Review

- $\langle u+x, v+y\rangle = \langle u,v\rangle + \langle x,v\rangle + \langle u,y\rangle + \langle x,y\rangle$

  Caution! $\langle u,v\rangle + \langle x,y\rangle \neq \langle u+x, v+y\rangle$ in general.

- Let A be an $m\times n$ matrix. $x \in \mathbb{R}^n$ $y \in \mathbb{R}^m$. $\langle Ax, y\rangle = y^T Ax = \sum_{r=1}^{m}\sum_{c=1}^{n} A_{rc} x_c y_r$.

- Let $f,g : \Omega \to \mathbb{R}^m$. $\Omega \subset \mathbb{R}^n$. $h(x) = \langle f(x), g(x)\rangle : \Omega \to \mathbb{R}$. Then

  $dh(x) = (f(x))^T dg(x) + (g(x))^T df(x)$.

  Proof. $h(x) = \langle f(x), g(x)\rangle = (f(x))^T g(x) = \sum_{i=1}^{m} f_i(x) g_i(x)$.

  $dh(x) = \sum_{i=1}^{m} f_i(x) dg_i(x) + g_i(x) df_i(x) = (f(x))^T dg(x) + (g(x))^T df(x)$.

  - A $n\times n$. $f(x) = \langle Ax, x\rangle$. Then $df(x) = (Ax)^T I + x^T A = x^T A^T + x^T A$.

    - If A is symmetric, then $df(x) = 2x^T A$. So, $\frac{\partial}{\partial x_j}\langle Ax, x\rangle = 2(Ax)_j$.

Second Derivatives for $f : D \to \mathbb{R}$

- Def: Let $f : D \to \mathbb{R}^m$ with $D \subseteq \mathbb{R}^n$ be differentiable. If $df : D \to \mathbb{R}^{m\times n}$ is differentiable at y, the its differential $d(df)(y)$ (if it exists) is called the **second derivative**, denoted $\underset{(m\times n)\times n}{d^2 f(y)}$.

  - If $m = 1$, then $\underset{n\times n}{d^2 f(y)}$ is an $n\times n$ matrix called the **Hessian** of f at y.

- Def: If f is differentiable and $\frac{\partial f}{\partial x_j}$ has a partial derivative $\frac{\partial}{\partial x_k}\left(\frac{\partial f}{\partial x_j}\right)$ at a point y, we say the **second order partial derivative** $\frac{\partial^2 f}{\partial x_k \partial x_j}$ exists at y and equals $\frac{\partial}{\partial x_k}\left(\frac{\partial f}{\partial x_j}\right)$.

$$\frac{\partial^2 f}{\partial x_j \partial x_k}(y) = \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_k}\right)(y) = \lim_{t\to 0}\frac{\frac{\partial f}{\partial x_k}(y+te_j) - \frac{\partial f}{\partial x_k}(y)}{t}$$

$$= \lim_{t\to 0}\lim_{s\to 0}\frac{1}{ts}\left(f(y+te_j+se_k) - f(y+se_k) - f(y+te_j) + f(y)\right)$$

$$= \lim_{t\to 0}\lim_{s\to 0}\frac{1}{ts}\Delta_{se_k}\Delta_{te_j} f(x) = \lim_{t\to 0}\lim_{s\to 0}\frac{1}{ts}\Delta_{te_j}\Delta_{se_k} f(x)$$

- Can be defined independent of the existence of $d^2 f$, as long as $\frac{\partial f}{\partial x_k}$ exists in a neighborhood of *y*.

- If $f : D \to \mathbb{R}$ with $D \subseteq \mathbb{R}^n$ has a second derivative at a point *y*, then all second-order partial derivatives exist at *y* and the Hessian $\underset{n\times n}{d^2 f}(y)$ matrix is $\left(d^2 f\right)_{jk} = \frac{\partial^2 f}{\partial x_k \partial x_j}$.

$df = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$. (In fact, it is $n\times 1$ (a row vector), but here, we treat it as a (column) vector in $\mathbb{R}^n$.

$$d\left(df\right) = \left(\frac{\partial}{\partial x_1} df \cdots \frac{\partial}{\partial x_n} df \right) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1} \\ \vdots & \frac{\partial^2 f}{\partial x_c \partial x_r} & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_n} \end{pmatrix}$$

- Let $f : D \to \mathbb{R}$ be continuous together with all partial derivatives of order one and two. Then $\left(d^2 f\right)_{jk} = \frac{\partial^2 f}{\partial x_j \partial x_k} = \frac{\partial^2 f}{\partial x_k \partial x_j} = \left(d^2 f\right)_{kj}$ $\forall j \forall k$; hence, the Hessian matrix is symmetric.

