Matrix Calculus

Lecture Notes

Ahmet Ademoglu, *PhD*
Bogazici University
Institute of Biomedical Engineering

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \ldots & \mathbf{a}_n \\ \downarrow & \downarrow & & \downarrow \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1r} \\ b_{21} & b_{22} & \cdots & b_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nr} \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \ldots & \mathbf{b}_r \\ \downarrow & \downarrow & & \downarrow \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{b}}_1 \longrightarrow \\ \hat{\mathbf{b}}_2 \longrightarrow \\ \vdots \\ \hat{\mathbf{b}}_n \longrightarrow \end{bmatrix}$$

Product of Matrices

$$\mathbf{AB} = \underset{\downarrow}{\mathbf{a}_1}\hat{\mathbf{b}}_1 \longrightarrow \; + \; \underset{\downarrow}{\mathbf{a}_2}\hat{\mathbf{b}}_2 \longrightarrow \; + \ldots + \; \underset{\downarrow}{\mathbf{a}_n}\hat{\mathbf{b}}_n \longrightarrow = \sum_{i=1}^{n} \mathbf{a}_i \hat{\mathbf{b}}_i$$

$$\{\mathbf{AB}\}_{ij} = \sum_k a_{ik} b_{kj}$$

$$\mathbf{a}^T\mathbf{B}\mathbf{c} = \sum_i \sum_j a_i b_{ij} c_j$$

$$\mathbf{ABC} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \dots \\ \downarrow & \downarrow & \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & \dots \\ b_{21} & b_{22} & \dots \\ \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} \mathbf{c}_1 \longrightarrow \\ \mathbf{c}_2 \longrightarrow \\ \vdots \end{bmatrix}$$

$$= \underset{\downarrow}{b_{11}\mathbf{a}_1}\mathbf{c}_1 \longrightarrow \quad + \quad \underset{\downarrow}{b_{12}\mathbf{a}_1}\mathbf{c}_2 \longrightarrow \quad + \quad \underset{\downarrow}{b_{21}\mathbf{a}_2}\mathbf{c}_1 \longrightarrow \quad + \quad \dots$$

Eigenvalue Problem : $\mathbf{Av} = \lambda \mathbf{v}$

$$\mathbf{A}\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_N \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_N \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_N \end{bmatrix}$$

$$\mathbf{AV} = \mathbf{V\Lambda} \longrightarrow \mathbf{A} = \mathbf{V\Lambda V}^{-1}$$

$$f(\mathbf{A}) = \mathbf{V}f(\mathbf{\Lambda})\mathbf{V}^{-1} \qquad e^{\mathbf{A}} = \mathbf{I} + \frac{1}{1!}\mathbf{A} + \frac{1}{2!}\mathbf{A}^2 + \cdots$$

Characteristic polynomial

$$|\lambda \mathbf{I} - \mathbf{A}| = \lambda^N + a_1\lambda^{N-1} + \cdots + a_{N-1}\lambda + a_N = 0$$

Every matrix satisfies its characteristic equation

$$\mathbf{A}^N + a_1\mathbf{A}^{N-1} + \cdots + a_{N-1}\mathbf{A} + a_N = \mathbf{0}$$

TRACE of $\mathbf{A}$

$$Trace(\mathbf{A}) = \sum_i a_{ii}$$

Useful Identities

- $Trace(\mathbf{AB}) = Trace(\mathbf{BA})$
- $Trace(\mathbf{ABC}) = Trace(\mathbf{CAB}) = Trace(\mathbf{BCA})$
- $\mathbf{x}^T\mathbf{Ay} = Trace(\mathbf{yx}^T\mathbf{A})$
- $Trace(\mathbf{A}) = \sum_i \lambda_i$

Singular Value Decomposition

Every matrix $\mathbf{A}$ can be decomposed into

$\mathbf{A}_{m\times n} = \mathbf{U}_{m\times m}\boldsymbol{\Sigma}_{m\times n}\mathbf{V}^T_{n\times n}$

$eig(\mathbf{A}\mathbf{A}^T) = \mathbf{U}\mathbf{D}\mathbf{U}^T \longrightarrow \mathbf{U}$: left eigenvectors

$\mathbf{u}_i^T\mathbf{u}_j = \delta_{ij} \longrightarrow \mathbf{U}^{-1} = \mathbf{U}^T$

$eig(\mathbf{A}^T\mathbf{A}) = \mathbf{V}\mathbf{D}\mathbf{V}^T \longrightarrow \mathbf{V}$: right eigenvectors

$\mathbf{v}_i^T\mathbf{v}_j = \delta_{ij} \longrightarrow \mathbf{V}^{-1} = \mathbf{V}^T$

$\boldsymbol{\Sigma}$: singular values $\longrightarrow \boldsymbol{\Sigma}^2 = \mathbf{D}$

$$\mathbf{A} = \begin{bmatrix} \mathbf{u}_1 & \dots & \mathbf{u}_r & \mathbf{u}_{r+1} & \dots & \mathbf{u}_m \\ \downarrow & \vdots & \downarrow & \downarrow & \vdots & \downarrow \end{bmatrix} \begin{bmatrix} \sigma_1 & & & \mathbf{0} \\ & \ddots & & \\ & & \sigma_r & \\ \mathbf{0} & & & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T & \longrightarrow \\ \vdots & \\ \mathbf{v}_r^T & \longrightarrow \\ \mathbf{v}_{r+1}^T & \longrightarrow \\ \vdots & \\ \mathbf{v_n}^T & \longrightarrow \end{bmatrix}$$

$\underbrace{Col(A)}$ $\underbrace{Null\ Space}$

Special Matrix Products

Kronecker Product of **A** and **B**

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1N}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{N1}\mathbf{B} & \cdots & a_{NN}\mathbf{B} \end{bmatrix}$$

Hadamard Product of **A** and **B**

$$\mathbf{A} \odot \mathbf{B} = \begin{bmatrix} a_{11}b_{11} & \cdots & a_{1N}b_{1N} \\ \vdots & \ddots & \vdots \\ a_{N1}b_{N1} & \cdots & a_{NN}a_{NN} \end{bmatrix}$$

