Technical Note:
The Navigation Handbook of Notation, Reference Frames, Linear Algebra, Probability Theory, and State Estimation.

Paul F. Roysdon, Ph.D.

I. Notation

A. Conventions and Symbols

Table I-A summarizes the notational conventions that are used throughout this text. A non-bold face symbol denotes a scalar quantity. A bold face symbol denotes either a vector (typically lower case) or a matrix (typically upper case). It is important to make the distinction between a *true* value, a *calculated*, *estimated*, or a *measured* value. As shown in Table I-A, the true value has no additional mark; the calculated value has a "hat" on it; the measured value has a "tilde" above it. The error is defined as the true value minus the estimated value. The error quantity is indicated with a δ, for example $\delta\boldsymbol{x} = \boldsymbol{x} - \hat{\boldsymbol{x}}$.

TABLE I

Notational conventions.

x	non-bold face variables denote *scalars*
$\boldsymbol{x}$	boldface lower-case denotes *vector* quantities
$\mathbf{X}$	boldface upper-case denotes *matrix* quantities
$\boldsymbol{x}$	true value of $\boldsymbol{x}$
$\hat{\boldsymbol{x}}$	calculated value of $\boldsymbol{x}$
$\tilde{\boldsymbol{x}}$	measured value of $\boldsymbol{x}$
$\delta\boldsymbol{x}$	error $\boldsymbol{x} - \hat{\boldsymbol{x}}$
$\mathbf{R}_a^b$	transformation matrix from reference frames a to b
$\boldsymbol{x}^a$	vector $\boldsymbol{x}$ represented with respect to frame a
$\mathbb{R}, \mathbb{R}^+, \mathbb{R}^n$	real numbers, reals greater than 0, n-tuples of reals
$\mathbb{N}$	natural numbers $\{0, 1, 2, \ldots\}$
$\mathbb{C}$	complex numbers
$\mathbb{Z}$	integer numbers
$(a\,..\,b), [a\,..\,b]$	open interval, closed interval
$\langle \ldots \rangle$	sequence (a list in which order matters)
$\{\ldots\}$	set (a list in which order does not matter)
$x_{i,j}$	row i and column j entry of matrix $\mathbf{X}$
$\mathbf{0}_{n\times m}$ or $\mathbf{0}$	zero matrix
$\mathbf{I}_{n\times n}$ or $\mathbf{I}$	identity matrix
$\lvert\mathbf{X}\rvert$	determinant of matrix $\mathbf{X}$
R, N	range space, null space
R_∞, N_∞	generalized range space and null space
$\mathcal{N}$	Normal or Gaussian random variable
$\mathcal{L}$	Laplace random variable

In navigation system theory, it is common to use several different reference frames. Vector quantities are transformed between these different frames using rotational transformations which can be represented as matrices. The notation $\mathbf{R}_a^b$ denotes the transformation matrix from reference frame a to reference frame b. When a vector is being "coordinatized" in frame a, the notation x^a is used. All of the symbols representing the different reference frames are listed in Table I-A. For example, the true rotation matrix from body frame to Earth frame would be denoted as $\mathbf{R}_b^e$. The estimated (or computed) rotation matrix from body frame to Earth frame would be denoted as $\hat{\mathbf{R}}_b^e$.

TABLE II

Reference frame symbol definitions.

a	accelerometer frame (non-orthogonal)
b	body frame
e	ECEF frame
g	gyro frame (non-orthogonal) or geographic
i	inertial frame
n	navigation frame (typically geographic)
p	platform frame
t	tangent plane frame

The text is concerned with estimating the state of a vehicle which evolves as a function of the continuous-time variable t. The estimation routines are implemented in discrete-time on digital computers. In addition, the estimation algorithm incorporates information from sensors that provide readings at discrete-time instants. Therefore, the estimates will evolve on the computer using a discrete-time increment T. The discrete-time increments are denoted $tk = kT$ for positive, integer values of k. Therefore, $x(tk) = x(t)|_{t=kT}$. Often, as a further shorthand, the notation x_k will be used instead of $x(t_k)$. When the measurements do not occur with a fixed sample period T (i.e., asynchronous measurements), the symbol y_k simply means the k-th measurement which will have occurred at time t_k. When considering state estimation, we are often concerned with specifying the state variables either just before or after a measurement that occurred at time t_k. We denote the value of the variable just prior to the measurement as x_k and the value of the variable just after the measurement as x_k^+.

B. Useful Constants

Table I-B presents various constants useful in navigation system design and analysis. The precision on several of the presented variables is critical to system performance. For example, the precision presented for π is that specified for the GPS system in the interface communication document [1]. Due to the radius of the satellite orbits, satellite position calculations are sensitive to this value.

C. Acronyms

The acronyms used in the text should be defined at their first usage which should be listed in the index.

TABLE III

IMPORTANT CONSTANTS RELATED TO GPS AND THE WGS-84 WORLD GEODETIC SYSTEM.

Symbol	Value	Units	Description
π	3.1415926535898		pi
ω_{ie}	7.2921151467×10^5	rad/s	Earth rotation rate
μ	3.986005×10^{14}	$\frac{m^3}{s^2}$	Univ. grav. constant
c	2.99792458×10^8	$\frac{m}{s}$	Speed of light
F	$4.442807633 \times 10^{10}$	$\frac{s}{\sqrt{m}}$	Earth flattening factor
a	6378137.0	m	Earth semi-major axis
b	6356752.3	m	Earth semi-minor axis
f_1	1575.42	M Hz	L1 carrier frequency
f_2	1227.60	M Hz	L2 carrier frequency
λ_1	$\frac{c}{f_1} \approx 19.0$	cm	L1 wavelength
λ_2	$\frac{c}{f_2} \approx 24.4$	cm	L2 wavelength
λ_w	$\frac{c}{f_1 - f_2} \approx 86.2$	cm	wide lane wavelength
λ_n	$\frac{c}{f_1 + f_2} \approx 10.7$	cm	narrow lane wavelength

D. Greek letters

The Greek letters used in the text are defined in Table I-D, with their proper pronounciation.

TABLE IV

GREEK LETTERS WITH PRONOUNCIATION.

