

4.

$$\mathbb{E}[\|X - \mu\|^2] = \mathbb{E}[(X - \mu)^T(X - \mu)] \quad (1)$$

$$\mathbb{E}[\|X - \mu\|^2] = \mathbb{E}[X^T X - 2X^T \mu + \mu^T \mu] \quad (2)$$

$$\mathbb{E}[\|X - \mu\|^2] = \mathbb{E}[X^T X] - \mu^T \mu \quad (3)$$

Now let us look at the term $\mathbb{E}[X^T X]$. It is equivalent to $\mathbb{E}[x_1^2 + x_2^2 + \dots + x_d^2]$. We know that expected value of the sum of variables is equivalent to the sum of expected values of individual variables:

$$\mathbb{E}[x_1^2 + x_2^2 + \dots + x_d^2] = \mathbb{E}[x_1^2] + \mathbb{E}[x_2^2] + \dots + \mathbb{E}[x_d^2] \quad (4)$$

On all of the terms we can apply the theorem of Variance Translation:

$$\text{Var}(X) = \mathbb{E}[X^2] - \mu^2 \quad (5)$$

$$\mathbb{E}[x_1^2] + \mathbb{E}[x_2^2] + \dots + \mathbb{E}[x_d^2] = (\text{Var}(x_1) + \mu_1^2) + (\text{Var}(x_2) + \mu_2^2) + \dots + (\text{Var}(x_d) + \mu_d^2) \quad (6)$$

We know that the diagonal values on the covariance matrix are variances for individual variables. The sum of them is equal to the trace of the matrix. If we also write the sum of individual μ values in matrix form, we get:

$$\mathbb{E}[X^T X] = \text{tr}(C) + \mu^T \mu \quad (7)$$

Substituting this into (3), we get:

$$\mathbb{E}[\|X - \mu\|^2] = \text{tr}(C) + \mu^T \mu - \mu^T \mu = \text{tr}(C) \quad (8)$$

This completes the proof.