Suppose that vector random variable \( \mathbf{X} = (X_1, X_2, \ldots, X_k)^T \) has a multivariate normal distribution with pdf given by

\[
f_{\mathbf{X}}(\mathbf{x}) = \left( \frac{1}{2\pi} \right)^{k/2} \frac{1}{|\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2} \mathbf{x}^T \Sigma^{-1} \mathbf{x} \right\}
\]

where \( \Sigma \) is the \( k \times k \) variance-covariance matrix (we can consider here the case where the expected value \( \mu \) is the \( k \times 1 \) zero vector; results for the general case are easily available by transformation).

Consider partitioning \( \mathbf{X} \) into two components \( \mathbf{X}_1 \) and \( \mathbf{X}_2 \) of dimensions \( d \) and \( k - d \) respectively, that is,

\[
\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}.
\]

We attempt to deduce

(a) the marginal distribution of \( \mathbf{X}_1 \), and

(b) the conditional distribution of \( \mathbf{X}_2 \) given that \( \mathbf{X}_1 = \mathbf{x}_1 \).

First, write

\[
\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}
\]

where \( \Sigma_{11} \) is \( d \times d \), \( \Sigma_{22} \) is \( (k - d) \times (k - d) \), \( \Sigma_{21} = \Sigma_{12}^T \), and

\[
\Sigma^{-1} = V = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}
\]

so that \( \Sigma V = I_k \) (\( I_r \) is the \( r \times r \) identity matrix) gives

\[
\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix} = \begin{bmatrix} I_d & 0 \\ 0 & I_{k-d} \end{bmatrix}
\]

and more specifically the four relations

\[
\begin{align*}
\Sigma_{11} V_{11} + \Sigma_{12} V_{21} &= I_d \quad (2) \\
\Sigma_{11} V_{12} + \Sigma_{12} V_{22} &= 0 \quad (3) \\
\Sigma_{21} V_{11} + \Sigma_{22} V_{21} &= 0 \quad (4) \\
\Sigma_{21} V_{12} + \Sigma_{22} V_{22} &= I_{k-d} \quad (5)
\end{align*}
\]

From the multivariate normal pdf in equation (1), we can re-express the term in the exponent as

\[
\mathbf{x}^T \Sigma^{-1} \mathbf{x} = \mathbf{x}_1^T V_{11} \mathbf{x}_1 + \mathbf{x}_1^T V_{12} \mathbf{x}_2 + \mathbf{x}_2^T V_{21} \mathbf{x}_1 + \mathbf{x}_2^T V_{22} \mathbf{x}_2.
\]

In order to compute the marginal and conditional distributions, we must complete the square in \( \mathbf{x}_2 \) in this expression. We can write

\[
\mathbf{x}^T \Sigma^{-1} \mathbf{x} = (\mathbf{x}_2 - \mu)^T M (\mathbf{x}_2 - \mu) + c
\]

and by comparing with equation (6) we can deduce that, for quadratic terms in \( \mathbf{x}_2 \),

\[
\mathbf{x}_2^T V_{22} \mathbf{x}_2 = \mathbf{x}_2^T M \mathbf{x}_2 \quad : \quad M = V_{22}
\]
for linear terms
\[ \varepsilon_2^T V_{21} \varepsilon_1 = -\varepsilon_2^T M m \quad \therefore \quad m = -V_{22}^{-1} V_{21} \varepsilon_1 \] (9)
and for constant terms
\[ \varepsilon_1^T V_{11} \varepsilon_1 = c + m^T M m \quad \therefore \quad c = \varepsilon_1^T (V_{11} - V_{21} V_{22}^{-1} V_{21}) \varepsilon_1 \] (10)
thus yielding all the terms required for equation (7), that is
\[ \varepsilon^T \Sigma^{-1} \varepsilon = (\varepsilon_2 + V_{22}^{-1} V_{21} \varepsilon_1)^T V_{22} (\varepsilon_2 + V_{22}^{-1} V_{21} \varepsilon_1) + \varepsilon_1^T (V_{11} - V_{21} V_{22}^{-1} V_{21}) \varepsilon_1, \] (11)
which, crucially, is a sum of two terms, where the first can be interpreted as a function of \( \varepsilon_2 \), given \( \varepsilon_1 \), and the second is a function of \( \varepsilon_1 \) only.

