Problem 1:

(a) To solve this problem, I took advantage of the fact that $z_{ij} = \boldsymbol{x}_i^T \boldsymbol{v}_j$ (1-2) and $\boldsymbol{v}_j^T \boldsymbol{v}_j = 1$ (3-4), as the problem statement suggested. Beyond the few points where I used those substitutions, the steps shown below use simple algebraic steps.

$$\begin{aligned}
\left\| \boldsymbol{x}_i - \sum_{j=1}^{k} z_{ij} \boldsymbol{v}_j \right\|_2^2 &= \left(\boldsymbol{x}_i - \sum_{j=1}^{k} z_{ij} \boldsymbol{v}_j \right)^T \left(\boldsymbol{x}_i - \sum_{j=1}^{k} z_{ij} \boldsymbol{v}_j \right) \\
&= \boldsymbol{x}_i^T \boldsymbol{x}_i - \boldsymbol{x}_i^T \sum_{j=1}^{k} z_{ij} \boldsymbol{v}_j - \boldsymbol{x}_i \left(\sum_{j=1}^{k} z_{ij} \boldsymbol{v}_j \right)^T + \left(\sum_{j=1}^{k} z_{ij} \boldsymbol{v}_j \right)^T \sum_{j=1}^{k} z_{ij} \boldsymbol{v}_j \\
&= \boldsymbol{x}_i^T \boldsymbol{x}_i - 2 \sum_{j=1}^{k} \boldsymbol{v}_j^T \boldsymbol{x}_i \boldsymbol{x}_i^T \boldsymbol{v}_j + \sum_{j=1}^{k} \boldsymbol{v}_j^T z_{ij}^T z_{ij} \boldsymbol{v}_j \\
&= \boldsymbol{x}_i^T \boldsymbol{x}_i - 2 \sum_{j=1}^{k} \boldsymbol{v}_j^T \boldsymbol{x}_i \boldsymbol{x}_i^T \boldsymbol{v}_j + \sum_{j=1}^{k} \boldsymbol{v}_j^T (\boldsymbol{x}_i^T \boldsymbol{v}_j)^T (\boldsymbol{x}_i^T \boldsymbol{v}_j) \boldsymbol{v}_j \qquad (1) \\
&= \boldsymbol{x}_i^T \boldsymbol{x}_i - 2 \sum_{j=1}^{k} \boldsymbol{v}_j^T \boldsymbol{x}_i \boldsymbol{x}_i^T \boldsymbol{v}_j \boldsymbol{v}_j + \sum_{j=1}^{k} \boldsymbol{v}_j^T \boldsymbol{v}_j^T \boldsymbol{x}_i \boldsymbol{x}_i^T \boldsymbol{v}_j \boldsymbol{v}_j \qquad (2) \\
&= \boldsymbol{x}_i^T \boldsymbol{x}_i - 2 \sum_{j=1}^{k} \boldsymbol{v}_j^T \boldsymbol{x}_i \boldsymbol{x}_i^T \boldsymbol{v}_j + \sum_{j=1}^{k} \boldsymbol{v}_j^T \boldsymbol{v}_j^T (\boldsymbol{x}_i \boldsymbol{x}_i^T) \boldsymbol{v}_j \boldsymbol{v}_j \\
&= \boldsymbol{x}_i^T \boldsymbol{x}_i - 2 \sum_{j=1}^{k} \boldsymbol{v}_j^T \boldsymbol{x}_i \boldsymbol{x}_i^T \boldsymbol{v}_j + \sum_{j=1}^{k} \boldsymbol{v}_j^T (\boldsymbol{x}_i \boldsymbol{x}_i^T) (\boldsymbol{v}_j^T \boldsymbol{v}_j) \boldsymbol{v}_j \qquad (3) \\
&= \boldsymbol{x}_i^T \boldsymbol{x}_i - 2 \sum_{j=1}^{k} \boldsymbol{v}_j^T \boldsymbol{x}_i \boldsymbol{x}_i^T \boldsymbol{v}_j + \sum_{j=1}^{k} \boldsymbol{v}_j^T (\boldsymbol{x}_i \boldsymbol{x}_i^T) (1) \boldsymbol{v}_j \qquad (4) \\
&= \boldsymbol{x}_i^T \boldsymbol{x}_i - 2 \sum_{j=1}^{k} \boldsymbol{v}_j^T \boldsymbol{x}_i \boldsymbol{x}_i^T \boldsymbol{v}_j + \sum_{j=1}^{k} \boldsymbol{v}_j^T \boldsymbol{x}_i \boldsymbol{x}_i^T \boldsymbol{v}_j \\
&= \boldsymbol{x}_i^T \boldsymbol{x}_i - \sum_{j=1}^{k} \boldsymbol{v}_j^T \boldsymbol{x}_i \boldsymbol{x}_i^T \boldsymbol{v}_j,
\end{aligned}$$

as desired.

