1. Let $A \in \Re^{nxm}, rank(A) = min\{m, n\}$ and the SVD decompostion of A $A = U\Sigma V^T$
where $U \in \Re^{nxn}, V \in \Re^{mxm}$ orthogonal matrices and $\Sigma \in \Re^{nxm}$ diagonal matrix.

   (a) $m < n$ need to prove $A^\dagger = (A^T A)^{-1} A^T$

$$A^T A = (U\Sigma V^T)^T U\Sigma V^T = V\Sigma^T U^T U\Sigma V^T = V\Sigma^T \Sigma V^T$$

$$\Sigma^T \Sigma \in \mathbb{R}^{mxm}, \quad [\Sigma^T \Sigma]_{ij} = \sum_{i=1}^{n} \Sigma^T_{ik} \Sigma_{kj} = \Sigma^2_{ii}$$

$$(A^T A)^{-1} = (V\Sigma^T \Sigma V^T)^{-1} = (V^T)^{-1} (\Sigma^T \Sigma)^{-1} V^{-1} = V(\Sigma^\dagger)(\Sigma^\dagger)^T V^T$$
$$(A^T A)^{-1} A^T = V(\Sigma^\dagger)(\Sigma^\dagger)^T V^T V \Sigma^T U^T = V\Sigma^\dagger (\Sigma^\dagger)^T \Sigma U^T = V\Sigma^\dagger U^T = A^\dagger$$

   (b) $n < m$ need to prove $A^\dagger = A^T (AA^T)^{-1}$

$$AA^T = U\Sigma V^T (U\Sigma V^T)^T = U\Sigma V^T V \Sigma^T U^T = U\Sigma\Sigma^T U^T$$

$$\Sigma\Sigma^T \in \mathbb{R}^{nxn}, \quad [\Sigma\Sigma^T]_{ij} = \sum_{i=1}^{m} \Sigma^T_{ik} \Sigma_{kj} = \Sigma^2_{ii}$$

$$((AA^T))^{-1} = (U\Sigma\Sigma^T U^T)^{-1} = (U^T)^{-1} (\Sigma\Sigma^T)^{-1} U^{-1} = U(\Sigma^\dagger)^T (\Sigma^\dagger) U^T$$
$$A^T (AA^T)^{-1} = V\Sigma^T U^T U (\Sigma^\dagger)^T \Sigma^\dagger U^T = V\Sigma^T \Sigma^\dagger (\Sigma^\dagger)^T U^T = V\Sigma^\dagger U^T$$

2. Let $f(x) : \mathbb{R}^n \to \mathbb{R}$, $f(x) = (y - Hx)^T W (y - Hx)$

$$\nabla f(x) = -2(y - Hx)^T W H = -2(y^T W H - x^T H^T W H) = 0$$
$$y^T W H = x^T H^T W H \Rightarrow Hx = y \Rightarrow x = yH^\dagger$$

3. $y \in \mathbb{R}^n$
Optimallity conditions:
Let f(x) be a differentiable and convex and C is a convex set, Then x is a global solution iff $x \in C$ and

$$(\nabla f(x))^T (z - x) \geq 0 \ \ \forall z \in C$$

   (a) $C = \{x \in \mathbb{R}^n : \|x\|_2 \leq 1\}$
   Let's guess the projection on the unit ball to be : $\frac{y}{||y||}$

$$P_C(y) = \underset{x \in C}{argmin} \|x - y\|_2$$

$$\underset{s.t\ \|x\|_2 \leq 1}{minimize} \|x - y\|_2 \Leftrightarrow \underset{s.t\ \|x\|_2 \leq 1}{minimize} (x - y)^T (x - y)$$

