1. let $x_1, x_2, \ldots, x_n$ are nonzero orhogonal vectors, show that they are linearly independent

假设 $x_1, x_2, \ldots, x_3$ 线性相关:
由 $x_1, x_2, \ldots, x_3$ 线性相关，得

$$k_1 x_1 + k_2 x_2 + \cdots + k_n x_n = 0$$

其中 $k_1, k_2, \ldots, k_n$ 不全为 0.
将上式左乘 x_1^T，得

$$k_1 x_1^T x_1 + k_2 x_1^T x_2 + \cdots + k_n x_1^T x_n = 0$$

由 $x_1, x_2, \ldots, x_n$ 两两正交，得 $x_i^T x_j = 0, \forall i, j \in [1, n], i \neq j$，因此

$$k_1 x_1^T x_1 + k_2 x_1^T x_2 + \cdots + k_n x_1^T x_n = k\|x_1\|_2 = 0$$

由 x_1 非零，得 $\|x_1\|_2 \neq 0$，即 $k_1 = 0$.

同理可得 $k_i = 0, \forall i \in [1, n]$，这与"$k_1, k_2, \ldots, k_n$ 不全为 0"矛盾，假设不成立.
所以 $x_1, x_2, \ldots, x_3$ 线性无关.

---

2 show $\det(\mathbf{I} + \mathbf{A}) = \det(\mathbf{I} + \mathbf{\Lambda})$ where $\mathbf{A} \triangleq \mathbf{P}\mathbf{\Lambda}\mathbf{P}^T$

由于

$$\mathbf{I} + \mathbf{A} = \mathbf{P}\mathbf{I}\mathbf{P}^T + \mathbf{P}\mathbf{\Lambda}\mathbf{P}^T = \mathbf{P}(\mathbf{I} + \mathbf{\Lambda})\mathbf{P}^T$$

有

$$\det(\mathbf{I} + \mathbf{A}) = \det(\mathbf{P}(\mathbf{I} + \mathbf{\Lambda})\mathbf{P}^T) = \det(\mathbf{P})\det(\mathbf{I} + \mathbf{\Lambda})\det(\mathbf{P}^T)$$

又由 $\mathbf{P}, \mathbf{P}^T$ 是单位正交矩阵，$\det(\mathbf{P}) = \det(\mathbf{P}^T) = 1$，得 $\det(\mathbf{I} + \mathbf{A}) = \det(\mathbf{I} + \mathbf{\Lambda})$.

---

3 prove $(\mathbf{A} + \mathbf{u}\mathbf{v}^T)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}}{1 + \mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}}$ where $\mathbf{v}^T\mathbf{A}^{-1}\mathbf{u} \neq 0$

要证

$$(\mathbf{A} + \mathbf{u}\mathbf{v}^T)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}}{1 + \mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}}$$

即证

$$(\mathbf{A} + \mathbf{u}\mathbf{v}^T)(\mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}}{1 + \mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}}) = \mathbf{I}$$

$(\mathbf{A} + \mathbf{u}\mathbf{v}^T)(\mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}}{1+\mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}}) = \mathbf{A}\mathbf{A}^{-1} + \mathbf{u}\mathbf{v}^T\mathbf{A}^{-1} - \frac{\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}}{1+\mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}} - \frac{\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}}{1+\mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}} = \mathbf{I} + \frac{\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}+\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}\mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}}{1+\mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}} - \frac{\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}}{1+\mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}} - \frac{\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^T\mathbf{A}^{-1}}{1+\mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}} = \mathbf{I} +$

$\frac{\mathbf{uv}^T\mathbf{A}^{-1}\mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}-\mathbf{uv}^T\mathbf{A}^{-1}\mathbf{uv}^T\mathbf{A}^{-1}}{1+\mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}} = \mathbf{I} + \frac{\mathbf{uv}^T\mathbf{A}^{-1}(\mathbf{v}^T\mathbf{A}^{-1}\mathbf{u})-\mathbf{u}(\mathbf{v}^T\mathbf{A}^{-1}\mathbf{u})\mathbf{v}^T\mathbf{A}^{-1}}{1+\mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}} = \mathbf{I} + \frac{(\mathbf{v}^T\mathbf{A}^{-1}\mathbf{u})(\mathbf{uv}^T\mathbf{A}^{-1}-\mathbf{uv}^T\mathbf{A}^{-1})}{1+\mathbf{v}^T\mathbf{A}^{-1}\mathbf{u}} = \mathbf{I}$

原命题得证.