(b) To solve this problem, I took advantage of the fact that $\sum_{i=1}^{n} \boldsymbol{v}_j^T \boldsymbol{x}_i \boldsymbol{x}_i^T \boldsymbol{v}_j = \boldsymbol{v}_j^T \boldsymbol{\Sigma} \boldsymbol{v}_j = \lambda_j \boldsymbol{v}_j^T \boldsymbol{v}_j$ (5-7), as the problem statement suggested, and that $\boldsymbol{v}_j^T \boldsymbol{v}_j = 1$ (7-8), from the previous problem. Beyond the few points where I used those substitutions, the steps shown below use simple algebraic steps.

$$\begin{aligned} J_k = \frac{1}{n} \sum_{i=1}^{n} \left(\boldsymbol{x}_i^T \boldsymbol{x}_i - \sum_{j=1}^{k} \boldsymbol{v}_j^T \boldsymbol{x}_i \boldsymbol{x}_i^T \boldsymbol{v}_j \right) &= \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i^T \boldsymbol{x}_i - \frac{1}{n} \sum_{i=1}^{n} \left(\sum_{j=1}^{k} \boldsymbol{v}_j^T \boldsymbol{x}_i \boldsymbol{x}_i^T \boldsymbol{v}_j \right) && (5) \\ &= \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i^T \boldsymbol{x}_i - \sum_{j=1}^{k} \boldsymbol{v}_j^T \boldsymbol{\Sigma} \boldsymbol{v}_j && (6) \\ &= \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i^T \boldsymbol{x}_i - \sum_{j=1}^{k} \lambda_j \boldsymbol{v}_j^T \boldsymbol{v}_j && (7) \\ &= \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i^T \boldsymbol{x}_i - \sum_{j=1}^{k} \lambda_j (1) && (8) \\ &= \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i^T \boldsymbol{x}_i - \sum_{j=1}^{k} \lambda_j. \end{aligned}$$

(c) Using our insights from part (b), notice that

$$J_d = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i^T \boldsymbol{x}_i - \sum_{j=1}^{d} \lambda_j,$$

which implies that

$$J_k = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i^T \boldsymbol{x}_i - \sum_{j=1}^{k} \lambda_j.$$

We can expand J_d, since $k < d$, as the problem statement suggests, to become

$$J_d = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i^T \boldsymbol{x}_i - \sum_{j=1}^{k} \lambda_j - \sum_{j=k+1}^{d} \lambda_j.$$

Furthermore, since $J_d = 0$, we know that

$$J_d = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i^T \boldsymbol{x}_i - \sum_{j=1}^{k} \lambda_j - \sum_{j=k+1}^{d} \lambda_j = 0.$$

Rearranging this, we see

$$\sum_{j=k+1}^{d} \lambda_j = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{x}_i^T \boldsymbol{x}_i - \sum_{j=1}^{k} \lambda_j = J_k,$$

as desired.

Problem 2:

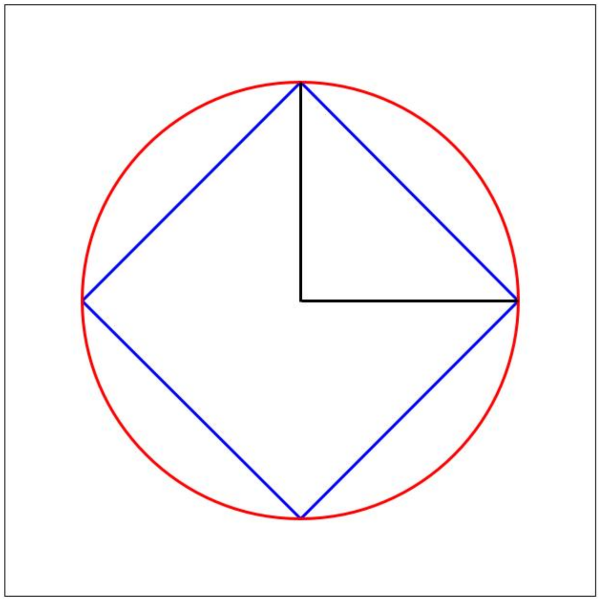

Figure 1: This figure shows both the norm-ball $B_k = \{\boldsymbol{x} : \|\boldsymbol{x}\|_1 \leq k\}$ in blue, and the norm-ball $A_k = \{\boldsymbol{x} : \|\boldsymbol{x}\|_2 \leq k\}$ in red, where $k = 1$ (and so too does the black line).

I wasn't able to solve the second part of this problem myself, so I referred to the solution sheet on the MATH 189 website. After reviewing that, I now understand that the optimization problem in question is equivalent to the Lagrangian expansion:

$$inf_x sup_{\lambda \geq 0} \mathcal{L}(\boldsymbol{x}, \lambda) = inf_x sup_{\lambda \geq 0} f(\boldsymbol{x}) + \lambda(\|\boldsymbol{x}\|_p - k).$$

Which we can flip the inf and sup to become:

$$sup_{\lambda \geq 0} inf_x f(\boldsymbol{x}) + \lambda(\|\boldsymbol{x}\|_p - k) = sup_{\lambda \geq 0} g(\lambda).$$

Optimizing $\boldsymbol{x}$ to minimize $f(\boldsymbol{x}) + \lambda(\|\boldsymbol{x}\|_p - k)$ is the same as optimizing $\boldsymbol{x}$ to minimize $f(\boldsymbol{x}) + \lambda \|\boldsymbol{x}\|_p$, thereby solving the optimization problem

$$\text{minimize: } f(\boldsymbol{x}) + \lambda \|\boldsymbol{x}\|_p$$

for some $\lambda \geq 0$, as desired.

Viewed this way, ℓ_1 regularization can be seen as projecting the optimal solution onto some ℓ_1 norm ball. The ℓ_1 norm-ball has less surface area along its faces than the ℓ_2 norm ball, meaning the optimal solution is more likely to land on an edge (where the x or y dimension of the solution is 0) in the ℓ_1 case than in the ℓ_2 case. This feature generalizes to higher dimensions too, meaning the ℓ_1 norm-ball projection is more likely to generate 0-weights dimensions, relative to the ℓ_2 and ℓ_n (where $n > 2$) cases.