Khatri-Rao Product of **A** and **B**

$$\mathbf{A} * \mathbf{B} = \begin{bmatrix} A(:,1) \otimes B(:,1) & A(:,2) \otimes B(:,2) \ldots & A(:,N) \otimes B(:,N) \end{bmatrix}$$

Vectorizing **A**

$$vec(\mathbf{A}) = \begin{bmatrix} A(:,1)^T & A(:,2)^T & \ldots & A(:,N)^T \end{bmatrix}^T$$

Kronecker Properties

- $(\mathbf{V} \otimes \mathbf{C})^{-1} = \mathbf{V}^{-1} \otimes \mathbf{C}^{-1}$
- $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$
- $(\mathbf{A} \otimes \mathbf{B})^T = (\mathbf{A}^T \otimes \mathbf{B}^T)$
- $Tr(\mathbf{A}^T\mathbf{B}) = vec(\mathbf{A})^T vec(\mathbf{B})$
- $vec(\mathbf{Y})^T(\mathbf{V} \otimes \mathbf{C})^{-1} vec(\mathbf{Y}) = vec(\mathbf{Y})^T \mathbf{V}^{-1} \otimes \mathbf{C}^{-1} vec(\mathbf{Y}) = Tr[\mathbf{C}^{-1}\mathbf{Y}^T\mathbf{V}^{-1}\mathbf{Y}]$
- $|\mathbf{V} \otimes \mathbf{C}| = |\mathbf{V}|^{rank(\mathbf{C})}|\mathbf{C}|^{rank(\mathbf{V})}$
- $vec(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})vec(\mathbf{B})$
- $Tr[\mathbf{A} \otimes \mathbf{B}] = Tr[\mathbf{A}]\,Tr[\mathbf{B}]$

Definitions

If $f(\mathbf{x})$ is a scalar; $\frac{\partial}{\partial \mathbf{x}} f(\mathbf{x}) = \begin{bmatrix} \frac{\partial}{\partial x_1} f(\mathbf{x}) \\ \frac{\partial}{\partial x_2} f(\mathbf{x}) \\ \vdots \\ \frac{\partial}{\partial x_n} f(\mathbf{x}) \end{bmatrix}$

If $f(\mathbf{x}) = [f_1(\mathbf{x})\ f_2(\mathbf{x}) \cdots f_m(\mathbf{x})]$ then

$$\frac{\partial}{\partial \mathbf{x}} f(\mathbf{x}) = \begin{bmatrix} \frac{\partial}{\partial x_1} f(\mathbf{x}) \\ \frac{\partial}{\partial x_2} f(\mathbf{x}) \\ \vdots \\ \frac{\partial}{\partial x_n} f(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial x_1} f_1(\mathbf{x}) & \frac{\partial}{\partial x_1} f_2(\mathbf{x}) & \cdots & \frac{\partial}{\partial x_1} f_m(\mathbf{x}) \\ \frac{\partial}{\partial x_2} f_1(\mathbf{x}) & \frac{\partial}{\partial x_2} f_2(\mathbf{x}) & & \frac{\partial}{\partial x_2} f_m(\mathbf{x}) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_n} f_1(\mathbf{x}) & \frac{\partial}{\partial x_n} f_2(\mathbf{x}) & \cdots & \frac{\partial}{\partial x_n} f_m(\mathbf{x}) \end{bmatrix}$$

Taylor Expansion of Multivariate Functions

$$f(\mathbf{x}+\delta\mathbf{x}) = f(\mathbf{x}) + \frac{1}{1!}\sum_{j=1}^{N}\frac{\partial f(\mathbf{x})}{\partial x_j}\delta x_j + \frac{1}{2!}\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{\partial^2 f(\mathbf{x})}{\partial x_j \partial x_j}\delta x_i \delta x_j + \frac{1}{3!}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\frac{\partial^3 f(\mathbf{x})}{\partial x_j \partial x_j \partial x_k}\delta x_i \delta x_j \delta x_k + \cdots$$

$$f(\mathbf{x}+\delta\mathbf{x}) = f(\mathbf{x}) + \mathbf{g}^T\delta\mathbf{x} + \frac{1}{2!}\delta\mathbf{x}^T H \delta\mathbf{x} + \cdots$$

$$\mathbf{g} = \frac{df(\mathbf{x})}{d\mathbf{x}} = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \frac{\partial f(\mathbf{x})}{\partial x_2} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_N} \end{bmatrix}, \quad \mathbf{H} = \frac{d\mathbf{g}}{d\mathbf{x}} \begin{bmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_N} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_N \partial x_1} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_N^2} \end{bmatrix}$$

$\mathbf{g}$: Jacobian $\qquad$ $\mathbf{H}$: Hessian

Matrix Derivatives

- $\frac{\partial \mathbf{u}^T\mathbf{A}\mathbf{v}}{\partial \mathbf{x}} = \frac{\partial \mathbf{u}}{\partial \mathbf{x}}\mathbf{A}\mathbf{v} + \frac{\partial \mathbf{v}}{\partial \mathbf{x}}\mathbf{A}^T\mathbf{u}$
  $\rightarrow \frac{\partial}{\partial \mathbf{x}}\mathbf{A}\mathbf{x} = \mathbf{A}^T$ and $\frac{\partial}{\partial \mathbf{x}}\mathbf{x}^T\mathbf{A} = \mathbf{A}$ and $\frac{\partial}{\partial \mathbf{x}}\mathbf{x}^T\mathbf{A}\mathbf{x} = (\mathbf{A}+\mathbf{A}^T)\mathbf{x}$
- $\frac{\partial}{\partial \mathbf{x}} tr(\mathbf{Y}) = tr(\frac{\partial}{\partial \mathbf{x}}\mathbf{Y})$
- $\frac{\partial}{\partial x}\mathbf{Y}^{-1} = -\mathbf{Y}^{-1}\left(\frac{\partial}{\partial x}\mathbf{Y}\right)\mathbf{Y}^{-1}$
- $\frac{\partial}{\partial \mathbf{X}} tr(\mathbf{Y}) = tr(\frac{\partial}{\partial \mathbf{X}}\mathbf{Y})$ if $\mathbf{Y} = \mathbf{X} \longrightarrow \frac{\partial}{\partial \mathbf{X}} tr(\mathbf{X}) = \mathbf{I}$
- $\frac{\partial}{\partial \mathbf{X}} tr(\mathbf{A}\mathbf{X}) = \mathbf{A}^T$
- $\frac{\partial}{\partial \mathbf{X}} tr(\mathbf{X}^\mathbf{T}\mathbf{X}) = 2\mathbf{X}^\mathbf{T}$
- $\frac{\partial}{\partial \mathbf{X}}|\mathbf{X}| = |\mathbf{X}|\mathbf{X}^{-T}$
- $\frac{\partial}{\partial \mathbf{X}}\log|\mathbf{X}| = \mathbf{X}^{-T}$
- $\frac{\partial}{\partial \mathbf{X}} tr(\mathbf{A}\mathbf{X}^{-1}\mathbf{B}) = -\mathbf{X}^{-T}\mathbf{A}^T\mathbf{B}^T\mathbf{X}^{-T}$