  - Mean value theorem: Regard $g(x)$ as $h(x_k)$, a function of x_k alone, with the other variables held fixed. If $h(x_k)$ is continuous on [0,*t*] and differentiable at every point in the interior. Then, $\exists t_1 \;\; 0 < t_1 < t$, $\frac{1}{t}\Delta_{te_k} g(x) = \frac{\partial g}{\partial x_k}(x + t_1 e_k)$.

  - $\forall t \forall s \exists t_1 \exists s_1 \;\; 0 < t_1 < t$, $0 < s_1 < s$, $\frac{1}{st}\Delta_{te_k}\Delta_{se_j} f(x) = \frac{\partial^2 f}{\partial x_j \partial x_k}\left(x + t_1 e_k + s_1 e_j\right)$

    Proof. Let $h(x) = \Delta_{se_j} f(x) = f\left(x + se_j\right) - f(x)$ differentiable, and $H(t) = h\left(x + te_k\right)$ differentiable. Then,
    $\frac{1}{st}\Delta_{te_k}\Delta_{se_j} f(x) = \frac{1}{st}\Delta_{te_k} h(x) = \frac{1}{st}\left(h\left(x + te_k\right) - h(x)\right) = \frac{1}{st}\left(H(t) - H(0)\right)$.
    By the mean value theorem, $\exists t_1 \;\; 0 < t_1 < t \;\; H(t) - H(0) = tH'(t_1)$. But
    $H'(t_1) = \lim_{l\to 0}\frac{H(t_1 + l) - H(t_1)}{l} = \lim_{l\to 0}\frac{h\left(x + t_1 e_k + le_k\right) - h\left(x + t_1 e_k\right)}{l} = \frac{\partial h}{\partial x_k}\left(x + t_1 e_k\right)$.

Thus, $\frac{1}{st}\Delta_{te_k}\Delta_{se_j} f(x) = \frac{1}{s}\frac{\partial h}{\partial x_k}(x + t_1 e_k)$. Note that by the mean value theorem, $\exists s_1$ $0 < s_1 < s$ $\frac{\partial h}{\partial x_k}(x) = \frac{\partial f}{\partial x_k}(x + se_j) - \frac{\partial f}{\partial x_k}(x) = \frac{\partial}{\partial x_j}\frac{\partial f}{\partial x_k}(x + s_1 e_j)$. Therefore, $\frac{1}{st}\Delta_{te_k}\Delta_{se_j} f(x) = \frac{1}{s}\frac{\partial h}{\partial x_k}(x + t_1 e_k) = \frac{\partial^2 f}{\partial x_j \partial x_k}(x + t_1 e_k + s_1 e_j)$.

- $\forall t \forall s \exists t_1 \exists t_2 \exists s_1 \exists s_2$ $0 < t_1, t_2 < t$ $0 < s_1, s_2 < s$ $\frac{\partial^2 f}{\partial x_j \partial x_k}(x + t_1 e_k + s_1 e_j) = \frac{1}{st}\Delta_{te_k}\Delta_{se_j} f(x) = \frac{1}{ts}\Delta_{se_j}\Delta_{te_k} f(x) = \frac{\partial^2 f}{\partial x_k \partial x_j}(x + t_2 e_k + s_2 e_j)$.

  Take limit as $s, t \to 0$, by the continuity of $\frac{\partial^2 f}{\partial x_j \partial x_k}$ and $\frac{\partial^2 f}{\partial x_k \partial x_j}$, we have their equality.

- Still works if only one of the mixed second partial derivatives is continuous.

- The **difference operator** Δ_u: $\Delta_u f(x) = f(x+u) - f(x)$.
  - $\Delta_v \Delta_u f(x) = \Delta_u \Delta_v f(x) = f(x + u + v) - f(x+v) - f(x+u) + f(x)$.

Higher Derivatives

- Def: A function is said to be of class C^k if all partial derivatives of orders up to k exist and are continuous.

