1.

Let

$$y = \Sigma^{-\frac{1}{2}}(x - \mu),$$

where $\Sigma^{\frac{1}{2}}$ is the unique symmetric positive definite square root of Σ satisfying $\Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}} = \Sigma.$

Then $x = \mu + \Sigma^{\frac{1}{2}}y,$ and the Jacobian determinant of this transformation is

$$\left|\frac{\partial x}{\partial y}\right| = \left|\Sigma^{\frac{1}{2}}\right| = |\Sigma|^{\frac{1}{2}}.$$

Substituting $x = \mu + \Sigma^{\frac{1}{2}}y$ gives:

$$\begin{aligned}\int_{\mathbb{R}^k} f(x)dx &= \frac{1}{(2\pi)^{\frac{k}{2}}|\Sigma|^{\frac{1}{2}}}\int_{\mathbb{R}^k} e^{\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]}dx \\ &= \frac{1}{(2\pi)^{\frac{k}{2}}|\Sigma|^{\frac{1}{2}}}\int_{\mathbb{R}^k} e^{\left[-\frac{1}{2}y^Ty\right]}|\Sigma|^{\frac{1}{2}}dy \\ &= \frac{1}{(2\pi)^{\frac{k}{2}}}\int_{\mathbb{R}^k} e^{-\frac{1}{2}\|y\|^2}dy.\end{aligned}$$

The integrand factorizes across coordinates:

$$\int_{\mathbb{R}^k} e^{-\frac{1}{2}\|y\|^2}dy = \prod_{i=1}^{k}\int_{-\infty}^{\infty} e^{-\frac{1}{2}y_i^2}dy_i = (2\pi)^{\frac{k}{2}}.$$

Therefore:

$$\int_{\mathbb{R}^k} f(x)dx = \frac{(2\pi)^{\frac{k}{2}}}{(2\pi)^{\frac{k}{2}}} = 1.$$

2.

(a)

Write the trace explicitly:

$$\operatorname{trace}(AB) = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} B_{ji}.$$

Then, for each entry A_{pq},

$$\frac{\partial}{\partial A_{pq}} \operatorname{trace}(AB) = B_{qp}.$$

Hence, the matrix of all partial derivatives satisfies:

$$\frac{\partial}{\partial A} \operatorname{trace}(AB) = B^T.$$

(b)

Using the cyclic invariance of the trace ($\operatorname{trace}(XYZ) = \operatorname{trace}(ZXY)$):

$$\operatorname{trace}(xx^T A) = \operatorname{trace}(Axx^T) = \sum_{i,j} A_{ij}\, x_j x_i = x^T A x.$$

(c)

Let $x_1, \dots, x_n \in \mathbb{R}^k$ be i.i.d. with density

$$f(x;\mu,\Sigma) = \frac{1}{(2\pi)^{\frac{k}{2}}|\Sigma|^{\frac{1}{2}}} e^{\left[-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right]},$$

where $\mu \in \mathbb{R}^{k}$ and $\Sigma \in \mathbb{R}^{k\times k}$ is symmetric positive definite ($\Sigma \succ 0$). The log-likelihood is

$$l(\mu,\Sigma) = \sum_{i=1}^{n} \log f(x_i;\mu,\Sigma) = -\frac{nk}{2}\log|\Sigma| - \frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu).$$

For convenience define $S_i \equiv (x_i-\mu)(x_i-\mu)^T$ and $S \equiv \sum_{i=1}^{n} S_i$. We will use trace identities:

$$(x-\mu)^T\Sigma^{-1}(x-\mu) = \mathrm{trace}(\Sigma^{-1}(x_i-\mu)(x_i-\mu)^T) = \mathrm{trace}(\Sigma^{-1}S_i).$$

Write the μ-dependent part:

$$l(\mu,\Sigma) = \text{const w.r.t.}\,\mu \quad -\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^T\Sigma^{-1}(x_i-\mu).$$

Differentiate (vector derivative). Use the identity for symmetric A:

$$\frac{\partial}{\partial\mu}(x_i-\mu)^T A(x_i-\mu) = -2A(x_i-\mu).$$

Since Σ^{-1} is symmetric,

$$\frac{\partial l}{\partial\mu} = -\frac{1}{2}\sum_{i=1}^{n}\left(-2\Sigma^{-1}(x_i-\mu)\right) = \Sigma^{-1}\sum_{i=1}^{n}(x_i-\mu).$$