$\frac{\partial \mathbf{u}^T \mathbf{A} \mathbf{v}}{\partial \mathbf{x}} = \frac{\partial \mathbf{u}}{\partial \mathbf{x}} \mathbf{A} \mathbf{v} + \frac{\partial \mathbf{v}}{\partial \mathbf{x}} \mathbf{A}^T \mathbf{u}$

$$\frac{\partial \mathbf{u}^T \mathbf{A} \mathbf{v}}{\partial \mathbf{x}} \longrightarrow \frac{\partial}{\partial x_m} \sum_i \sum_j u_i A_{ij} v_j = \sum_i \sum_j \frac{\partial u_i}{\partial x_m} A_{ij} v_j + \sum_i \sum_j u_i A_{ij} \frac{\partial v_j}{\partial x_m}$$

$$= \sum_i \frac{\partial u_i}{\partial x_m} \sum_j A_{ij} v_j + \sum_j \frac{\partial v_j}{\partial x_m} \sum_i A_{ij} u_i$$

$$\longrightarrow \frac{\partial \mathbf{u}}{\partial \mathbf{x}} \mathbf{A} \mathbf{v} + \frac{\partial \mathbf{v}}{\partial \mathbf{x}} \mathbf{A}^T \mathbf{u}$$

$\frac{\partial Tr(\mathbf{X})}{\partial \mathbf{X}} = \mathbf{I}$

$$\frac{\partial Tr(\mathbf{X})}{\partial \mathbf{X}} = \mathbf{I} \longrightarrow \sum_i \frac{\partial x_{ii}}{\partial x_{mn}}$$

If $m = n = i$ then $\frac{\partial x_{ii}}{\partial x_{mn}} = \frac{\partial x_{ii}}{\partial x_{ii}} = 1$, otherwise $\frac{\partial x_{ii}}{\partial x_{mn}} = 0$

$$\sum_i \frac{\partial x_{ii}}{\partial x_{mn}} \longrightarrow \mathbf{I} = \frac{\partial Tr(\mathbf{X})}{\partial \mathbf{X}}$$

$$\frac{\partial Tr(\mathbf{AX}^{-1}\mathbf{B})}{\partial \mathbf{X}} = -\mathbf{X}^{-T}\mathbf{A}^T\mathbf{B}^T\mathbf{X}^{-T}$$

$$\sum_{i,p,q} a_{ip}\{\mathbf{X}^{-1}\}_{pq} b_{qi} = -\sum_{i,p,q} a_{ip} \sum_{r,s} \{\mathbf{X}^{-1}\}_{pr} x_{rs} \{\mathbf{X}^{-1}\}_{sq} b_{qi}$$

$$\frac{\partial Tr(\mathbf{AX}^{-1}\mathbf{B})}{\partial \mathbf{X}} = -\sum_{i,p,q} a_{ip} \sum_{r,s} \{\mathbf{X}^{-1}\}_{pr} \frac{\partial x_{rs}}{\partial x_{mn}} \{\mathbf{X}^{-1}\}_{sq} b_{qi} =$$

$$-\sum_{i,p,q} a_{ip}\{\mathbf{X}^{-1}\}_{pm}\{\mathbf{X}^{-1}\}_{nq} b_{qi}$$

$$= -\sum_{i}\{\mathbf{AX}^{-1}\}_{im}\{\mathbf{X}^{-1}\mathbf{B}\}_{ni} = -\sum_{i}\{\mathbf{X}^{-T}\mathbf{A}^T\}_{mi}\{\mathbf{BX}^{-T}\}_{in} =$$

$$-\mathbf{X}^{-T}\mathbf{A}^T\mathbf{B}^T\mathbf{X}^{-T}$$

Inverse and Determinant

$$\mathbf{X}^{-1} = \frac{1}{|\mathbf{X}|} adj(\mathbf{X})$$
$$adj(\mathbf{X})^T = cofactor(\mathbf{X})$$
$$C_{ij} : cofactor(\mathbf{X})_{ij}$$
$$C_{ij} = (-1)^{i+j} M_{ij}$$
$$M_{ij} : Minor(\mathbf{X})_{ij}$$
$$M_{ij} = |X| \text{ with } i\text{th row and } j\text{th column deleted}$$
$$|X| = X_{i1}C_{i1} + X_{i2}C_{i2} + \cdots + X_{in}C_{in} = \sum_{j=1}^{n} X_{ij}C_{ij}$$
$$\left[\frac{\partial|\mathbf{X}|}{\partial\mathbf{X}}\right]_{kl} = \frac{\partial}{\partial X_{kl}}\{X_{i1}C_{i1} + X_{i2}C_{i2} + \cdots X_{in}C_{in}\}, \text{ choose } i = k$$
$$\frac{\partial}{\partial X_{kl}}\{X_{k1}C_{k1} + X_{k2}C_{k2} + \cdots X_{kn}C_{kn}\} = C_{kl} = cofactor(\mathbf{X})_{kl}$$
$$\frac{\partial|\mathbf{X}|}{\partial\mathbf{X}} = |\mathbf{X}|\mathbf{X}^{-1}$$