  $\Rightarrow$ full derivatives upto order k exist and are continuous.

- Def: f is C^∞ if f is in C^k for all finite k.

- Notation: Assume f is $C^{|a|}$. $a = (a_1, a_2, \ldots, a_n)$ is called a **multi-index**, each a_j being a non-negative integer; $\left(\frac{\partial}{\partial x}\right)^a f = \left(\frac{\partial}{\partial x_1}\right)^{a_1}\left(\frac{\partial}{\partial x_2}\right)^{a_2}\cdots\left(\frac{\partial}{\partial x_n}\right)^{a_n} f$; and $|a| = a_1 + a_2 + \ldots + a_n$ is the order of the partial derivatives.

- Assume f is $C^{|a|+|b|}$. $\left(\frac{\partial}{\partial x}\right)^a\left(\frac{\partial}{\partial x}\right)^b f = \left(\frac{\partial}{\partial x}\right)^{a+b} f$.

Quadratic form $Q_A(u) = \langle Au, u\rangle$

- Def: A **quadratic form** on $\mathbb{R}^n$ is a function of the form $\langle Au, u\rangle$ where A is a symmetric $n \times n$ matrix. Denote this by $Q_A(u)$.

  It is said to be

  non-negative definite if $\langle Au, u\rangle \geq 0$ $\forall u \in \mathbb{R}^n$

  positive definite if $\langle Au, u\rangle > 0$ $\forall u \in \mathbb{R}^n \setminus \{0\}$

  non-positive definite if $\langle Au, u\rangle \leq 0$ $\forall u \in \mathbb{R}^n$

negative definite if $\langle Au,u\rangle < 0 \ \ \forall u \in \mathbb{R}^n \setminus \{0\}$

- $\langle Au,u\rangle = \sum_{k=1}^{n}\sum_{i=1}^{n} A_{k,i}u_iu_k = \langle u, Au\rangle$. (Not require symmetry of A)
- $\langle Au,u\rangle = \langle u, Au\rangle$
  - $= \sum_{k=1}^{n}\sum_{i=1}^{n} A_{k,i}u_iu_k$
  - is continuous on $\mathbb{R}^n$.

    Proof. $\langle Ax,x\rangle = \sum_{k=1}^{n}\sum_{i=1}^{n} A_{k,i}x_ix_k$.
  - attains its minimum and maximum on compact set $|u| = 1$. $\Rightarrow$ critical point.
  - $dQ_A(x) = 2x^T A = 2(Ax)^T$.
  - $\frac{\partial}{\partial x_j}\langle Ax,x\rangle = \sum_{i=1}^{n} A_{ji}x_i + \sum_{k=1}^{n} A_{kj}x_k \underset{\text{Symmetric } A}{=} 2\sum_{k=1}^{n} A_{jk}x_k = 2(Ax)_j$.
- Def: An eigenvector u for a matrix A with eigenvalue l is a <u>non-zero</u> solution of $Au = lu$.
  - So, eigenvectors are nonzero.

- A quadratic form $\langle Au,u\rangle$ is **positive definite**

Equivalent statements (iff)

1) Def: $\langle Au,u\rangle > 0 \ \ \forall u \in \mathbb{R}^n \setminus \{0\}$.
2) $\langle Au,u\rangle > 0 \ \ \forall u$ in the unit sphere $|u| = 1$.
3) $\exists e > 0$ such that $\langle Au,u\rangle \ge e|u|^2 \ \ \forall u \in \mathbb{R}^n$.
4) All eigenvalues of A are positive.

Properties (implication)

a) Main diagonal entries are positive.

b) $\exists c > 0$ such that if B is symmetric and $|A - B| < c$, then Q_B is positive definite.

Proof 2). "$\Rightarrow$" all u in the unit sphere is also in $\mathbb{R}^n \setminus \{0\}$. "$\Leftarrow$" For any nonzero v, have $\langle Av,v\rangle = |v|^2 \left\langle A\frac{v}{|v|}, \frac{v}{|v|}\right\rangle > 0$.

Proof 3) "$\Leftarrow$" Because $e > 0$, and $|u|^2 \ge 0$ with equality iff $u = 0$, $\langle Au,u\rangle \ge e|u|^2 \ge 0$ with equality iff $u = 0$. "$\Rightarrow$" By continuity of $\langle Au,u\rangle$ on compact set $|u| = 1$, $\langle Au,u\rangle$ attains minimum e ; > 0 because anywhere on the sphere, $\langle Au,u\rangle > 0$. For general nonzero v,

$\left\langle A\frac{v}{|v|},\frac{v}{|v|}\right\rangle \geq e > 0$.