α	alpha *AL-fuh*
β	beta *BAY-tuh*
γ, Γ	gamma *GAM-muh*
δ, Δ	delta *DEL-tuh*
ϵ	epsilon *EP-suh-lon*
ζ	zeta *ZAY-tuh*
η	eta *AY-tuh*
θ, Θ	theta *THAY-tuh*
ι	iota *eye-OH-tuh*
κ	kappa *KAP-uh*
λ, Λ	lambda *LAM-duh*
μ	mu *MEW*
ν	nu *NEW*
ξ, Ξ	xi *KSIGH*
o	omicron *OM-uh-CRON*
π, Π	pi *PIE*
ρ	rho *ROW*
σ, Σ	sigma *SIG-muh*
τ	tau *TOW (as in cow)*
υ, Υ	upsilon *OOP-suh-LON*
ϕ, Φ	phi *FEE, or FI (as in hi)*
χ	chi *KI (as in hi)*
ψ, Ψ	psi *SIGH, or PSIGH*
ω, Ω	omega *oh-MAY-guh*

II. COORDINATES AND NOTATIONS

The coordinates used in this text are given in the following sections.

A. ECI

An inertial frame is a reference frame in which Newtons laws of motion apply. An inertial frame is therefore not accelerating, but may be in uniform linear motion. The origin of the inertial coordinate system is arbitrary, and the coordinate axis may point in any three mutually perpendicular directions. All inertial sensors produce measurements relative to an inertial frame, resolved along the instrument sensitive axis.

For discussion purposes it is sometimes convenient to define an Earth centered inertial (ECI) frame which at a specified initial time has its origin coincident with the center of mass of the Earth, see Figure 2.3. At the same initial time, the inertial x and z axes point toward the vernal equinox and along the Earth spin axis, respectively. The y-axis is defined to complete the right-handed coordinate system. The axes define an orthogonal coordinate system. Note that the ECEF frame, defined in Section 2.2.2, rotates with respect to this ECI frame with angular rate ω_i^e; therefore, in the ECI frame the angular rate vector is $\boldsymbol{\omega}_i^e = [0, 0, \omega_i^e]^\intercal$.

B. ECEF

This frame has its origin fixed to the center of the Earth. Therefore, the axes rotate relative to the inertial frame with frequency

$$\omega_{ie} \approx \left(\frac{1 + 365.25\ \text{cycle}}{(365.25)(24)\ \text{hr}}\right)\left(\frac{2\pi\ \text{rad/cycle}}{3600\ \text{sec/hr}}\right)$$
$$= 7.292115 \times 10^5\ \frac{\text{rad}}{\text{sec}} \tag{1}$$

due to the 365.25 daily Earth rotations per year plus the one annual revolution about the sun. Relative to inertial frame, the Earth rotational rate vector expressed relative to the ECEF axes is $\boldsymbol{\omega}_{ie}^e = [0, 0, 1]^\intercal \omega_{ie}$

The Earths geoid is usually approximated as an ellipsoid of revolution about its minor axis. A consistent set of Earth shape (i.e., ellipsoid) and gravitation model parameters must be used in any given application. Therefore, the value for ω_{ie} in eqn. (2.2) should only be considered as an approximated value. Earth shape and gravity models are discussed in Section 2.3. Due to the Earth rotation, the ECEF frame-of-reference is not an inertial reference frame. Two common coordinate systems for the ECEF frame-of-reference are discussed in Section 2.3.

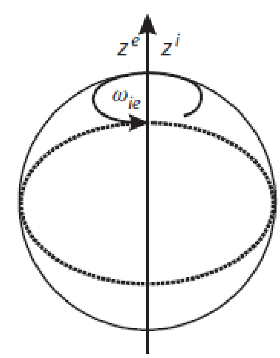

Fig. 1. Rotation of the ECEF frame with respect to an Earth-centered inertial frame. The vectors x_i, y_i, and z_i, define the axes of the ECI frame. The vector x_e defines the x-axis of the ECEF frame [2].

C. Geographic LLH

The geographic frame is defined locally, relative to the Earths geoid. The origin of the geographic frame moves with the system and is defined as the projection of the platform origin P onto the reference ellipsoid, see the left portion of Figure 2.2. The geographic z-axis points toward the interior of the ellipsoid along the ellipsoid normal. The x-axis points toward true north (i.e., along the projection of the Earth angular rate vector ω_{ie} onto the plane orthogonal to the z-axis). The y-axis points east to complete the orthogonal, right-handed rectangular coordinate system.

Since the origin of the geographic frame travels along with the vehicle, the axes of the frame rotate as the vehicle moves either north or east. The rotation rate is discussed in Example 2.6. Because the geographic frame rotates with respect to inertial space, the geographic frame is not an inertial frame.

Two additional points are worth specifically stating. First, true north and magnetic north usually are distinct directions. Second, as illustrated by the exaggerated ellipse in the left portion of Figure 2.2, the normal to a reference ellipsoid (approximate Earth geoid) does not pass through the center of the ellipsoid, unless the platform origin P is at the equator or along the Earth spin axis.

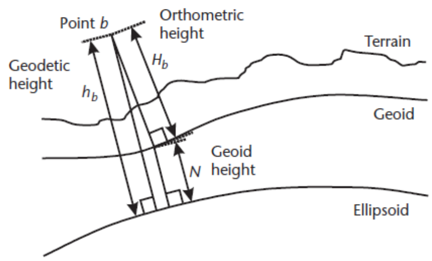

Fig. 2. Add text here...

D. Local Tangent Plane NED

The local geodetic frame is the north, east, down rectangular coordinate system we often refer to in our everyday life (see Figure 3). It is determined by fitting a tangent plane to the geodetic reference ellipse at a point of interest. The tangent plane is attached to a fixed point on the surface of the Earth at some convenient point for local measurements. This point is the origin of the local frame. The x-axis points to true north. The z-axis points toward the interior of the Earth, perpendicular to the reference ellipsoid. The y-axis completes the right-handed coordinate system, pointing east.

For a stationary system, located at the origin of the tangent frame, the geographic and tangent plane frames coincide. When a system is in motion, the tangent plane origin is fixed, while the geographic frame origin is the projection of the platform origin onto the reference ellipsoid of the Earth. The tangent frame system is often used for local navigation (e.g., navigation relative to a runway).