MATRIX INVERSION LEMMA

Some Useful Identities

$$(\mathbf{I}+\mathbf{P})^{-1} = (\mathbf{I}+\mathbf{P})^{-1}(\mathbf{I}+\mathbf{P}-\mathbf{P}) = \mathbf{I}-(\mathbf{I}+\mathbf{P})^{-1}\mathbf{P} \qquad (1)$$

$$\mathbf{P}+\mathbf{PQP} = \mathbf{P}(\mathbf{I}+\mathbf{QP}) = (\mathbf{I}+\mathbf{PQ})\mathbf{P}$$

$$(\mathbf{I}+\mathbf{PQ})^{-1}\mathbf{P} = \mathbf{P}(\mathbf{I}+\mathbf{QP})^{-1} \qquad (2)$$

$$(\mathbf{A}+\mathbf{BCD})^{-1} = \left(\mathbf{A}\left[\mathbf{I}+\mathbf{A}^{-1}\mathbf{BCD}\right]\right)^{-1} = \left[\mathbf{I}+\mathbf{A}^{-1}\mathbf{BCD}\right]^{-1}\mathbf{A}^{-1}$$

calling $\mathbf{P}=\mathbf{A}^{-1}\mathbf{BCD}$ and using (1)

$$\left[\mathbf{I}-\left(\mathbf{I}+\mathbf{A}^{-1}\mathbf{BCD}\right)^{-1}\mathbf{A}^{-1}\mathbf{BCD}\right]\mathbf{A}^{-1} = \mathbf{A}^{-1}-(\mathbf{I}+\mathbf{A}^{-1}\mathbf{BCD})^{-1}\mathbf{A}^{-1}\mathbf{BCDA}^{-1}$$

calling $\mathbf{P}=\mathbf{A}^{-1}$ and $\mathbf{Q}=\mathbf{BCD}$ and using (2)

$$(\mathbf{A}+\mathbf{BCD})^{-1} = \mathbf{A}^{-1}-\mathbf{A}^{-1}(\mathbf{I}+\mathbf{BCDA}^{-1})^{-1}\mathbf{BCDA}^{-1}$$

Using (2) again $\quad (\mathbf{A}+\mathbf{BCD})^{-1} = \mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{B}(\mathbf{I}+\mathbf{CDA}^{-1}\mathbf{B})^{-1}\mathbf{CDA}^{-1}$

or $\quad (\mathbf{A}+\mathbf{BCD})^{-1} = \mathbf{A}^{-1}-\mathbf{A}^{-1}\mathbf{BC}(\mathbf{I}+\mathbf{DA}^{-1}\mathbf{BC})^{-1}\mathbf{DA}^{-1}$

Woodbury Identity

In a special case where $\mathbf{B} \to \mathbf{u}$, $\mathbf{C} = 1$, $\mathbf{D} = \mathbf{v}^H$
the Matrix Inversion Lemma leads to

$$(\mathbf{A} + \mathbf{u}\mathbf{v}^H)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^H\mathbf{A}^{-1}}{1+\mathbf{v}^H\mathbf{A}^{-1}\mathbf{u}}$$

Efficient way to compute the determinant

Suppose $\mathbf{A}$ is an invertible n-by-n matrix and $\mathbf{U}$, $\mathbf{V}$ are n-by-m matrices. Then
$|\mathbf{A} + \mathbf{U}\mathbf{V}^T| = |\mathbf{I}_m + \mathbf{V}^T\mathbf{A}^{-1}\mathbf{U}||\mathbf{A}|$
Given additionally an invertible m-by-m matrix $\mathbf{W}$, the relationship can also be expressed as
$|\mathbf{A} + \mathbf{U}\mathbf{W}\mathbf{V}^T| = |\mathbf{W}^{-1} + \mathbf{V}^T\mathbf{A}^{-1}\mathbf{U}||\mathbf{W}||\mathbf{A}|$

Inverse and Determinant with Submatrices

$$\begin{bmatrix} \mathbf{A} & \mathbf{C} \\ \mathbf{B} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{C}\mathbf{F_D}^{-1}\mathbf{B}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{C}\mathbf{F_D}^{-1} \\ -\mathbf{F_D}^{-1}\mathbf{B}\mathbf{A}^{-1} & \mathbf{F_D}^{-1} \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{F_A}^{-1} & -\mathbf{F_D}^{-1}\mathbf{C}\mathbf{D}^{-1} \\ -\mathbf{C}^{-1}\mathbf{D}\mathbf{F_A}^{-1} & \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{B}\mathbf{F_A}^{-1}\mathbf{C}\mathbf{D}^{-1} \end{bmatrix}$$

$$\mathbf{F_A} = [\mathbf{A} - \mathbf{C}\mathbf{D}^{-1}\mathbf{B}] \qquad \mathbf{F_D} = [\mathbf{D} - \mathbf{B}\mathbf{A}^{-1}\mathbf{C}]$$

where $\mathbf{F_D}$ and $\mathbf{F_D}$ are the Schur's complement of $\mathbf{A}$ and $\mathbf{D}$, respectively.
Determinant of a block diagonal matrix is

$$\begin{vmatrix} \mathbf{A} & \mathbf{C} \\ \mathbf{B} & \mathbf{D} \end{vmatrix} = \left| \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{B} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{A}^{-1}\mathbf{C} \\ \mathbf{0} & \mathbf{D} - \mathbf{B}\mathbf{A}^{-1}\mathbf{C} \end{bmatrix} \right| = |\mathbf{A}| \cdot |\mathbf{D} - \mathbf{B}\mathbf{A}^{-1}\mathbf{C}| = |\mathbf{A}| \cdot |\mathbf{F_D}|$$

$$\begin{vmatrix} \mathbf{A} & \mathbf{C} \\ \mathbf{B} & \mathbf{D} \end{vmatrix} = \left| \begin{bmatrix} \mathbf{I} & \mathbf{C} \\ \mathbf{0} & \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{A} - \mathbf{C}\mathbf{D}^{-1}\mathbf{B} & \mathbf{0} \\ \mathbf{D}^{-1}\mathbf{B} & \mathbf{I} \end{bmatrix} \right| = |\mathbf{D}| \cdot |\mathbf{A} - \mathbf{C}\mathbf{D}^{-1}\mathbf{B}| = |\mathbf{D}| \cdot |\mathbf{F_A}|$$

