Exercise 1: Lagrange Multipliers (10 + 10 P)

Let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N \in \mathbb{R}^d$ be a dataset of N data points. We consider the objective function

$$J(\boldsymbol{\theta}) = \sum_{k=1}^{N} \|\boldsymbol{\theta} - \boldsymbol{x}_k\|^2$$

to be minimized with respect to the parameter $\boldsymbol{\theta} \in \mathbb{R}^d$. In absence of constraints, the parameter $\boldsymbol{\theta}$ that minimizes this objective is given by the empirical mean $\boldsymbol{m} = \frac{1}{N}\sum_{k=1}^{N} \boldsymbol{x}_k$. However, this is generally not the case when the parameter $\boldsymbol{\theta}$ is constrained.

(a) Using the method of Lagrange multipliers, *find* the parameter $\boldsymbol{\theta}$ that minimizes $J(\boldsymbol{\theta})$ subject to the constraint $\boldsymbol{\theta}^\top \boldsymbol{b} = 0$, with $\boldsymbol{b}$ some unit vector in $\mathbb{R}^d$. Give a geometrical interpretation to your solution.

$$J(\vartheta) = \sum_{k=1}^{N} \|\vartheta - x_k\|^2 \qquad m = \frac{1}{N}\sum_{k=1}^{N} x_k \qquad \vartheta^T b = 0.$$

$$\mathcal{L}(\vartheta, \lambda) = \sum_{k=1}^{N} \|\vartheta - x_k\|^2 + \lambda \vartheta^T b$$

$$\nabla_{\vartheta_i} \mathcal{L}(\vartheta, \lambda) = \frac{\partial}{\partial \vartheta_i}\left(\sum_{k=1}^{N}\left[\sum_{j=1}^{d} (\vartheta_j - x_k)^2\right] + \lambda \vartheta^T b\right)$$

$$= \sum_{k=1}^{N} 2(\vartheta_i - x_k) + \lambda b_i$$

$$= 2N\vartheta_i - 2\sum_{k}^{N} x_k + \lambda b_i$$

Necessary condition: $\nabla_\vartheta \mathcal{L}(\vartheta, \lambda) = 0.$

$$\nabla_\vartheta \mathcal{L} = 2N\vartheta - 2\sum x_k + \lambda b = 0$$

$$\Rightarrow \vartheta - m + \frac{\lambda}{2N} b = 0$$

$$\Rightarrow \vartheta = m - \frac{\lambda}{2N} b$$

Applying the constraint we have:

$$-\vartheta^T b = -m^T b - \frac{\lambda}{2N} \underbrace{b\, b^T}_{=1 \text{ because } b \text{ is unitary}} = 0$$

$$\Rightarrow \frac{\lambda}{2N} = m^T b$$

$$\Rightarrow \vartheta = m + (m^T b) b = m + m^T b b$$

The geometric interpretation is given by the constraint $\vartheta^T b = 0$, which entails that we are considering only the hyperplane orthogonal to b for our optimisation.

(b) Using the same method, *find* the parameter $\boldsymbol{\theta}$ that minimizes $J(\boldsymbol{\theta})$ subject to $\|\boldsymbol{\theta}-\mathbf{c}\|^2 = 1$, where $\boldsymbol{c}$ is a vector in $\mathbb{R}^d$ different from $\boldsymbol{m}$. Give a geometrical interpretation to your solution.

$$g(\vartheta, \lambda) = \|\vartheta - c\|^2 - 1$$

$$\mathcal{L}(\vartheta, \lambda) = \sum_{k=1}^{N} \|\vartheta - x_k\|^2 + \lambda\left(\|\vartheta - c\|^2 - 1\right)$$

$$\nabla_{\vartheta_i} \mathcal{L}(\vartheta, \lambda) = \sum_{k=1}^{N} 2(\vartheta_i - x_k) + 2\lambda(\vartheta_i - c)$$

$$= 2N\vartheta_i - 2\sum_k x_k + 2\lambda\vartheta_i - 2c\lambda$$

$$\nabla_\vartheta \mathcal{L} = \vartheta - 2m + \frac{\lambda}{N}\vartheta - \frac{c\lambda}{N} = 0$$

$$\vartheta\left(1 + \frac{\lambda}{N}\right) - 2m - \frac{c\lambda}{N} = 0$$

$$\vartheta\left(1 + \frac{\lambda}{N}\right) - 2m - \frac{c\lambda}{N} + c - c = 0$$

$$\vartheta\left(1 + \frac{\lambda}{N}\right) - c\left(1 + \frac{\lambda}{N}\right) - 2m = 0$$

$$(\vartheta - c)\left(1 + \frac{\lambda}{N}\right) = 2m - c$$

$$\left(1 + \frac{\lambda}{N}\right)^2 = \|2m - c\|^2$$

$$\Rightarrow 1 + \frac{\lambda}{N} = \pm\|2m - c\|$$

$$\frac{\lambda}{N} = \pm\|2m - c\| - 1$$

$$\Rightarrow (\vartheta - c)\left(1 \pm \|2m - c\| - 1\right) = 2m - c$$

$$\vartheta = \pm\frac{2m - c}{\|2m - c\|} + c$$

The geometric interpretation is given by the constraint $\|\vartheta - c\|^2 = 1$, which entails that we are considering only the unit hypersphere centered in c for our optimisation.