E. Body Frame

In navigation applications, the objective is to determine the position and velocity of a vehicle based on measurements from various sensors attached to a platform on the vehicle. This motivates the definition of vehicle and instrument frames-of-reference and their associated coordinate systems.

The body frame is rigidly attached to the vehicle of interest, usually at a fixed point such as the center of gravity. Picking the center of gravity as the location of the body frame

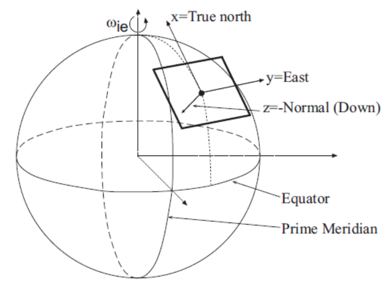

Fig. 3. Local geodetic or tangent plane reference coordinate system in relation to the ECEF frame [2].

origin simplifies the derivation of the kinematic equations [90] and is usually convenient for control system design. The u-axis is defined in the forward direction of the vehicle. The w-axis is defined pointing to the bottom of the vehicle and the v-axis completes the right-handed orthogonal coordinate system. The axes directions so defined (see Figure 2.5) are not unique, but are typical in aircraft and underwater vehicle applications. In this text, the above definitions will be used. In addition, the notation $[u, v, w]$ for the vehicle axes unit vectors has been used instead of $[x, y, z]$, as the former is more standard.

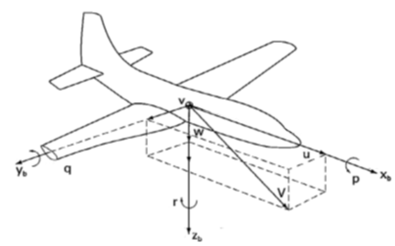

Fig. 4. Vehicle (body) coordinate system.

As indicated in Figure 4, the rotation rate vector of the body frame relative to inertial space, resolved along the body axis is denoted by $\boldsymbol{\omega}_{ib}^{b} = [p, q, r]^\intercal$ where p is the angular rate about the u-axis (i.e., roll rate), q is the angular rate about the v-axis (i.e., pitch rate), and r is the angular rate about the w-axis (i.e., yaw rate). Each angular rate is positive in the right-hand sense. The body frame is not an inertial frame-of-reference.

III. Attitude Representations

The attitude of a rigid body can be represented by rotation matrix, quaternion and Euler angles.

A. Rotation Matrix

The rotation matrix $\mathbf{R}$ evolves in the special orthogonal group of degree three:

$$SO(3) = \{\mathbf{R} \in \mathbb{R}^{3\times 3} | \mathbf{R}^\intercal \mathbf{R} = \mathbf{R}\mathbf{R}^\intercal = \mathbf{I}_3, det(\mathbf{R}) = 1\} \quad (2)$$

B. Quaternion

This section review the basics of quaternion that are useful for navigation and control. The details of the quaternion algebra can be found in [3]. The notation of quaternion used in this thesis follows the JPL convention, instead of the Hamilton convention. A good reference about the discussion of the two notations is [4].

1) Unit Quaternion: Unit quaternion ($\bar{\mathbf{q}}$) uses four parameters to represent a rotation matrix. Such representation is globally non-singular. Unit quaternion is evolving in the three-sphere $\mathbb{S}^3$, embedded in $\mathbb{R}^4$, $\mathbb{S}^3 = \{\bar{\mathbf{q}} \in \mathbb{R}^4 | \bar{\mathbf{q}}^\intercal \bar{\mathbf{q}} = 1\}$. The quaternion is a unit quaternion if it satisfies:

$$|\bar{\mathbf{q}}| = \sqrt{\bar{\mathbf{q}}^\intercal \bar{\mathbf{q}}} = \sqrt{q_1^2 + |\mathbf{q}|^2} = 1 \tag{3}$$

where $\mathbf{q} = [q_2, q_3, q_4]^\intercal$. Usually, the unit quaternion is written as:

$$\begin{aligned} \bar{\mathbf{q}} &= \begin{bmatrix} q_1 \\ \mathbf{q} \end{bmatrix} &(4)\\ &= \begin{bmatrix} \hat{\mathbf{k}} \sin(\theta/2) \\ \cos(\theta/2) \end{bmatrix} &(5)\end{aligned}$$

In this notation, the unit vector $\hat{\mathbf{k}}$ describes the rotation axis and θ is the angle of rotation. Note that, one attitude configuration has two quaternion representations: $\bar{\mathbf{q}}$ and $-\bar{\mathbf{q}}$. They are differed by the rotating directions around the rotation axis to reach the target configuration.

2) Quaternion Algebra: We define the skew-symmetric matrix $\lfloor \boldsymbol{x} \times \rfloor$ for any $\boldsymbol{x} \in \mathbb{R}^3$ as

$$\lfloor \boldsymbol{x} \times \rfloor = \begin{bmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{bmatrix} \tag{6}$$

For the convenience, the skew-symmetric matrix is also denoted as $S(\boldsymbol{x})$ in this thesis. So for any two vectors $\boldsymbol{x}, \mathbf{y} \in \mathbb{R}^3$, the cross product can be denoted as $\boldsymbol{x} \times \mathbf{y} = \lfloor \boldsymbol{x} \times \rfloor \mathbf{y}$. Also we have $\lfloor \boldsymbol{x} \times \rfloor^\intercal = \lfloor -\boldsymbol{x} \times \rfloor$.

The corresponding rotation matrix for a given quaternion is

$$\begin{aligned} \mathbf{R}(\bar{\mathbf{q}}) &= (2q_1^2 - 1)\mathbf{I}_{3\times3} - 2q_1 \lfloor \mathbf{q} \times \rfloor + 2\mathbf{q}\mathbf{q}^\intercal &(7)\\ &= \mathbf{I}_{3\times3} - 2q_1 \lfloor \mathbf{q} \times \rfloor + 2 \lfloor \mathbf{q} \times \rfloor^2 &(8)\end{aligned}$$

The multiplication of two quaternions is defined as

$$\bar{\mathbf{q}} \otimes \bar{\mathbf{p}} = \begin{bmatrix} q_1 \mathbf{p} + p_1 \mathbf{q} - \mathbf{q} \times \mathbf{p} \\ q_1 p_1 - \mathbf{q}^\intercal \mathbf{p} \end{bmatrix} \tag{9}$$