It follows by the assignment $\begin{bmatrix} \mathbf{A} & \mathbf{C} \\ \mathbf{B} & \mathbf{D} \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{X} \\ \mathbf{X}^T & -\mathbf{D}^{-1} \end{bmatrix}$ that

$$|\mathbf{A} + \mathbf{X}\mathbf{D}\mathbf{X}^T| = |\mathbf{A}| \cdot |\mathbf{D}| \cdot |\mathbf{D}^{-1} + \mathbf{X}^T\mathbf{A}^{-1}\mathbf{X}|$$

Norm of a matrix **A** or a vector **a**

- L_0 Norm : $||\mathbf{A}||_0 = \sum_{ij} |a_{ij}|^0$
- L_1 Norm : $||\mathbf{A}||_1 = \sum_{ij} |a_{ij}|$
- L_2 Norm of a vector **a** $\rightarrow ||\mathbf{a}||_2 = \sqrt{\sum_i a_i^2}$
- Frobenius Norm of a matrix **A** $\rightarrow ||\mathbf{A}||_F = \sqrt{\sum_{ij} a_{ij}^2}$
- L_∞ Norm : $||\mathbf{A}||_\infty = \max_{ij} |a_{ij}|$
- $L_{2,1}$ Norm of a matrix **A** $= \sum_i ||\mathbf{A}(i,:)||_2$
- $L_{1,2}$ Norm of a matrix **A** $= \sqrt{\sum_i ||\mathbf{A}(i,:)||_1^2}$
- Nuclear norm of a matrix **A** $||\mathbf{A}||_* = \sum_i \lambda_i \; \lambda_i = \{singular\ values\ of\ \mathbf{A}\}$

Solution of Linear System of Equations $\mathbf{Ax} = \mathbf{b}$

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m &= b_1 \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2m}x_m &= b_2 \\
\vdots \qquad\qquad & \vdots \quad \vdots \\
a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m &= b_n
\end{aligned}
$$

$$\mathbf{A}_{m \times n}\mathbf{x}_n = \mathbf{b}_m$$

- Least Squares Solution
- Minimum Norm Solution
- Regularized Least Squares Solution

Solution of Linear System of Equations $\mathbf{Ax} = \mathbf{b}$

- $m < n$ Overdetermined Case $\longrightarrow$ Least Squares Solution
  $\frac{\partial}{\partial \mathbf{x}} |\mathbf{Ax} - \mathbf{b}|^2 = 0,$
  $\frac{\partial}{\partial \mathbf{x}} \left(\mathbf{x}^T \mathbf{A}^T \mathbf{Ax} - \mathbf{x}^T \mathbf{A}^T \mathbf{b} - \mathbf{b}^T \mathbf{Ax} + \mathbf{b}^T \mathbf{b} \right) = 2\mathbf{A}^T \mathbf{Ax} - \mathbf{A}^T \mathbf{b} - \mathbf{A}^T \mathbf{b} = 0$
  $$\mathbf{x} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}$$
  If $\frac{\partial}{\partial \mathbf{x}} (\mathbf{Ax} - \mathbf{b})^T \mathbf{W} (\mathbf{Ax} - \mathbf{b}) = 0$ then it is Weighted Least Squares Solution
  $$\mathbf{x} = (\mathbf{A}^T \mathbf{WA})^{-1} \mathbf{A}^T \mathbf{Wb}$$
- $m > n$ Underdetermined Case $\longrightarrow$ Minimum Norm Solution
  $\frac{\partial}{\partial \mathbf{x}} \left(\mathbf{x}^T \mathbf{x} + \lambda^T (\mathbf{Ax} - \mathbf{b}) \right) = 2\mathbf{x} + \mathbf{A}^T \lambda = 0$ and $\lambda = -2(\mathbf{AA}^T)^{-1} \mathbf{b}$
  $$\mathbf{x} = \mathbf{A}^T (\mathbf{AA}^T)^{-1} \mathbf{b}$$
- $m > n \longrightarrow$ Regularized Least Squares Solution
  $\frac{\partial}{\partial \mathbf{x}} \left(|\mathbf{Ax} - \mathbf{b}|^2 + \mu \mathbf{x}^T \mathbf{x} \right) = 0$
  $$\mathbf{x}_\mu = (\mathbf{A}^T \mathbf{A} + \mu \mathbf{I})^{-1} \mathbf{A}^T \mathbf{b}$$
- $m > n \longrightarrow$ Smoothness Regularization
  $\frac{\partial}{\partial \mathbf{x}} \left(|\mathbf{Lx}|^2 + \lambda^T (\mathbf{Ax} - \mathbf{b}) \right) = 0$, $\mathbf{L}$: The Laplacian Operator
  $$\mathbf{x} = (\mathbf{L}^T \mathbf{L})^{-1} \mathbf{A}^T (\mathbf{A} (\mathbf{L}^T \mathbf{L})^{-1} \mathbf{A}^T)^{-1} \mathbf{b}$$
- $m > n \rightarrow$ General Minimum Norm Solution with Constraint $\mathbf{Cx} - \mathbf{d}$
  $\frac{\partial}{\partial \mathbf{x}} \left(|\mathbf{Ax} - \mathbf{b}|^2 + \lambda^T (\mathbf{Cx} - \mathbf{d}) \right) = 0 \longrightarrow \mathbf{x} = ?$

A 2D INVERSE PROBLEM WITH SMOOTHNESS CONSTRAINT

A 2−D image vectorized as $\mathbf{x}$ is transformed through a spatial transfer function $\mathbf{L}$ and observed as $\mathbf{v}$. The inverse problem to estimate $\mathbf{x}$ from $\mathbf{v} = \mathbf{L}\mathbf{x}$ is usually a heavily underdetermined system whose regularization requires choosing the smoothest possible solution by minimizing the cost function

$$\mathcal{L} = (\mathbf{v} - \mathbf{L}\mathbf{x})^T \mathbf{C}_e^{-1} (\mathbf{v} - \mathbf{L}\mathbf{x}) + \lambda \mathbf{x}^T \mathbf{M}^T \mathbf{M}\mathbf{x}$$

where $\mathbf{L}$ is the 2-D Laplacian operator minimizing the output of the spatial high-pass portion of the signal and the error $\mathbf{e} = \mathbf{v} - \mathbf{L}\mathbf{x}$ is multivariate normal *i.e.* $\mathcal{N}(0, \mathbf{C}_e)$