Proof a) Let $x = e^{(j)} \neq 0$. Then $\langle Ax,x\rangle = A_{jj} > 0$. Converse is not ture.

- Positive definite matrices form an open set in the space of symmetric matrices.
  - If $\langle Au,u\rangle$ is positive definite, then
    - so is $\langle Bu,u\rangle$ for all symmetric matrices B sufficiently close to A.
    - $\exists c > 0$ such that if B is symmetric and $|A - B| < c$, then Q_B is positive definite.

      Proof b). $-\langle (B-A)u,u\rangle \leq |\langle (B-A)u,u\rangle| \leq |(B-A)u||u| \leq |B-A||u|^2$. Therefore, $\langle Bu,u\rangle = \langle Au,u\rangle + \langle (B-A)u,u\rangle \geq e|u|^2 - |B-A||u|^2 = (e - |B-A|)|u|^2$. Choose $|B-A| = \frac{e}{2}$.
  - For a C^2 function f, if $d^2 f(y)$ is positive definite, then $\exists d$ such that $\forall x$ $|x-y| < d \Rightarrow$ $d^2 f(x)$ is positive definite.

    Proof. $\exists c > 0$ such that if $|d^2 f(x) - d^2 f(y)| < c$, then $d^2 f(x)$ is positive definite. By continuity of $d^2 f$ at y, $\forall c$ $\exists d$ such that $|x-y| < d \Rightarrow |d^2 f(x) - d^2 f(y)| < c$.
- non-negative definite if $\langle Au,u\rangle \geq 0$ $\forall u \in \mathbb{R}^n$ $\equiv$ all eigenvalues are non-negative.

  positive definite if $\langle Au,u\rangle > 0$ $\forall u \in \mathbb{R}^n \setminus \{0\}$ $\equiv$ all eigenvalues are positive.

  non-positive definite if $\langle Au,u\rangle \leq 0$ $\forall u \in \mathbb{R}^n$ $\equiv$ all eigenvalues are non-positive.

  negative definite if $\langle Au,u\rangle < 0$ $\forall u \in \mathbb{R}^n \setminus \{0\}$ $\equiv$ all eigenvalues are negative.

  Proof. By spectral theorem, $x = \sum_{j=1}^{n} \langle x,u^{(j)}\rangle u^{(j)}$, $\langle u^{(j)},u^{(k)}\rangle = d(j,k)$.

  $Ax = \sum_{j=1}^{n} \langle x,u^{(j)}\rangle Au^{(j)} = \sum_{j=1}^{n} \langle x,u^{(j)}\rangle l_j u^{(j)}$. Thus, $\langle Ax,x\rangle$

  $= \left\langle \sum_{j=1}^{n} \langle x,u^{(j)}\rangle l_j u^{(j)}, \sum_{k=1}^{n} \langle x,u^{(k)}\rangle u^{(k)} \right\rangle = \sum_{k=1}^{n}\sum_{j=1}^{n} \langle x,u^{(j)}\rangle \langle x,u^{(k)}\rangle l_j \langle u^{(j)},u^{(k)}\rangle$

  $= \sum_{j=1}^{n} l_j \langle x,u^{(j)}\rangle^2$. $\sum_{j=1}^{n} l_j t_j^2 > 0 \,\forall \vec{t} \neq 0 \Leftrightarrow l_j > 0 \,\forall j$.

Maximum and minimum of $f : D \to \mathbb{R}$ with $D \subset \mathbb{R}^n$

- Let $f : D \to \mathbb{R}$ for $D \subseteq \mathbb{R}^n$, and let y be a point in the interior of D.

If f assumes its maximum or minimum value at y and f is differentiable at y, then $df(y) = \nabla f(y) = 0$ $\left(\frac{\partial f}{\partial x_k}(y) = 0 \forall\ k = 1, \ldots, n\right)$.

Proof. Let $g(t) = f(y + te_j)$. By chain rule, $g'(0)$ exists and $= df(y + 0e_j)e_j$ $= \frac{\partial f}{\partial x_j}(y)$. Because $g(t)$ attains its max or min at $t = 0$, $g'(0) = 0$.