---

4 compute the first and second derivative of $g(t) \triangleq f(\mathbf{x} + t(\mathbf{y} - \mathbf{x}))$

令 $\mathbf{z} \triangleq \mathbf{x} + t(\mathbf{y} - \mathbf{x})$，则

$g'(t) = \sum_i \frac{\partial f(z)}{\partial zi}\frac{\partial z_i}{\partial t} = \nabla f(\mathbf{x} + \mathbf{t}(\mathbf{y} - \mathbf{x}))^T \cdot (\mathbf{y} - \mathbf{x})$

$g''(t) = \frac{d}{dt}(\nabla f(\mathbf{z})) \cdot (\mathbf{y} - \mathbf{x}) + \nabla f(\mathbf{z}) \cdot \frac{d}{dt}(\mathbf{y} - \mathbf{x}) = diag(\nabla^2 f(\mathbf{x} + \mathbf{t}(\mathbf{y} - \mathbf{x})))^T \cdot (\mathbf{y} - \mathbf{x})$

---

5 compute the gradient and Hessian of $f(x, y) \triangleq \frac{y^2}{x}$ where $x > 0$

$\nabla f(x, y) = \left(\frac{\partial f(x,y)}{\partial x}, \frac{\partial f(x,y)}{\partial y}\right)^T = \left(-\frac{y^2}{x^2}, \frac{2y}{x}\right)^T$

$\nabla^2 f(x, y) = \begin{pmatrix} \frac{2y^2}{x^3} & -\frac{2y}{x^2} \\ -\frac{2y}{x^2} & \frac{2}{x} \end{pmatrix}$

---

6 compute the gradient and Hessian of $f(\mathbf{x}) \triangleq \log\sum_{k=1}^n \exp(x_k)$

$\nabla f(\mathbf{x}) = \left(\frac{exp(x1)}{\sum_{k=1}^n \exp(xk)}, \frac{exp(x2)}{\sum_{k=1}^n \exp(xk)}, \dots, \frac{exp(xn)}{\sum_{k=1}^n \exp(xk)}\right)^T$

$\nabla^2 f(\mathbf{x}) = \begin{pmatrix} \frac{\exp(x1)\left(\sum_{k=1}^n \exp(xk)\right)-\exp(x1)^2}{\left(\sum_{k=1}^n \exp(xk)\right)^2} & \frac{-\exp(x1)\exp(x2)}{\left(\sum_{k=1}^n \exp(xk)\right)^2} & \cdots & \frac{-\exp(x1)\exp(xn)}{\left(\sum_{k=1}^n \exp(xk)\right)^2} \\ \frac{-\exp(x2)\exp(x1)}{\left(\sum_{k=1}^n \exp(xk)\right)^2} & \frac{\exp(x2)\left(\sum_{k=1}^n \exp(xk)\right)-\exp(x2)^2}{\left(\sum_{k=1}^n \exp(xk)\right)^2} & \cdots & \frac{-\exp(x2)\exp(xn)}{\left(\sum_{k=1}^n \exp(xk)\right)^2} \\ \vdots & \vdots & & \vdots \\ \frac{-\exp(xn)\exp(x1)}{\left(\sum_{k=1}^n \exp(xk)\right)^2} & \frac{-\exp(xn)\exp(x2)}{\left(\sum_{k=1}^n \exp(xk)\right)^2} & \cdots & \frac{\exp(xn)\left(\sum_{k=1}^n \exp(xk)\right)-\exp(xn)^2}{\left(\sum_{k=1}^n \exp(xk)\right)^2} \end{pmatrix}$

---

7 Assume that one makes m measurements for each of n objects, and collect these data in columns $b_1, b_2, \dots, b_n \in \mathbb{R}^m$. For example, $m = 3, n = 8$

	P1	P2	P3	P4	P5	P6	P7	P8
age	22	30	23	23	22	21	22	21
weight	10.4	12.2	10.5	10.9	9	12.5	11.5	10.2
shoe size	7	8	7	7	8	8	9	7

If one want to distinguish these 8 people by a linear combination of the 3 measurements, what would be a best possible combination?