The quaternion multiplication is distributive and associative but not commutative. The multiplication of quaternions is analogous to the multiplication of rotation matrix, in the same order:

$$\mathbf{R}(\bar{\mathbf{q}})\mathbf{R}(\bar{\mathbf{p}}) = \mathbf{R}(\bar{\mathbf{q}} \otimes \bar{\mathbf{p}}) \tag{10}$$

The inverse of quaternion is defined as

$$\bar{\mathbf{q}}^{-1} = \begin{bmatrix} q_1 \\ -\mathbf{q} \end{bmatrix} \tag{11}$$

We define the identity quaternion as $\bar{\mathbf{q}}_o = [1 \;\; \mathbf{0}]^\intercal$. For any given quaternion $\bar{\mathbf{q}}$, we have

$$\bar{\mathbf{q}} \otimes \bar{\mathbf{q}}^{-1} = \bar{\mathbf{q}}^{-1} \otimes \bar{\mathbf{q}} = \begin{bmatrix} 1 \\ \mathbf{0} \end{bmatrix} \tag{12}$$

From a given vector $\mathbf{p}$, we define its quaternion form as

$$\bar{\mathbf{q}}_{\mathbf{p}} = \begin{bmatrix} 0 \\ \mathbf{p} \end{bmatrix} \tag{13}$$

The transformation of a vector from frame a to frame b using quaternion is given by:

$$\begin{aligned} {}^b\bar{\mathbf{q}}_{\mathbf{p}} &= {}^b_a\bar{\mathbf{q}} \otimes {}^a\bar{\mathbf{q}}_{\mathbf{p}} \otimes {}^b_a\bar{\mathbf{q}}^{-1} &(14)\\ &= \begin{bmatrix} 0 \\ {}^b_a\mathbf{R}\,{}^a\mathbf{p} \end{bmatrix} &(15)\end{aligned}$$

The corresponding formula for skew-symmetric matrix $\lfloor \boldsymbol{x} \times \rfloor$ is

$$\lfloor {}^b\boldsymbol{x} \times \rfloor = {}^b_a\mathbf{R} \lfloor {}^a\boldsymbol{x} \times \rfloor {}^b_a\mathbf{R}^\intercal \tag{16}$$

3) Quaternion Kinematics: Here we use $\{G\}$ to represent inertial frame and $\{B\}$ to represent body frame. For the angular velocity, we use $\boldsymbol{\omega}_{GB}^B$ to denote the angular velocity of body frame w.r.t. inertial frame expressed in body frame.

The kinematic equation for rotation ${}^B_G\mathbf{R}$ is:

$$\begin{aligned} {}^B_G\dot{\mathbf{R}} &= -\lfloor \boldsymbol{\omega}_{GB}^B \times \rfloor {}^B_G\mathbf{R} &(17)\\ {}^B_G\dot{\bar{\mathbf{q}}} &= \frac{1}{2}\bar{\mathbf{q}}_{\boldsymbol{\omega}_{GB}^B} \otimes {}^B_G\bar{\mathbf{q}} &(18)\\ &= \frac{1}{2}\boldsymbol{\Omega}(\boldsymbol{\omega}_{GB}^B) {}^I_G\bar{\mathbf{q}} &(19)\\ &= \frac{1}{2}\begin{bmatrix} \lfloor \mathbf{q} \times \rfloor + q_1 \mathbf{I} \\ -\mathbf{q}^\intercal \end{bmatrix} \boldsymbol{\omega}_{GB}^B &(20)\end{aligned}$$

where

$$\boldsymbol{\Omega}(\boldsymbol{\omega}) = \begin{bmatrix} -\lfloor \boldsymbol{\omega} \times \rfloor & \boldsymbol{\omega} \\ -\boldsymbol{\omega}^\intercal & \mathbf{0} \end{bmatrix} \tag{21}$$

The kinematic equation for rotation ${}^G_B\mathbf{R}$ is:

$$\begin{aligned} {}^G_B\dot{\mathbf{R}} &= {}^G_B\mathbf{R} \lfloor \boldsymbol{\omega}_{GB}^B \times \rfloor &(22)\\ {}^G_B\dot{\bar{\mathbf{q}}} &= \frac{1}{2} {}^G_B\bar{\mathbf{q}} \otimes \bar{\mathbf{q}}_{-\boldsymbol{\omega}_{GB}^B} &(23)\\ &= \frac{1}{2}\begin{bmatrix} \lfloor \mathbf{q} \times \rfloor - q_1 \mathbf{I} \\ \mathbf{q}^\intercal \end{bmatrix} \boldsymbol{\omega}_{GB}^B &(24)\end{aligned}$$

The derivation can refer to Section 2.6.1 of [2] for the dynamics of rotation matrix and Section 2.4 in page 16 of [3] for the dynamics of quaternion.

Let $\boldsymbol{\Phi} = \begin{bmatrix} \lfloor \mathbf{q} \times \rfloor + q_1 \mathbf{I} \\ -\mathbf{q}^\intercal \end{bmatrix}$, then we have $\boldsymbol{\Phi}^\intercal \boldsymbol{\Phi} = \mathbf{I}_{3\times3}$. The proof will make use of this property: $\lfloor \boldsymbol{x} \times \rfloor^2 = \boldsymbol{x}\boldsymbol{x}^\intercal - \|\boldsymbol{x}\|^2 \mathbf{I}$ (see eqn. (55) in [3]).