- Let g be C^2. If $g'(t_0) = 0$ and $g''(t) > 0\ \forall t\ 0 < |t - t_0| < e$, then $\forall t\ 0 < |t - t_0| < e$, $g(t) > g(t_0)$.

  Proof. Consider t, $t_0 < t < e + t_0$. Then, from MVT, $\exists t_1 \exists t_2\ t_0 < t_2 < t_1 < t < e + t_0$,

  $$g(t) - g(t_0) = g'(t_1)(t - t_0) = \left(g'(t_1) - \underbrace{g'(t_0)}_{0}\right)(t - t_0) = \underbrace{g''(t_2)}_{>0}\underbrace{(t_1 - t_0)}_{>0}\underbrace{(t - t_0)}_{>0} > 0.$$

- $f : D \to \mathbb{R}$ be C^2 with open $D \subseteq \mathbb{R}^n$. Let $g(t) = f(y + tu)$, then

  $$dg(t) = \underbrace{df(y + tu)}_{1\times n}u = u^T(df(y + tu))^T$$

  $$d^2g(t) = u^T d\left(df(y + tu)\right)^T = u^T d^2 f(y + tu)u = \langle d^2 f(y + tu)u, u\rangle.$$

  $$\boxed{g(t) = f(y + tu) \Rightarrow d^2g(t) = \langle d^2 f(y + tu)u, u\rangle}.$$

- Def: Let $f : D \to \mathbb{R}^m$ with open $D \subseteq \mathbb{R}^n$. $y \in D$, and $df(y)$ exists. y is called a **critical point** (of f on D) if $df(y) = 0_{m\times n}$.

• Let $f : D \to \mathbb{R}$ be C^2 with open $D \subseteq \mathbb{R}^n$. Let $y \in D$ be a critical point $\left(\frac{\partial f}{\partial x_j}(y) = 0, \forall j\right)$.
1) If y is a **local minimum** $\cup$ $(\exists d > 0\ \forall x \in B_d(y)\ f(x) \ge f(y))$, then $d^2 f(y)$ is **non-negative definite** $(\langle d^2 f(y)u, u\rangle \ge 0 \forall u)$.
2) If y is a local maximum $\cap$, then $d^2 f(y)$ is non-positive definite.
3) If $d^2 f(y)$ is **positive definite**, then y is a **strict local minimum**.
4) If $d^2 f(y)$ is negative definite, then y is a strict local maximum.

- Let $g(t) = f(y + tu)$, then $g'(0) = 0$.

Proof 1). $\forall u \neq 0$, consider $g(t) = f(y+tu)$, $|t| < \frac{d}{|u|}$. Because y has local min at y, g has local min at 0, and thus, $g''(0) \geq 0$. Because $g''(0) = \langle d^2 f(y)u, u\rangle$, $d^2 f(y)$ is non-negative definite.

Proof 3). First, note that $\exists d$ such that $\forall x \; |x-y| < d \Rightarrow d^2 f(x)$ is positive definite. Consider any $x \neq y$ in this ball. Then, $\exists u \; |u| = 1 \; \exists t_0 \in (0,d)$ such that $x = y + t_0 u$. Let $g(t) = f(y+tu)$. So, we have $\forall t \; |t| < d \quad g''(t) = \langle d^2 f(y+tu)u, u\rangle > 0$. Thus, $f(x) = g(t_0) > g(0) = f(y)$.

Note: If limit to $g(t) = f(y+tu)$ for any nonzero direction u. Then, have $g'(0) = 0$ and $g''(0) > 0$, which implies $\exists e_u > 0 \;\; 0 < |t| < e_u \Rightarrow g(t) > g(0)$. But $\left(\bigcup_u \{y + tu : 0 < |t| < e_u\}\right)$ does not necessarily constitute a neighborhood of y. Need $\inf_u \{e_u\} > 0$.