$f(x) = (x-y)^T(x-y)$
$\nabla f(x) = 2(x-y)$
$(\nabla f(x))^T(y-x) = 2(x-y)^T(y-x)$
Now let's make sure that $x = \frac{y}{||y||}$ satisfies the optimallity conditions, so we need to show:

$$\forall z \in C, \quad 2(\frac{y}{||y||} - y)^T(z - \frac{y}{||y||}) \geq 0$$

$$\Leftrightarrow (\frac{y}{||y||} - y)^T(z - \frac{y}{||y||}) \geq 0 \Leftrightarrow \left[\frac{y^T z}{||y||} - 1 - y^T z + ||y||\right] \geq 0 \Leftrightarrow$$

$$y^T z \left(\frac{1-||y||}{||y||}\right) - (1-||y||) \geq 0 \Leftrightarrow (1-||y||)\left(\frac{y^T z - ||y||}{||y||}\right) \geq 0 \Leftrightarrow$$

$$(1-||y||)\left(y^T z - ||y||\right) \geq 0$$

For $||y|| \leq 1$ it's clear that $P_c(y) = y$, let's concentrate on the case where $||y|| > 1 \Rightarrow 1 - ||y|| \leq 0$
We know $||z|| \leq 1$ and from Cauchy-Shwarz $|y^T z| \leq ||z|| \cdot ||y|| \leq ||y||$
$\Rightarrow y^T z - ||y|| \leq 0$
So finally we know that

$$(1-||y||)\left(y^T z - ||y||\right) \geq 0$$

(b) The projection is $P_c(y)_i = min(1, |x_i|)sign(x_i)$

4. $f(x) : \mathbb{R}^2 \to \mathbb{R}, \;\; f(x) = \frac{1}{2}(x_1^2 + \gamma x_2^2), \;\; \gamma > 0$

   (a) $\nabla f(x) = \frac{1}{2}\begin{bmatrix} 2x_1 \\ 2\gamma x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ \gamma x_2 \end{bmatrix}$, $\nabla^2 f(x) = \begin{bmatrix} 1 & 0 \\ 0 & \gamma \end{bmatrix}$ since $\gamma > 0$ we know that $\nabla^2 f(x) \succeq 0$ and so f(x) is convex.
   To find the optimal $x^* = argmin_x f(x)$ we require that:

$$\nabla f(x) = 0 \Rightarrow \begin{bmatrix} x_1 \\ \gamma x_2 \end{bmatrix} = 0 \Rightarrow x^* = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

   (b) $x_0 = \begin{bmatrix} \gamma \\ 1 \end{bmatrix}, \;\; \nabla f(x^{i-1}) = \begin{bmatrix} x_1^{i-1} \\ \gamma x_2^{i-1} \end{bmatrix}$

$$t_i = \underset{t}{argmin} f(x_{i-1} - t\nabla f(x_{i-1})) = \underset{t}{argmin} f(x_{i-1} - t\begin{bmatrix} x_1^{i-1} \\ \gamma x_2^{i-1} \end{bmatrix}) =$$

$$\underset{t}{argmin} f(\begin{bmatrix} x_1^{i-1} - t x_1^{i-1} \\ x_2^{i-1} - \gamma t x_2^{i-1} \end{bmatrix}) = \underset{t}{argmin} \frac{1}{2}\left[(x_1^{i-1} - t x_1^{i-1})^2 + \gamma(x_2^{i-1} - \gamma t x_2^{i-1})^2\right] =$$

$$\underset{t}{argmin} \frac{1}{2}\left[(x_1^{i-1})^2 - 2t(x_1^{i-1})^2 + t^2(x_1^{i-1})^2 + \gamma(x_2^{i-1})^2 - 2\gamma^2 t(x_2^{i-1})^2 + \gamma^3 t^2(x_2^{i-1})^2)\right] =$$

$$\underset{t}{argmin}\frac{1}{2}\left[(x_1^{i-1})^2+\gamma(x_2^{i-1})^2-2t\left[(x_1^{i-1})^2+\gamma^2(x_2^{i-1})^2\right]+t^2\left[(x_1^{i-1})^2+\gamma^3(x_2^{i-1})^2\right]\right]$$