The attitude error is expressed in *error quaternion* form:

$$\begin{aligned} \delta\bar{\mathbf{q}} &= \bar{\mathbf{q}} \otimes \hat{\bar{\mathbf{q}}}^{-1} &(25)\\ &= \begin{bmatrix} \cos(\delta\theta/2) \\ \hat{\mathbf{k}} \sin(\delta\theta/2) \end{bmatrix} &(26)\\ &\simeq \begin{bmatrix} 1 \\ \frac{1}{2}\delta\boldsymbol{\theta} \end{bmatrix} &(27)\end{aligned}$$

The corresponding rotation matrix is

$$\mathbf{R}(\delta\bar{\mathbf{q}}) \simeq \mathbf{I}_{3\times 3} - \lfloor \delta\boldsymbol{\theta}\times \rfloor \tag{28}$$

C. Euler Angles

In air force convention, Euler angles parameterize the rotation matrix ${}^I_G\mathbf{R}$ by three rotation angles: roll(ϕ), pitch(θ) and yaw(ψ). However, such minimum representation of rotation suffers from the singularity problem. The Euler angles represent the rotation ${}^I_G\mathbf{R}$ that rotates the global frame to the IMU frame and are followed by the right hand rule.

$$\mathbf{R}_z(\psi) = \begin{bmatrix} \cos(\psi) & \sin(\psi) & 0 \\ -\sin(\psi) & \cos(\psi) & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{29}$$

$$\mathbf{R}_y(\theta) = \begin{bmatrix} \cos(\theta) & 0 & -\sin(\theta) \\ 0 & 1 & 0 \\ \sin(\theta) & 0 & \cos(\theta) \end{bmatrix} \tag{30}$$

$$\mathbf{R}_x(\phi) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\phi) & \sin(\phi) \\ 0 & -\sin(\phi) & \cos(\phi) \end{bmatrix} \tag{31}$$

Hence we have:

$${}^I_G\mathbf{R} = \mathbf{R}_x(\phi)\mathbf{R}_y(\theta)\mathbf{R}_z(\psi) \tag{32}$$

D. Conversion Between Representations

Conversions to/from Euler, quaternion, and rotation matrix are provided in the following subsections III-D.1 thru III-D.6.

1) Euler to Quaternion: Euler angle vector to quaternion conversion using the *XYZ* sequence:

$$\mathbf{q} = [q_1, q_2, q_3, q_4]^\intercal \tag{33}$$

where

$$\begin{aligned} q_1 &= \cos\frac{\phi}{2}\cos\frac{\theta}{2}\cos\frac{\psi}{2} + \sin\frac{\phi}{2}\sin\frac{\theta}{2}\sin\frac{\psi}{2} \\ q_2 &= \sin\frac{\phi}{2}\cos\frac{\theta}{2}\cos\frac{\psi}{2} - \cos\frac{\phi}{2}\sin\frac{\theta}{2}\sin\frac{\psi}{2} \\ q_3 &= \cos\frac{\phi}{2}\sin\frac{\theta}{2}\cos\frac{\psi}{2} + \sin\frac{\phi}{2}\cos\frac{\theta}{2}\sin\frac{\psi}{2} \\ q_4 &= \cos\frac{\phi}{2}\cos\frac{\theta}{2}\sin\frac{\psi}{2} - \sin\frac{\phi}{2}\sin\frac{\theta}{2}\cos\frac{\psi}{2} \end{aligned}$$

Note: this is the unit quaternion.

2) Euler to Rotation Matrix: Convert Euler angles to a rotation matrix using the *XYZ* sequence:

$$\mathbf{R}_n^b = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \tag{34}$$

where

$$\begin{aligned} r_{11} &= \cos\psi\cos\theta \\ r_{12} &= \sin\psi\cos\theta \\ r_{13} &= -\sin\theta \\ r_{21} &= -\sin\psi\cos\phi + \cos\psi\sin\theta\sin\phi \\ r_{22} &= \cos\psi\cos\phi + \sin\psi\sin\theta\sin\phi \\ r_{23} &= \cos\theta\sin\phi \\ r_{31} &= \sin\psi\sin\phi + \cos\psi\sin\theta\cos\phi \\ r_{32} &= -\cos\psi\sin\phi + \sin\psi\sin\theta\cos\phi \\ r_{33} &= \cos\theta\cos\phi \end{aligned}$$

Note: this is the *navigation* to *body* frame rotation matrix.

3) Quaternion to Euler: Convert quaternion to Euler angles

$$\phi = \text{atan2}\left(2(q_1q_2 + q_3q_4), 1 - 2(q_2^2 + q_3^2)\right) \tag{35}$$

$$\theta = \text{asin}\left(2(q_1q_3 - q_4q_2)\right) \tag{36}$$

$$\psi = \text{atan2}\left(2(q_1q_4 + q_2q_3), 1 - 2(q_3^2 + q_4^2)\right) \tag{37}$$

Note: these are the *body* frame Euler angles.

4) Quaternion to Rotation Matrix: Convert a quaternion to a rotation matrix, where the normalized quaternion is

$$\bar{\mathbf{q}} = \frac{\mathbf{q}}{\|\mathbf{q}\|_2} \tag{38}$$

and the rotation matrix is

$$\begin{aligned} \mathbf{R}_n^b = &\left(\bar{\mathbf{q}}_1^2 - \bar{\mathbf{q}}_{\{2,3,4\}}^\intercal \bar{\mathbf{q}}_{\{2,3,4\}}\right)\mathbf{I}_3 \\ &+ 2(\bar{\mathbf{q}}_{\{2,3,4\}}\bar{\mathbf{q}}_{\{2,3,4\}}^\intercal) + 2\ \bar{\mathbf{q}}_1 \left[\bar{\mathbf{q}}_{\{2,3,4\}}\times\right] \end{aligned} \tag{39}$$

Note 1: parenthesize $\left(\bar{\mathbf{q}}_{\{2,3,4\}}\bar{\mathbf{q}}_{\{2,3,4\}}^\intercal\right)$ to ensure result is Hermitian.
Note 2: this is the *navigation* to *body* frame rotation matrix.

5) Rotation Matrix to Euler: Convert rotation matrix to Euler angles using the *XYZ* sequence:

$$\phi = \text{atan2}\left(\mathbf{R}_n^b(2,3), \mathbf{R}_n^b(3,3)\right) \tag{40}$$

$$\theta = -\text{atan2}\left(\mathbf{R}_n^b(1,3), \sqrt{1 - \mathbf{R}_n^b(1,3)^2}\right) \tag{41}$$

$$\psi = \text{atan2}\left(\mathbf{R}_n^b(1,2), \mathbf{R}_n^b(1,1)\right) \tag{42}$$

Note: these are the *body* frame Euler angles.