- Eigenvalue of matrix A are the roots of the **characteristics polynomial** $p(l) = \det(lI - A)$.
  - $p(l) = \prod_{j=1}^{n}(l - l_j) = \prod_{k=1}^{n} a_k l^k$ with $a_n = 1$.
  - If $l_j > 0 \;\; \forall j$ then the signs of a_k alternate.
  - If $l_j < 0 \;\; \forall j$ then $a_k > 0 \; \forall k$.
- A is **symmetric**

$\equiv$ $A = A^T$

$\equiv$ $\forall \vec{x} \forall \vec{y} \;\; \langle A\vec{x}, \vec{y}\rangle = \langle \vec{x}, A\vec{y}\rangle$

Proof. "$\Rightarrow$" $\langle A\vec{x}, \vec{y}\rangle = \vec{x}^T A^T \vec{y}$, and $\langle \vec{x}, A\vec{y}\rangle = \vec{x}^T A\vec{y}$. "$\Leftarrow$" Let $\vec{x} = e_j$, and $\vec{y} = e_k$, then $\langle \vec{x}, A\vec{y}\rangle = A_{jk}$, and $\langle A\vec{x}, \vec{y}\rangle = (A^T)_{jk} = A_{kj}$.

- $\langle Ax, y\rangle = \sum_{i=1}^{n} \langle y, u^{(i)}\rangle \langle x, u^{(i)}\rangle l_i$

$\Rightarrow$ (**Spectral Theorem**) there exists a complete set of eigenvectors: $\exists$ an orthonormal basis $u^{(1)}, \ldots, u^{(n)}$ of $\mathbb{R}^n$ with $Au^{(k)} = l_k u^{(k)}$.

- $\langle u^{(j)}, u^{(i)}\rangle = d(i,j) = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$
- $\forall x \in \mathbb{R}^n \quad x = \sum_{i=1}^{n} \langle x, u^{(i)}\rangle u^{(i)}$

$\equiv$ A is diagonalizable by an orthogonal matrix.

$\Rightarrow$ $\left\langle Ax,y\right\rangle = \sum_{i=1}^{n}\left\langle y,u^{(i)}\right\rangle\left\langle x,u^{(i)}\right\rangle l_i$.

Proof

$$\left\langle Ax,y\right\rangle = \left\langle A\sum_{i=1}^{n}\left\langle x,u^{(i)}\right\rangle u^{(i)}, \sum_{j=1}^{n}\left\langle y,u^{(j)}\right\rangle u^{(j)}\right\rangle = \sum_{i=1}^{n}\sum_{j=1}^{n}\left\langle y,u^{(j)}\right\rangle\left\langle x,u^{(i)}\right\rangle\left\langle Au^{(i)},u^{(j)}\right\rangle$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n}\left\langle y,u^{(j)}\right\rangle\left\langle x,u^{(i)}\right\rangle\left\langle l_i u^{(i)},u^{(j)}\right\rangle = \sum_{i=1}^{n}\sum_{j=1}^{n}\left\langle y,u^{(j)}\right\rangle\left\langle x,u^{(i)}\right\rangle l_i\left\langle u^{(i)},u^{(j)}\right\rangle$$

$$= \sum_{i=1}^{n}\left\langle y,u^{(i)}\right\rangle\left\langle x,u^{(i)}\right\rangle l_i,$$

$\Rightarrow$ If u is an eigenvector of A, then $\forall x\ \left\langle x,u\right\rangle = 0 \Rightarrow \left\langle Ax,u\right\rangle = 0$. $\left(x\in u^{\perp} \Rightarrow Ax\in u^{\perp}\right)$.

Proof. $\left\langle Ax,u\right\rangle = x^T A^T u = x^T Au = x^T\left(lu\right) = l\,x^T u = 0$.

$\Rightarrow$ A is Nondegenerate $\left(\forall x\neq 0\ \exists y\neq 0 \text{ such that } \left\langle Ax,y\right\rangle\neq 0\right)$ iff all the eigenvalues of A are non-zero

Proof. "$\Rightarrow$" For each j, let $x = u^{(j)}$. $\exists y\in\mathbb{R}^n\setminus\{0\}$ such that $\left\langle Ax,y\right\rangle = \sum_{i=1}^{n}\left\langle y,u^{(i)}\right\rangle\left\langle x,u^{(i)}\right\rangle l_i \neq 0.$ Because $\left\langle x,u^{(i)}\right\rangle = \left\langle u^{(j)},u^{(i)}\right\rangle = d\left(i,j\right)$, we then have $\left\langle Ax,y\right\rangle = l_j\left\langle y,u^{(j)}\right\rangle\neq 0$, which implies $l_j\neq 0$. "$\Leftarrow$" $\forall x\neq 0\ \ \exists j\in\{1,\ldots,n\}$ such that $\left\langle x,u^{(j)}\right\rangle\neq 0$. (Otherwise $\forall j\ \left\langle x,u^{(j)}\right\rangle = 0$ would imply $x = \sum_{j=1}^{n}\left\langle x,u^{(j)}\right\rangle u^{(j)} = 0$.)