Find minimum by deriving the function by t:

$$-\left[(x_1^{i-1})^2+\gamma^2(x_2^{i-1})^2\right]+t\left[(x_1^{i-1})^2+\gamma^3(x_2^{i-1})^2\right]=0$$

$$t\left[(x_1^{i-1})^2+\gamma^3(x_2^{i-1})^2\right]=\left[(x_1^{i-1})^2+\gamma^2(x_2^{i-1})^2\right]$$

$$t_i=\frac{(x_1^{i-1})^2+\gamma^2(x_2^{i-1})^2}{(x_1^{i-1})^2+\gamma^3(x_2^{i-1})^2}$$

$$x_i=x^{i-1}-t_i\nabla f(x^{i-1})=x^{i-1}-\frac{(x_1^{i-1})^2+\gamma^2(x_2^{i-1})^2}{(x_1^{i-1})^2+\gamma^3(x_2^{i-1})^2}\begin{bmatrix}x_1^{i-1}\\ \gamma x_2^{i-1}\end{bmatrix}=$$

$$\begin{bmatrix}x_1^{i-1}-x_1^{i-1}\frac{(x_1^{i-1})^2+\gamma^2(x_2^{i-1})^2}{(x_1^{i-1})^2+\gamma^3(x_2^{i-1})^2}\\ \\ x_2^{i-1}-\gamma x_2^{i-1}\frac{(x_1^{i-1})^2+\gamma^2(x_2^{i-1})^2}{(x_1^{i-1})^2+\gamma^3(x_2^{i-1})^2}\end{bmatrix}=\begin{bmatrix}x_1^{i-1}\left[1-\frac{(x_1^{i-1})^2+\gamma^2(x_2^{i-1})^2}{(x_1^{i-1})^2+\gamma^3(x_2^{i-1})^2}\right]\\ \\ x_2^{i-1}\left[1-\gamma\frac{(x_1^{i-1})^2+\gamma^2(x_2^{i-1})^2}{(x_1^{i-1})^2+\gamma^3(x_2^{i-1})^2}\right]\end{bmatrix}=$$

$$\begin{bmatrix}x_1^{i-1}\left[\frac{(x_1^{i-1})^2+\gamma^3(x_2^{i-1})^2-(x_1^{i-1})^2-\gamma^2(x_2^{i-1})^2}{(x_1^{i-1})^2+\gamma^3(x_2^{i-1})^2}\right]\\ \\ x_2^{i-1}\left[\frac{(x_1^{i-1})^2+\gamma^3(x_2^{i-1})^2-\gamma(x_1^{i-1})^2-\gamma^3(x_2^{i-1})^2}{(x_1^{i-1})^2+\gamma^3(x_2^{i-1})^2}\right]\end{bmatrix}=$$

$$=\begin{bmatrix}x_1^{i-1}\left[\frac{\gamma^3(x_2^{i-1})^2-\gamma^2(x_2^{i-1})^2}{(x_1^{i-1})^2+\gamma^3(x_2^{i-1})^2}\right]\\ \\ x_2^{i-1}\left[\frac{(x_1^{i-1})^2-\gamma(x_1^{i-1})^2}{(x_1^{i-1})^2+\gamma^3(x_2^{i-1})^2}\right]\end{bmatrix}=\begin{bmatrix}x_1^{i-1}\left[\frac{(\gamma-1)\gamma^2(x_2^{i-1})^2}{(x_1^{i-1})^2+\gamma^3(x_2^{i-1})^2}\right]\\ \\ x_2^{i-1}\left[\frac{(1-\gamma)(x_1^{i-1})^2}{(x_1^{i-1})^2+\gamma^3(x_2^{i-1})^2}\right]\end{bmatrix}=$$

$$=\frac{(\gamma-1)}{(x_1^{i-1})^2+\gamma^3(x_2^{i-1})^2}\begin{bmatrix}\gamma^2x_1^{i-1}(x_2^{i-1})^2\\ \\ -x_2^{i-1}(x_1^{i-1})^2\end{bmatrix}=\frac{(\gamma-1)x_2^{i-1}x_1^{i-1}}{(x_1^{i-1})^2+\gamma^3(x_2^{i-1})^2}\begin{bmatrix}\gamma^2x_2^{i-1}\\ -x_1^{i-1}\end{bmatrix}$$

$$x_i=\frac{(\gamma-1)x_2^{i-1}x_1^{i-1}}{(x_1^{i-1})^2+\gamma^3(x_2^{i-1})^2}\begin{bmatrix}\gamma^2x_2^{i-1}\\ -x_1^{i-1}\end{bmatrix}$$