6) Rotation Matrix to Quaternion: Convert a rotation matrix to a normalized quaternion. There are four possible solutions. First, find $max(T_i)$ for $\{T_i | i = 1, \ldots, 4\}$, where

$$T_1 = 1 + \mathbf{R}_b^n(1,1) + \mathbf{R}_b^n(2,2) + \mathbf{R}_b^n(3,3) \tag{43}$$

$$T_2 = 1 + \mathbf{R}_b^n(1,1) - \mathbf{R}_b^n(2,2) - \mathbf{R}_b^n(3,3) \tag{44}$$

$$T_3 = 1 - \mathbf{R}_b^n(1,1) + \mathbf{R}_b^n(2,2) - \mathbf{R}_b^n(3,3) \tag{45}$$

$$T_4 = 1 - \mathbf{R}_b^n(1,1) - \mathbf{R}_b^n(2,2) + \mathbf{R}_b^n(3,3) \tag{46}$$

Then solve for $\bar{\mathbf{q}}$ based on the i^{th} case of $max(T_i)$:
case 1:

$$q_1 = 0.5 * \sqrt{max(T_i)} \tag{47}$$

$$q_2 = (\mathbf{R}_b^n(3,2) - \mathbf{R}_b^n(2,3))/4/q_1 \tag{48}$$

$$q_3 = (\mathbf{R}_b^n(1,3) - \mathbf{R}_b^n(3,1))/4/q_1 \tag{49}$$

$$q_4 = (\mathbf{R}_b^n(2,1) - \mathbf{R}_b^n(1,2))/4/q_1 \tag{50}$$

case 2:

$$q_2 = 0.5 * \sqrt{max(T_i)} \tag{51}$$
$$q_1 = (\mathbf{R}_b^n(3,2) - \mathbf{R}_b^n(2,3))/4/q_2 \tag{52}$$
$$q_3 = (\mathbf{R}_b^n(1,2) + \mathbf{R}_b^n(2,1))/4/q_2 \tag{53}$$
$$q_4 = (\mathbf{R}_b^n(3,1) + \mathbf{R}_b^n(1,3))/4/q_2 \tag{54}$$

case 3:

$$q_3 = 0.5 * \sqrt{max(T_i)} \tag{55}$$
$$q_1 = (\mathbf{R}_b^n(1,3) - \mathbf{R}_b^n(3,1))/4/q_3 \tag{56}$$
$$q_2 = (\mathbf{R}_b^n(1,2) + \mathbf{R}_b^n(2,1))/4/q_3 \tag{57}$$
$$q_4 = (\mathbf{R}_b^n(2,3) + \mathbf{R}_b^n(3,2))/4/q_3 \tag{58}$$

case 4:

$$q_4 = 0.5 * \sqrt{max(T_i)} \tag{59}$$
$$q_1 = (\mathbf{R}_b^n(2,1) - \mathbf{R}_b^n(1,2))/4/q_4 \tag{60}$$
$$q_2 = (\mathbf{R}_b^n(1,3) + \mathbf{R}_b^n(3,1))/4/q_4 \tag{61}$$
$$q_3 = (\mathbf{R}_b^n(2,3) + \mathbf{R}_b^n(3,2))/4/q_4 \tag{62}$$

Note: this result is the unit quaternion.

IV. Linear Algebra

The following subsections contain useful relations from Linear Algebra.

A. Vector Properties

First define a vector space $\mathcal{V}$, where $\mathcal{V}_0$ is a vector subspace of $\mathcal{V}$, *iff*:

- $\mathcal{V}_0$ is non-empty
- $\mathcal{V}_0$ is closed under multiplication
- $\mathcal{V}_0$ is closed under addition

A basis of subspace $\mathcal{V}_0$ is a linearly-independent spanning set of $\mathcal{V}$.

The following properties of a vector space hold $\forall \mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathcal{V}$ and $a, b \in \mathbb{R}$:

$$\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w} \tag{63}$$
$$\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u} \tag{64}$$
$$\exists \mathbf{0} \in \mathcal{V} : \mathbf{v} + \mathbf{0} = \mathbf{v},\ \forall \mathbf{v} \in \mathcal{V} \tag{65}$$
$$\forall \mathbf{v} \in \mathcal{V},\ \exists(-\mathbf{v}) \in \mathcal{V} : \mathbf{v} + (-\mathbf{v}) = \mathbf{0} \tag{66}$$
$$(ab)\mathbf{v} = a(b\mathbf{v}) \tag{67}$$
$$1\mathbf{v} = \mathbf{v} \tag{68}$$
$$a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v} \tag{69}$$
$$(a + b)\mathbf{v} = a\mathbf{v} + a\mathbf{v} \tag{70}$$

B. Products of Vectors

The *scalar product* (also called the inner product or dot product) of two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^{m \times 1}$ is defined as

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^\intercal \cdot \mathbf{v} = \sum_{i=1}^{m} \mathbf{u}_i \mathbf{v}_i \tag{71}$$

The Euclidean norm of a vector is the scalar product of a vector with itself

$$\mathbf{u} \cdot \mathbf{u} = ||\mathbf{u}||^2 \tag{72}$$

the multiplicative property is

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{v} \cdot \mathbf{u} = \mathbf{u}^\intercal \cdot \mathbf{v} = \mathbf{v}^\intercal \cdot \mathbf{u} \neq \mathbf{u} \cdot \mathbf{v}^\intercal \tag{73}$$

the associative property is

$$\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w} \tag{74}$$

The vector $\mathbf{u}$ is a *unit vector* if $\|\mathbf{u}\| = 1$. A vector has both magnitude and direction, with magnitude defined by $\|\mathbf{u}\|$ and direction defined by $\frac{\mathbf{u}}{\|\mathbf{u}\|}$.

The angle α between two vectors $\mathbf{u}$ and $\mathbf{v}$ is defined as

$$\mathbf{u} \cdot \mathbf{v} = \|\mathbf{u}\| \|\mathbf{v}\| \cos(\alpha) \tag{75}$$

Two vectors are orthogonal if and only if $\mathbf{u} \cdot \mathbf{v} = 0$.