Hence we have an immediate factorization of the full joint pdf using the chain rule for random variables;
\[ f_{X}(\varepsilon) = f_{X_2|X_1}(\varepsilon_2|\varepsilon_1) f_{X_1}(\varepsilon_1) \] (12)
where
\[ f_{X_2|X_1}(\varepsilon_2|\varepsilon_1) \propto \exp \left\{ -\frac{1}{2} (\varepsilon_2 + V_{22}^{-1} V_{21} \varepsilon_1)^T V_{22} (\varepsilon_2 + V_{22}^{-1} V_{21} \varepsilon_1) \right\} \] (13)
giving that
\[ X_2 | X_1 = \varepsilon_1 \sim N \left( -V_{22}^{-1} V_{21} \varepsilon_1, V_{22}^{-1} \right) \] (14)
and
\[ f_{X_1}(\varepsilon_1) \propto \exp \left\{ -\frac{1}{2} \varepsilon_1^T (V_{11} - V_{21} V_{22}^{-1} V_{21}) \varepsilon_1 \right\} \] (15)
giving that
\[ X_1 \sim N \left( 0, (V_{11} - V_{21} V_{22}^{-1} V_{21})^{-1} \right). \] (16)
But, from equation (3), \( \Sigma_{12} = -\Sigma_{11} V_{12} V_{22}^{-1} \), and then from equation (2), substituting in \( \Sigma_{12} \),
\[ \Sigma_{11} V_{11} - \Sigma_{11} V_{12} V_{22}^{-1} V_{21} = I_d \quad \therefore \quad \Sigma_{11} = (V_{11} - V_{12} V_{22}^{-1} V_{21})^{-1} = (V_{11} - V_{21} V_{22}^{-1} V_{21})^{-1}. \]
Hence, by inspection of equation (16), we conclude that
\[ X_1 \sim N \left( 0, \Sigma_{11} \right). \] (17)
that is, we can extract the \( \Sigma_{11} \) block of \( \Sigma \) to define the marginal variance-covariance matrix of \( X_1 \).

Using similar arguments, we can define the conditional distribution from equation (14) more precisely.
First, from equation (3), \( V_{12} = -\Sigma_{11}^{-1} \Sigma_{12} V_{22} \), and then from equation (5), substituting in \( V_{12} \)
\[ -\Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} V_{22} + \Sigma_{22} V_{22} = I_{k-d} \quad \therefore \quad V_{22}^{-1} = \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} = \Sigma_{22} - \Sigma_{12} \Sigma_{11}^{-1} \Sigma_{12}. \]
Finally, from equation (3), taking transposes on both sides, we have that \( V_{21} \Sigma_{11} + V_{22} \Sigma_{21} = 0 \). Then pre-multiplying by \( V_{22}^{-1} \), and post-multiplying by \( \Sigma_{11}^{-1} \), we have
\[ V_{22}^{-1} V_{21} + \Sigma_{21} \Sigma_{11}^{-1} = 0 \quad \therefore \quad V_{22}^{-1} V_{21} = -\Sigma_{21} \Sigma_{11}^{-1}, \]
so we have, substituting into equation (14), that
\[ X_2 | X_1 = \varepsilon_1 \sim N \left( \Sigma_{21} \Sigma_{11}^{-1} \varepsilon_1, \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \right). \] (18)
Thus any marginal, and any conditional distribution of a multivariate normal joint distribution is also multivariate normal, as the choices of \( X_1 \) and \( X_2 \) are arbitrary.