Let $y = u^{(j)}$. Then $\left\langle Ax,y\right\rangle = \sum_{i=1}^{n}\left\langle u^{(j)},u^{(i)}\right\rangle\left\langle x,u^{(i)}\right\rangle l_i = \underbrace{\left\langle x,u^{(j)}\right\rangle}_{\neq 0}\underbrace{l_j}_{\neq 0}\neq 0$.

- **Rayleigh quotient**: $R\left(x\right) = \dfrac{\left\langle Ax,x\right\rangle}{\left\langle x,x\right\rangle} : \mathbb{R}^n\setminus\{0\}\to\mathbb{R}$.
  - $R\left(cx\right) = R\left(x\right)$. Therefore, $R\left(x\right) = R\left(\dfrac{x}{\|x\|}\right)$ for $x\neq 0$.
  - Let $x\neq 0$ in $\mathbb{R}^n$ be a critical point for $R\left(x\right) = \dfrac{\left\langle Ax,x\right\rangle}{\left\langle x,x\right\rangle}$. Then x is an eigenvector for A, with the corresponding eigenvalue = $R\left(x\right)$.

Proof. $\frac{\partial R}{\partial x_j}(x) = \frac{|x|^2 \frac{\partial}{\partial x_j}\langle Ax, x\rangle - \langle Ax, x\rangle \frac{\partial}{\partial x_j}|x|^2}{|x|^4} = \frac{|x|^2 2(Ax)_j - \langle Ax, x\rangle 2x_j}{|x|^4}$. So, for critical point x, $\frac{\partial R}{\partial x_j}(x) = 0$ implies $(Ax)_j = \frac{\langle Ax, x\rangle}{\langle x, x\rangle} x_j \ \forall j$. Thus, $Ax = \frac{\langle Ax, x\rangle}{\langle x, x\rangle} x$.

- Attains its maximum value when x is an eigenvector corresponding to the largest eigenvalue.

- $g(t) = f(y+tu) = f\begin{pmatrix} y_1 + tu_1 \\ \vdots \\ y_n + tu_n \end{pmatrix} : \mathbb{R} \to \mathbb{R}$.
  - $g'(t) = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j}(y+ut)u_j$.
  - $g''(t) = \sum_{j=1}^{n}\sum_{k=1}^{n} \frac{\partial^2 f}{\partial x_k \partial x_j}(y+ut)u_k u_j = \langle d^2 f(y+ut)u, u\rangle = Q_{d^2 f(y+ut)}(u)$.

- For $n = 2$, $m = 1$, i.e. $f(x,y)$, C^2. First find x_0 where $\frac{\partial f}{\partial x}(x_0, y_0) = \frac{\partial f}{\partial y}(x_0, y_0) = 0$. Then, find $H = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial y \partial x} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix}(x_0, y_0)$. l_1 and l_2 be the eigenvalues of H. Then,
  - $\det(H) = l_1 l_2$

    Proof. $\det\left(lI - \begin{bmatrix} a & c \\ c & b \end{bmatrix}\right) = l^2 - \underbrace{(a+b)}_{l_1+l_2} l + \underbrace{(ab - c^2)}_{l_1 l_2} = 0$.
  - If $\det(H) < 0$, then (x_0, y_0) is a saddle point (neither local min nor max).

    Because l_1, l_2 don't have the same sign.
  - If $\det(H) > 0$, then (x_0, y_0) is a strict local min or max.

    Because l_1, l_2 either both positive or both negative.
    - In addition, if entries on diagonal > 0, then (x_0, y_0) is a strict local min
    - In addition, if entries on diagonal < 0, then (x_0, y_0) is a strict local min
    - Note: the sign of all entries on diagonal will be the same.
  - Hard to decide when $\det(H) = 0$.

If $f(x,y) = g(x) + h(y)$, then if x_0 is a strict local min of $g(x)$ and y_0 is a strict local min of $h(y)$, then (x_0, y_0) is a local min of $f(x,y)$.