A function $\mathcal{V} \mapsto \mathbb{R}$ is a norm *iff*, $\forall \mathbf{v} \in \mathcal{V}$, the all three properties must be satisfied

$$1.\ \|\mathbf{v}\| \geq 0 \text{ and},\ \|\mathbf{v}\| = 0 \iff \mathbf{v} = \mathbf{0} \tag{76}$$
$$2.\ \|c\mathbf{v}\| = |c| \cdot \|\mathbf{v}\|,\ \ c \in \mathbb{R} \tag{77}$$
$$3.\ \|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\| \tag{78}$$

Some useful vector norms are

$$\|\mathbf{u}\|_p = \left(\sum_{i=1}^{m} |u_i|^p \right)^{1 \backslash p},\ \ p \geq 1 \tag{79}$$
$$\|\mathbf{u}\|_1 = \sum_{i=1}^{m} |u_i| \tag{80}$$
$$\|\mathbf{u}\|_2 = \sqrt{\mathbf{u}^\intercal \mathbf{u}} \tag{81}$$
$$\|\mathbf{u}\|_\infty = \max_i \{|u_i|\} \tag{82}$$

The Cauchy-Schwartz Inequality for $\mathbf{u}, \mathbf{v} \in \mathbb{R}^{m \times 1}$:

$$|\mathbf{u}^\intercal \mathbf{v}| \leq \|\mathbf{u}\|_2 \cdot \|\mathbf{v}\|_2 \tag{83}$$

The *vector product* (also called the outer-product) of two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^{3 \times 1}$ is defined as

$$\mathbf{u} \times \mathbf{v} = \begin{vmatrix} i & j & k \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} \tag{84}$$
$$= \begin{bmatrix} u_2 v_3 - u_3 v_2 \\ u_3 v_1 - u_1 v_3 \\ u_1 v_2 - u_2 v_1 \end{bmatrix} \tag{85}$$
$$= \begin{bmatrix} 0 & -u_3 & u_2 \\ u_3 & 0 & -u_1 \\ -u_2 & u_1 & 0 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \tag{86}$$

where $| \cdot |$ denotes a determinant, and i, j, and k are unit vectors pointing along the principal axes of the reference frame. Notice $\mathbf{u} \times \mathbf{u} = 0$ and $\mathbf{u} \times \mathbf{v} = -\mathbf{v} \times \mathbf{u}$.

It is sometimes convenient to express the vector product of eqn. (84) in the matrix form

$$\mathbf{u} \times \mathbf{v} = \mathbf{U}\mathbf{v} \tag{87}$$

where

$$\mathbf{U} = \begin{bmatrix} 0 & -u_3 & u_2 \\ u_3 & 0 & -u_1 \\ -u_2 & u_1 & 0 \end{bmatrix} \tag{88}$$

This skew symmetric form of $\mathbf{u}$ is convenient in analysis. The matrix $\mathbf{U}$ will be denoted $\mathbf{U} = [\mathbf{u}\times]$. The vector product has the following properties:

$$\mathbf{u} \times \mathbf{u} = \mathbf{0} \tag{89}$$
$$\mathbf{u} \times \mathbf{v} = -\mathbf{v} \times \mathbf{u} \tag{90}$$
$$\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) \neq (\mathbf{u} \times \mathbf{v}) \times \mathbf{w} \tag{91}$$
$$\mathbf{u} \times (\mathbf{v} \times \mathbf{w}) = (\mathbf{u} \cdot \mathbf{w})\mathbf{v} - (\mathbf{u} \cdot \mathbf{v})\mathbf{w} \tag{92}$$
$$\mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) = \mathbf{v} \cdot (\mathbf{w} \times \mathbf{u}) \tag{93}$$
$$= \mathbf{w} \cdot (\mathbf{u} \times \mathbf{v}) \tag{94}$$
$$(\mathbf{u} \times \mathbf{v}) \cdot (\mathbf{w} \times \mathbf{z}) = (\mathbf{u} \cdot \mathbf{w})(\mathbf{v} \cdot \mathbf{z}) - (\mathbf{v} \cdot \mathbf{w})(\mathbf{u} \cdot \mathbf{z}). \tag{95}$$

C. Products of Matrices

Matrix multiplication has several useful formulas, where $\mathbf{A}, \mathbf{B}, \mathbf{C} \in \mathbb{R}^{m \times m}$. The associative property is

$$(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC}). \tag{96}$$

The distributive property is

$$(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC} \tag{97}$$
$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1} \tag{98}$$
$$(\mathbf{ABC}\ldots)^{-1} = \ldots \mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1} \tag{99}$$
$$(\mathbf{A}^\intercal)^{-1} = (\mathbf{A}^{-1})^\intercal \tag{100}$$
$$(\mathbf{A} + \mathbf{B})^\intercal = \mathbf{A}^\intercal + \mathbf{B}^\intercal \tag{101}$$
$$(\mathbf{AB})^\intercal = \mathbf{B}^\intercal\mathbf{A}^\intercal \tag{102}$$
$$(\mathbf{ABC}\ldots)^\intercal = \ldots \mathbf{C}^\intercal\mathbf{B}^\intercal\mathbf{A}^\intercal \tag{103}$$
$$(\mathbf{A}^H)^{-1} = (\mathbf{A}^{-1})^H \tag{104}$$
$$(\mathbf{A} + \mathbf{B})^H = \mathbf{A}^H + \mathbf{B}^H \tag{105}$$
$$(\mathbf{AB})^H = \mathbf{B}^H\mathbf{A}^H \tag{106}$$
$$(\mathbf{ABC}\ldots)^H = \ldots \mathbf{C}^H\mathbf{B}^H\mathbf{A}^H. \tag{107}$$

Matrix multiplication does not commute: $\mathbf{AB} \neq \mathbf{BA}$. Matrix division is undefined. When $\mathbf{A} \in \mathbb{R}^{n \times m}$ and $\mathbf{B} \in \mathbb{R}^{m \times p}$, then their product $\mathbf{C} = \mathbf{AB} \in \mathbb{R}^{n \times p}$ is defined as $c_{i,j} = \sum_{k=1}^{m} a_{i,k}b_{k,j}$. Note that the number of columns in $\mathbf{A}$ and the number of rows in $\mathbf{B}$, referred to as the inner dimension of the product, must be identical; otherwise the two matrices cannot be multiplied.

A square matrix $\mathbf{A}$ is an orthogonal matrix if and only if $\mathbf{AA}^\intercal = \mathbf{A}^\intercal\mathbf{A}$ is a diagonal matrix. A square matrix $\mathbf{A}$ is an orthonormal matrix if and only if $\mathbf{AA}^\intercal = \mathbf{A}^\intercal\mathbf{A} = \mathbf{I}$. Let $\mathbf{R}_a^b$ represent the vector transformation from reference frame a to reference frame b. The inverse transformation from reference frame b to reference frame a can be shown to be $\mathbf{R}_b^a = (\mathbf{R}_a^b)^{-1} = (\mathbf{R}_a^b)^\intercal$. Then $(\mathbf{R}_a^b)^\intercal\mathbf{R}_a^b = \mathbf{R}_a^b(\mathbf{R}_a^b)^\intercal = \mathbf{I}$. This shows that $\mathbf{R}_a^b$ and $\mathbf{R}_b^a$ are an orthonormal matrix.