The optimal solution is given by

$$\mathbf{T} = (\mathbf{L}^T \mathbf{C}_e^{-1} \mathbf{L} + \lambda \mathbf{M}^T \mathbf{M})^{-1} \mathbf{L}^T \mathbf{C}_e^{-1}$$

or by using the *Matrix Inversion Lemma*

$$\mathbf{T} = (\mathbf{M}^T \mathbf{M})^{-1} \mathbf{L}^T (\mathbf{L}(\mathbf{M}^T \mathbf{M})^{-1} \mathbf{L}^T + \lambda \mathbf{C}_e)^{-1}$$

(*See the following MATLAB code for its implementation*).

```
% SIMULATION OF A 2D INVERSE PROBLEM WITH SMOOTHNESS + DEPTH CONSTRAINT
DIM = [32 32 1]; S = 40; % number of sensors
Gain=1; X = zeros(DIM(1),DIM(2));
RANGE = round ( [DIM(1)/6+[1:5]   DIM(1)/1.5+[1:6] ] );
X( RANGE,RANGE) = 1; % create an image
X = spm_conv(X,3,3); % smooth the image
L = Gain*randn(S,DIM(1)*DIM(2)); % Generate Lead Field Matrix
V = L*X(:); % Calculate the forward problem to obtain the sensor data
Vn = V + randn(S,1); % V + E : Add noise to sensor data
C_e = eye(S,S); % Noise covariance
% Inversion : V = LX + E,  Minimize { (V - L*X)^T*C_e_inv*(V - L*X)^T + lambda*X^T*HTH*X }
% HTH : a priori constraint, C_e_inv : Inverse of the noise covariance matrix E distributed as N(0,C_e)
% T = inv(L'*C_e_inv*L + lambda*HTH)*L'*C_e_inv
% HTH : Laplacian Operator
% Calculate the Adjacency matrix of the 2D data
mat = reshape([1:DIM(1)*DIM(2)],DIM(1),DIM(2));              % neigboring graph
[r,c] = size(mat);                        %# Get the matrix size
diagVec1 = repmat([ones(c-1,1); 0],r,1);  %# Make the first diagonal vector
%#   (for horizontal connections)
diagVec1 = diagVec1(1:end-1);             %# Remove the last value
diagVec2 = ones(c*(r-1),1);               %# Make the second diagonal vector
%#   (for vertical connections)
adj = diag(diagVec1,1)+...                %# Add the diagonals to a zero matrix
diag(diagVec2,c);
adj = adj+adj.' ;                         %'# Add the matrix to a transposed
M = (adj - 4*eye(size(adj)));  % Laplace operator
W = eye(size(M)); % no depth weighting %W=  diag((diag(L'*L).^( 0.5)));   % Depth Weighting
MW = M*W; % combined prior : Laplace + Depth Weighting
HTH_inv = pinv (MW'*MW);  lambda = 1.5;
T = HTH_inv*L'*inv(L*HTH_inv*L' + lambda * C_e); % Using Matrix Inversion Lemma
X_e = reshape(T*Vn,DIM(1),DIM(2)); X_ls = reshape (pinv(L)*Vn,DIM(1),DIM(2));
subplot(3,1,1); imagesc(X);subplot(3,1,2); imagesc(X_e); subplot(3,1,3); imagesc(X_ls);
```

Questions for Self Study

1. Show the following relations:
   - $\mathbf{a}^T\mathbf{b} = tr(\mathbf{b}\mathbf{a}^T)$
   - $tr(\mathbf{AB}) = tr(\mathbf{BA})$
   - $\frac{\partial}{\partial \mathbf{x}} tr(\mathbf{Y}) = tr(\frac{\partial}{\partial \mathbf{x}}\mathbf{Y})$
   - $\frac{\partial}{\partial \mathbf{X}} tr(\mathbf{Y}) = tr(\frac{\partial}{\partial \mathbf{X}}\mathbf{Y})$
   - $\frac{\partial}{\partial \mathbf{X}} tr(\mathbf{X}) = \mathbf{I}$
   - $\frac{\partial}{\partial \mathbf{X}} tr(\mathbf{AX}) = \mathbf{A}^T$
   - $\frac{\partial}{\partial \mathbf{X}} tr(\mathbf{X}^\mathbf{T}\mathbf{X}) = 2\mathbf{X}^\mathbf{T}$
   - $\frac{\partial}{\partial \mathbf{X}} \log|\mathbf{X}| = \mathbf{X}^{-T}$
   - $\frac{\partial}{\partial \mathbf{X}} Trace[\mathbf{A}^T\mathbf{X}^{-1}\mathbf{B}] = -\mathbf{X}^{-T}\mathbf{A}\mathbf{B}^T\mathbf{X}^{-T}$
2. Show that the determinant of a matrix **A** is the product of its eigenvalues *i.e.* $|\mathbf{A}| = \prod \lambda_i$
3. Using Matrix Inversion Lemma and Kronecker properties
   $\left(\mathbf{V}_1^{-1} \otimes \mathbf{X}^T\mathbf{C}_1^{-1}\mathbf{X} + \mathbf{V}_2^{-1} \otimes \mathbf{C}_2^{-1}\right)^{-1}(\mathbf{V}_1^{-1} \otimes \mathbf{X}^T\mathbf{C}_1^{-1})$
   $= (\mathbf{V}_2 \otimes \mathbf{C}_2\mathbf{X}^T)\left(\mathbf{V}_2 \otimes \mathbf{X}\mathbf{C}_2\mathbf{X}^T + \mathbf{V}_1 \otimes \mathbf{C}_1\right)^{-1}$

Questions for Self Study

3. Show that the eigenvectors $\mathbf{v}_i$ of a symmetric matrix $\mathbf{A}$ are orthogonal if its eigenvalues are distinct *i.e.* $\lambda_i \neq \lambda_j$.
4. Compute $f = \sin(A)$ where $\mathbf{A}$ is given as $\mathbf{A} = \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix}$.
5. Show that $\frac{\partial x_i}{\partial y_i} = \mathbf{J} = \mathbf{U}^T$ for a transformation given as $\mathbf{x} - \mu = \mathbf{U}\mathbf{y}$
6. Show that the maximum likelihood solution for *i.i.d.* data $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$, of mean and covariance are;

