given as $\alpha_k = -\frac{\nabla^T f(\mathbf{x}^k)\mathbf{d}^k}{(\mathbf{d}^k)^T A\mathbf{d}^k}$.

Let $\mathbf{x}^0 \in \Re^n$ be the initial point. We will prove that the sequence $\{\mathbf{x}^k\}$ generated by the repeated application of the conjugate gradient step above, for increasing values of k, converges to the solution $\mathbf{x}^*$ of the problem (3) in at most n steps (for step (b) onwards, provide brief justification):

(a) Prove that the directions $\{\mathbf{d}^0, \mathbf{d}^1, \ldots, \mathbf{d}^{(n-1)}\}$ are linearly independent.

(2 Marks)

Ans: Suppose $d^0, d^1 \ldots d^{(n-1)}$ were not linearly independent. Then there would exist

$\alpha_1, \alpha_2 \ldots \alpha^{(n-1)}$

$$d^0 = \sum_{i=1}^{n-1} \alpha_i d_i$$

Premultiplying both sides by $(d^j)^T A$ for $j = 1 \ldots n-1$,

$$(d^j)^T A d^0 = \sum_{i=1}^{n-1} \alpha_i (d^j)^T A d^i \quad \text{for } j = 1 \dots n-1$$

$$\Longrightarrow \quad \alpha_i (d^i)^T A d^i = 0 \quad \text{for } i = 1 \dots n-1$$

(since $(d^j)^T A d^i = 0$ $\forall i \neq j$)

$$\Longrightarrow \quad \alpha_i = 0 \quad \text{for } i = 1 \dots n-1$$

(since $(d^i)^T A d^i \neq 0$)

Which contradicts our assumption.

$\therefore$ $d^0, d^1 \dots d^{n-1}$ are linearly independent

(b) Since the directions $\{\mathbf{d}^0, \mathbf{d}^1, \dots, \mathbf{d}^{(n-1)}\}$ are linearly independent, we can write the following for some choice of scalars $\gamma_0, \gamma_1, \dots, \gamma_{n-1}$.

$$\mathbf{x}^* - \mathbf{x}^0 = \sum_{i=0}^{n-1} \gamma_i d^i$$

(1 Marks)

(c) By premultiplying both sides of this inequality by $\left(\mathbf{d}^k\right)^T A$ and using properties determined so far, we obtain the following expression for γ_k:

$$\gamma_k = \ldots \frac{(d^k)^T A (x^* - x^0)}{(d^k)^T A (d^k)} \ldots$$

(1 Marks)

(d) $\mathbf{x}^k$ can be expressed in terms of $\mathbf{x}^0, \mathbf{d}^0, \mathbf{d}^1, \ldots, \mathbf{d}^{(k-1)}$, *etc.* as:

$$\mathbf{x}^k = \ldots x^0 + \sum_{i=0}^{k-1} \alpha_i d^i \ldots$$

(1 Marks)

(e) By premultiplying the expression by $\left(\mathbf{d}^k\right)^T A$ and using the properties determined so far, we have:

$$\left(\mathbf{d}^k\right)^T A(\mathbf{x}^k - \mathbf{x}^0) = \ldots 0 \ldots$$

(1 Mark)

(f) And therefore

$$\left(\mathbf{d}^k\right)^T A(\mathbf{x}^* - \mathbf{x}^0) = \ldots (d^k)^T A (x^* - x^k) \ldots$$

$$= (d^k)^T (b - Ax^k)$$

$$= \gamma_k (d^k)^T A (d^k)$$

$$\Rightarrow \gamma_k = \frac{(d^k)^T (b - Ax^k)}{(d^k)^T A d^k} = \alpha_k.$$

(1 Mark)

(g) Thus $\gamma_k = \ldots \alpha_k \ldots$, which establishes the result.
(1 Mark)

Note $\nabla f(x^k) = Ax^k - b \Rightarrow \alpha_k = \dfrac{(d^k)^T (b - Ax^k)}{(d^k)^T A d^k}.$