1

a) $\|x_i - \sum_{j=1}^{k} z_{ij} v_j\|_2^2 = (x_i - \sum_{j=1}^{k} z_{ij} v_j)^T (x_i - \sum_{j=1}^{k} z_{ij} v_j)$

$= x_i^T x_i - \sum_{j=1}^{k} z_{ij} v_j^T x_i - x_i^T \sum_{j=1}^{k} z_{ij} v_j + (\sum_{j=1}^{k} z_{ij} v_j)^T (\sum_{j=1}^{k} z_{ij} v_j)$

$= x_i^T x_i - 2\sum_{j=1}^{k} z_{ij} v_j^T x_i + (\sum_{j=1}^{k} z_{ij} v_j^T) \cdot (\sum_{j=1}^{k} z_{ij} v_j)$

$= x_i^T x_i - 2\sum_{j=1}^{k} z_{ij} v_j^T x_i + \sum_{j=1}^{k} v_j^T z_{ij}^T z_{ij} v_j$

$= x_i^T x_i - 2\sum_{j=1}^{k} z_{ij} v_j^T x_i + \sum_{j=1}^{k} v_j^T x_i x_i^T v_j$

$= x_i^T x_i - 2\sum_{j=1}^{k} v_j^T x_i x_i^T v_j + \sum_{j=1}^{k} v_j^T x_i x_i^T v_j$

$= x_i^T x_i - \sum_{j=1}^{k} v_j^T x_i x_i^T v_j$

b) $J_k = \frac{1}{n} \sum_{i=1}^{n} (x_i^T x_i - \sum_{j=1}^{k} v_j^T x_i x_i^T v_j)$

$= \frac{1}{n} \sum_{i=1}^{n} x_i^T x_i - \sum_{j=1}^{k} v_j^T \frac{1}{n} (\sum_{i=1}^{n} x_i x_i^T) v_j$

$= \frac{1}{n} \sum_{i=1}^{n} x_i^T x_i - \sum_{j=1}^{k} v_j^T \Sigma v_j$

$= \frac{1}{n} \sum_{i=1}^{n} x_i^T x_i - \sum_{j=1}^{k} v_j^T \Sigma v_j$

$= \frac{1}{n} \sum_{i=1}^{n} x_i^T x_i - \sum_{j=1}^{k} \lambda_j$

c) $J_d = 0 \qquad \sum_{j=1}^{d} \lambda_j = \frac{1}{n} \sum_{i=1}^{n} x_i^T x_i$

$J_k = \frac{1}{n} \sum_{i=1}^{n} x_i^T x_i - \sum_{j=1}^{k} \lambda_j + \sum_{j=k+1}^{d} \lambda_j = \sum_{j=k+1}^{d} \lambda_j$

2. From optimization prob! min $(f(x)$ subj to $\|x\|_p \le k)$

→ Form Lagrangian

→ $L(x, \lambda) = f(x) + \lambda (\|x\|_p - k)$ where λ is multiplier

We have an optimization problem where we want to minimize some function f(x) subject to a constraint that the p-norm of x must be less than or equal to some value k.
To solve this, we can construct something called a Lagrangian by adding a new term that includes a Lagrange multiplier lambda multiplied by the constraint equation.
This converts the constrained optimization into an unconstrained one where we minimize f(x) plus an additional term proportional to the norm of x.
Setting lambda >= 0 and finding the point where the gradient of the Lagrangian is zero enforces the original constraint in the solution.
Adding an L1 regularization term lambda*||x||_1 tends to make more coefficients zero due to its linear slope, making the solution more sparse and interpretable. In contrast, L2 regularization with lambda*||x||_2 shrinks coefficients more evenly without setting them exactly to zero, giving a less sparse solution. This difference comes from the sharp corners of the L1 norm at zero versus the smooth contour of the L2 norm.