Set $\frac{\partial l}{\partial \mu} = 0$. Because Σ^{-1} is invertible,

$$\sum_{i=1}^{n} (x_i - \mu) = 0 \quad \Longrightarrow \quad \hat{\mu} = \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

Check: second derivative (Hessian) with respect to μ is

$$\frac{\partial^2 l}{\partial \mu \partial \mu^T} = -n\Sigma^{-1},$$

which is negative definite because $\Sigma^{-1} \succ 0$. Thus the critical point $\hat{\mu}$ is a (strict) local maximum in μ, and because the objective is concave in μ it is the global maximizer for fixed Σ.

We now maximize with respect to $\Sigma \succ 0$. Use the log-likelihood written in trace form:

$$l(\mu, \Sigma) = -\frac{n}{2} \log|\Sigma| - \frac{1}{2} \sum_{i=1}^{n} \operatorname{trace}(\Sigma^{-1} S_i) + \text{const}.$$

(“const” depends on data but not on Σ.)

We use standard matrix derivative facts (for Σ symmetric and $\Sigma \succ 0$):

$$\frac{\partial}{\partial \Sigma} \log|\Sigma| = (\Sigma^{-1})^T = \Sigma^{-1}, \qquad \frac{\partial}{\partial \Sigma} \operatorname{trace}(\Sigma^{-1} S) = -(\Sigma^{-1} S \Sigma^{-1})^T = -\Sigma^{-1} S \Sigma^{-1}.$$

(Transposes drop because Σ^{-1} and S are symmetric when appropriate; the formulas above are general.)

Differentiate:

$$\frac{\partial l}{\partial \Sigma} = -\frac{n}{2}\Sigma^{-1} + \frac{1}{2}\sum_{i=1}^{n} \Sigma^{-1} S_i \Sigma^{-1}.$$

Set equal to zero:

$$-\frac{n}{2}\Sigma^{-1} + \frac{1}{2}\sum_{i=1}^{n} \Sigma^{-1} S_i \Sigma^{-1} = 0.$$

Multiply on the left and right by Σ (which is invertible):

$$-nI + \sum_{i=1}^{n} S_i \Sigma^{-1} = 0.$$

Right-multiply by Σ or rearrange to

$$\sum_{i=1}^{n} S_i = n\Sigma.$$

Therefore the stationary point for Σ is

$$\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} (x_i - \hat{u})(x_i - \hat{u})^T.$$

(We substituted $u = \hat{u}$ because $\hat{u}$ was found earlier.)

We must argue the stationary point is actually the global maximum on

domain $\Sigma \succ 0$.

Two standard justifications:

1. Concavity in $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma^{-1}}$. The log-likelihood $l(\mu, \Sigma)$ is a concave function of μ (quadratic, negative definite Hessian) and is concave in the precision matrix $\Lambda = \Sigma^{-1}$ (linear in $\log\det$ term becomes $\log\det(\Lambda^{-1}) = -\log\det\Lambda$ plus linear term $\mathrm{trace}(\Lambda S)$; more directly, l is concave in Λ). Therefore the stationary point in the domain $\Sigma \succ 0$ is the unique global maximum.

2. Second-variation in $\boldsymbol{\Sigma}$. One can compute the second variation of l in directions H with small symmetric H and show it is negative definite at the solution; equivalently, the Hessian operator of l on the space of symmetric matrices is negative definite at the stationary point. This confirms a (strict) local maximum; combined with concavity argument above yields global maximality.

Hence, $\hat{\mu}, \hat{\Sigma}$ above are the MLEs.

- The MLE for the covariance uses factor $1/n$. The unbiased sample covariance uses $1/(n-1)$; $\frac{1}{n-1}\sum(x_i - \bar{x})(x_i - \bar{x})^T$, but that is **not** the MLE.
- Domain: the MLE $\hat{\Sigma}$ is positive semidefinite; with nondegenerate data it is positive definite. If data are degenerate (e.g., points lie in a lower-dimensional

affine subspace), $\hat{\Sigma}$ may be singular and the MLE over $\Sigma \succ 0$ does not exist (or is on the boundary).

- As a final boxed summary:

$$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad \hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^T$$