   $\mu_{ML} = \frac{1}{N} \sum_{n=1}^{N} \mathbf{x}_n$ and $\Sigma_{ML} = \frac{1}{N} \sum_{n=1}^{N} (\mathbf{x}_n - \mu)(\mathbf{x}_n - \mu)^T$
7. Show that the optimal solution for $\mathcal{L} = (\mathbf{v} - \mathbf{A}\mathbf{x})^T \mathbf{C}_e^{-1} (\mathbf{v} - \mathbf{A}\mathbf{x}) + \lambda \mathbf{x}^T \mathbf{L}^T \mathbf{L}\mathbf{x}$ is given by $\mathbf{x} = (\mathbf{A}^T \mathbf{C}_e^{-1} \mathbf{A} + \lambda \mathbf{L}^T \mathbf{L})^{-1} \mathbf{A}^T \mathbf{C}_e^{-1} \mathbf{v}$.
8. Reexpress $||\mathbf{X}||_2^2, ||\mathbf{X}||_{1,2}^2, ||\mathbf{X}||_{2,1}$ so that they can be differentiated with respect to $\mathbf{X}$ and determine these derivative expressions.

   Example :

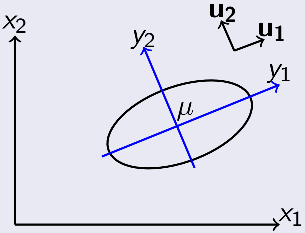
   $||\mathbf{X}||_2^2 = trace(\mathbf{X}^T \mathbf{X})$ and $\frac{\partial ||\mathbf{X}||_2^2}{\partial \mathbf{X}} = 2\mathbf{X}^T$
9. Prove the identity

   $(\mathbf{L}^T \mathbf{X} \mathbf{L} + \mathbf{Y})^{-1} \mathbf{L}^T \mathbf{X} = \mathbf{Y}^{-1} \mathbf{L}^T (\mathbf{L} \mathbf{Y}^{-1} \mathbf{L}^T + \mathbf{X}^{-1})^{-1}$

Multivariate Gaussian Distribution

$$\mathcal{N}(\mathbf{x}|\mu,\Sigma) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)}$$

Geometry of Multivariate Gaussian



$$\Delta^2 = (\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)$$

$$\Sigma^{-1} = \sum_{i=1}^{D} \frac{1}{\lambda_i}\mathbf{u}_i\mathbf{u}_i^T$$

$$y_i = \mathbf{u}_i^T(\mathbf{x}-\mu)$$

$$\Delta^2 = \sum_{i=1}^{D} \frac{y_i^2}{\lambda_i}$$

Show that $E\{\mathbf{x}\} = \mu$ for a Multivariate Gaussian.

$$E\{\mathbf{x}\} = \int \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)}\mathbf{x}d\mathbf{x} = \int \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}\mathbf{z}^T\Sigma^{-1}\mathbf{z}}(\mathbf{z}+\mu)d\mathbf{z}$$

$$= \int \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}\mathbf{z}^T\Sigma^{-1}\mathbf{z}}\mathbf{z}d\mathbf{z} + \mu\int \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}\mathbf{z}^T\Sigma^{-1}\mathbf{z}}d\mathbf{z} = 0 + \mu\cdot 1 = \mu$$

Show that $Var\{\mathbf{x}\} = \Sigma$ for a Multivariate Gaussian.

$\Sigma = \sum_{i=1}^{D} \lambda_i \mathbf{u}_i \mathbf{u}_i^T, \quad \mathbf{x} - \mu = \mathbf{U}\mathbf{y} = \sum_{i=1}^{D} \mathbf{u}_i y_i \quad \frac{\partial x_i}{\partial y_j} = \mathbf{J}_{ij} = \mathbf{U}_{ji}, \quad |\mathbf{J}| = |\mathbf{U}^T|$ since $\mathbf{U}^T = \mathbf{U}^{-1}$

$\mathbf{U}^T\mathbf{U} = \mathbf{I}$ and $|\mathbf{U}^T\mathbf{U}| = |\mathbf{U}^T||\mathbf{U}| = 1, \ |\mathbf{J}| = 1 \quad dx_i = |\frac{\partial x_i}{\partial y_j}|dy_j$ or $d\mathbf{x} = |\mathbf{J}|d\mathbf{y} = d\mathbf{y}$

Also $p(\mathbf{y}) = |\mathbf{J}|p(\mathbf{x})$

$$E\{(\mathbf{x}-\mu)(\mathbf{x}-\mu)^T\} = \int \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)} (\mathbf{x}-\mu)(\mathbf{x}-\mu)^T d\mathbf{x}$$

$$= \frac{1}{(2\pi)^{D/2}\prod_{p=1}^{D}\lambda_p^{1/2}} \int \sum_{i,j} \mathbf{u}_i y_i y_j \mathbf{u}_j^T e^{-\frac{1}{2}\sum_{k=1}^{D}\frac{y_k^2}{\lambda_k}} d\mathbf{y}$$

$$= \frac{1}{(2\pi)^{D/2}\prod_{p=1}^{D}\lambda_p^{1/2}} \left[\mathbf{u}_1\mathbf{u}_1^T \int y_1^2 e^{\frac{y_1^2}{\lambda_1}} dy_1 \int e^{\frac{y_2^2}{\lambda_2}} dy_2 \ldots\right] +$$

$$\left[\mathbf{u}_2\mathbf{u}_2^T \int e^{\frac{y_1^1}{\lambda_1}} dy_1 \int y_2^2 e^{\frac{y_2^2}{\lambda_2}} dy_2 \ldots\right] + \cdots + \left[\mathbf{u}_D\mathbf{u}_D^T \int e^{\frac{y_1^1}{\lambda_1}} dy_1 \ldots \int y_D^2 e^{\frac{y_D^2}{\lambda_D}} dy_D\right]$$

$$= \mathbf{u}_1\mathbf{u}_1^T\lambda_1 + \mathbf{u}_2\mathbf{u}_2^T\lambda_2 + \cdots = \sum_{i=1}^{D} \lambda_i \mathbf{u}_i \mathbf{u}_i^T = \Sigma$$

Maximum Likelihood Estimation

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} \exp\left\{ -\frac{1}{2}(\mathbf{x}-\mu)^T \Sigma^{-1} (\mathbf{x}-\mu) \right\} d\mathbf{x}$$

log-likelihood of a multivariate Gaussian distribution

$\log p(\mathbf{x}|\mu, \Sigma) = -\frac{D}{2}\log(2\pi) - \frac{1}{2}|\Sigma| - \frac{1}{2}(\mathbf{x}-\mu)^T \Sigma^{-1}(\mathbf{x}-\mu)$

$$\frac{\partial \log p(\mathbf{x})}{\partial \mathbf{x}} = -\frac{1}{2}\frac{\partial}{\partial \mathbf{x}}[-\mu^T \Sigma^{-1}\mathbf{x} + \mathbf{x}^T \Sigma^{-1}\mathbf{x} - \mathbf{x}^T \Sigma^{-1}\mu] = 0 \rightarrow \mathbf{x} = \mu$$

Maximum Likelihood Estimation

Given i.i.d. data $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)^T$ the log-likelihood function is given by

$$\ln p(\mathbf{X}|\mu, \Sigma) = -\frac{ND}{2}\ln(2\pi) - \frac{N}{2}\ln|\Sigma| - \frac{1}{2}\sum_{n=1}^{N}(\mathbf{x}_n - \mu)^T \Sigma^{-1}(\mathbf{x_n} - \mu)$$

$$\frac{\partial}{\partial \mu}\ln p(\mathbf{X}|\mu, \Sigma) = \sum_{n=1}^{N}\Sigma^{-1}(\mathbf{x} - \mu) = 0 \to \mu_{ml} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}_n$$

$$\frac{\partial}{\partial \Sigma}\ln p(\mathbf{X}|\mu, \Sigma) = 0 = -\frac{N}{2}\Sigma^{-T} + \frac{1}{2}\sum_{n=1}^{N}\Sigma^{-T}(\mathbf{x}_n - \mu)x_n - \mu)^T\Sigma^{-T}$$

$$\Sigma = \tfrac{1}{N}\sum_{n=1}^{N}(\mathbf{x}_n - \mu)(\mathbf{x}_n - \mu)^T$$

Self Study : Determine $p(y) = \int p(y|x)p(x)dx$

$p(y|x) = \mathcal{N}(y|Hx, \Lambda^{-1})$ and $p(x) = \mathcal{N}(x|0, \Phi^{-1})$

$p(y) = (2\pi)^{-N/2}|\Lambda|^{1/2}(2\pi)^{-M/2}|\Phi|^{1/2} \int e^{-\frac{1}{2}(y-Hx)^T\Lambda(y-Hx)}e^{-\frac{1}{2}x^T\Phi x}dx$

$y^T\Lambda y + x^TH^T\Lambda Hx - y^T\Lambda Hx - x^TH^T\Lambda y + x^T\Phi x$

Let's have the x terms to be a complete square as $(x-m_x)^T\Lambda_x(x-m_x)$ and identify Λ_x and m_x as

$\Lambda_x = H^T\Lambda H + \Phi$ and $\Lambda_x m_x = H^T\Lambda y$ or $m_x = \Lambda_x^{-1}H^T\Lambda y$

Then all the terms can be reexpressed as

$$x^T\Lambda_x x - \underbrace{y^T\Lambda H\Lambda_x^{-1}}_{m_x^T}\Lambda_x x - x^T\Lambda_x\underbrace{\Lambda_x^{-1}H^T\Lambda}_{m_x} - y + m_x^T\Lambda_x m_x - m_x^T\Lambda_x m_x + y^T\Lambda y$$

Then the first 4 terms can be reduced to $(x-m_x)^T\Lambda_x(x-m_x)$

and the last two $-(y^T\Lambda H\Lambda_x^{-1})\Lambda_x(\Lambda_x^{-1}H^T\Lambda y) + y^T\Lambda y = y^T(\Lambda - \Lambda H\Lambda_x^{-1}H^T\Lambda)y$

The integral over the x terms yield $\int e^{-\frac{1}{2}(x-m_x)^T\Lambda_x(x-m_x)}dx = (2\pi)^{M/2}|\Lambda_x|^{-1/2}$

and the remaing y terms form a m.v. gaussian with

$\Lambda_y = \Lambda - \Lambda H\Lambda_x^{-1}H^T\Lambda = \Lambda - \Lambda H(\Phi + H^T\Lambda H)^{-1}H^T\Lambda = \Lambda - \Lambda H(I + \Phi^{-1}H^T\Lambda H)^{-1}\Phi^{-1}H^T\Lambda$

By using Matrix Inversion Lemma $(A+BCD)^{-1} = A^{-1} - A^{-1}B(I + CDA^{-1}B)^{-1}CDA^{-1}$

with $\Lambda \longrightarrow A^{-1}$, $B \longrightarrow H$, $C \longrightarrow \Phi^{-1}$ and $D \longrightarrow H^T$

$\Lambda_y = (\Lambda^{-1} + H\Phi^{-1}H^T)^{-1}$

$p(y) = (2\pi)^{-\frac{N}{2}}|\Lambda|^{\frac{1}{2}}(2\pi)^{-\frac{M}{2}}|\Phi|^{1/2}e^{-\frac{1}{2}y^T\Lambda_y y}(2\pi)^{\frac{M}{2}}|\Lambda_x|^{-1/2}$

$|\Lambda^{-1}||\Phi^{-1}||\Lambda_x| = |\Lambda^{-1}||\Phi^{-1}||H^T\Lambda H + \Phi|$

Using the identity $|A + XDX^T| = |A|\cdot|D|\cdot|D^{-1} + X^TA^{-1}X|$

with $A \longrightarrow \Lambda^{-1}$ $D \longrightarrow \Phi^{-1}$, $X \longrightarrow H$ we get $|\Lambda^{-1}||\Phi^{-1}||H^T\Lambda H + \Phi|^{-1} = |\Lambda^{-1} + H\Phi^{-1}H^T| = |\Lambda_y^{-1}|$

$p(y) = (2\pi)^{-N/2}|\Lambda_y|e^{-\frac{1}{2}y^T\Lambda_y y}$