D. Matrix Norms

A matrix norm is a mapping which fulfills 1-3

$$1.\ \|\mathbf{A}\| \geq 0 \text{ and, } \|\mathbf{A}\| = 0 \iff \mathbf{A} = \mathbf{0} \tag{108}$$
$$2.\ \|c\mathbf{A}\| = |c| \cdot \|\mathbf{A}\|,\ c \in \mathbb{R} \tag{109}$$
$$3.\ \|\mathbf{A} + \mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\| \tag{110}$$

An induced norm is a matrix norm induced by a vector norm

$$\|\mathbf{A}\| = \sup\{\|\mathbf{A}\boldsymbol{x}\| \mid \|\boldsymbol{x}\| = 1\} \tag{111}$$

where $\|\cdot\|$ on the left side is the induced matrix norm, while $\|\cdot\|$ on the right side denotes the vector norm. For induced norms it holds that

$$1.\ \|\mathbf{I}\| = 1 \tag{112}$$
$$2.\ \|\mathbf{A}\boldsymbol{x}\|_p \leq \|\mathbf{A}\|_p \cdot \|\boldsymbol{x}\|_p,\ \forall \mathbf{A}, \boldsymbol{x} \tag{113}$$
$$3.\ \|\mathbf{AB}\|_p \leq \|\mathbf{A}\|_p \cdot \|\mathbf{B}\|_p,\ \forall \mathbf{A}, \mathbf{B} \tag{114}$$

where eq. (114) is known as the Submultiplicative Property of Induced Norms. Induced norms have the following properties:

$$\|\mathbf{A}\|_p = \max_{x \neq 0} \frac{\|\mathbf{A}\boldsymbol{x}\|_p}{\|\boldsymbol{x}\|_p} \tag{115}$$
$$= \max_{x \neq 0} \left\| \mathbf{A} \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_p} \right\|_p \tag{116}$$
$$= \max_{\|\mathbf{y}\|_p = 1} \|\mathbf{A}\mathbf{y}\|_p \tag{117}$$

Some useful matrix norms are

$$\|\mathbf{A}\|_p = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^p \right)^{1\backslash p} \tag{118}$$
$$\|\mathbf{A}\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^{m} |a_{ij}| \tag{119}$$
$$\|\mathbf{A}\|_2 = \sqrt{\lambda_{\text{max}}(\mathbf{A}^\intercal\mathbf{A})} \tag{120}$$
$$\|\mathbf{A}\|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^{n} |a_{ij}| \tag{121}$$
$$\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2} = \sqrt{\text{Tr}(\mathbf{AA}^H)} \tag{122}$$
$$\|\mathbf{A}\|_{\text{max}} = \max_{ij} |a_{ij}| \tag{123}$$

where λ_{max} is the maximum eigenvalue of $(\mathbf{A}^\intercal\mathbf{A})$.

E. Condition Number

The condition number is a measure of how sensitive a function is to errors in the input. A problem with a low condition number is said to be well-conditioned, while a problem with a high condition number is said to be ill-conditioned.

The condition number, $\kappa(\mathbf{A})$, for the general matrix $\mathbf{A}$, is

$$\kappa(\mathbf{A}) = \frac{\sigma_{\text{max}}(\mathbf{A})}{\sigma_{\text{min}}(\mathbf{A})} \tag{124}$$

where $\sigma_{\text{max}}(\mathbf{A})$ and $\sigma_{\text{min}}(\mathbf{A})$ are the maximum and minimum singular values of $\mathbf{A}$.

If $\mathbf{A}$ is a symmetric, positive definite matrix,

$$\kappa(\mathbf{A}) = \frac{|\lambda_{\text{max}}(\mathbf{A})|}{|\lambda_{\text{min}}(\mathbf{A})|} \tag{125}$$

where $\lambda_{\text{max}}(\mathbf{A})$ and $\lambda_{\text{min}}(\mathbf{A})$ are the maximum and minimum eigenvalues of $\mathbf{A}$.

If $\mathbf{A}$ is unitary, then

$$\kappa(\mathbf{A}) = 1 \tag{126}$$

If the condition number is ϵ (some small value) larger than one, the matrix is well conditioned, and therefore its inverse can be computed with good accuracy.

F. Trace Operator

The trace of a square matrix $\mathbf{A} \in \mathbb{R}^{m \times m}$ is defined as

$$\text{Tr}(\mathbf{A}) = \sum_{i=1}^{m} a_{ii} \tag{127}$$

Notice:

$$\text{Tr}(\mathbf{A}) = \text{Tr}(\mathbf{A}^\intercal) \tag{128}$$
$$\text{Tr}(\mathbf{A}+\mathbf{B}) = \text{Tr}(\mathbf{A}) + \text{Tr}(\mathbf{B}) \tag{129}$$
$$\text{Tr}(\mathbf{AB}) = \text{Tr}(\mathbf{BA}) \tag{130}$$
$$\text{Tr}(\mathbf{ABC}) = \text{Tr}(\mathbf{BCA}) = \text{Tr}(\mathbf{CAB}) \tag{131}$$
$$\mathbf{v}^\intercal\mathbf{v} = \text{Tr}(\mathbf{v}\mathbf{v}^\intercal) \tag{132}$$
$$\text{Tr}(\mathbf{v}\mathbf{v}^\intercal) = \text{Tr}(\mathbf{v}^\intercal\mathbf{v}) = ||\mathbf{v}||^2 \tag{133}$$
$$\text{Tr}(\mathbf{A}) = \sum_{i=1}^{m} \lambda_i, \quad \lambda_i = \text{eig}(\mathbf{A}) \tag{134}$$

where $\mathbf{v} \in \mathbb{R}^{m \times 1}$, and the eigenvalues $\boldsymbol{\lambda} \in \mathbb{R}^{m \times 1}$.

G. Definiteness of Matrices

For the symmetric matrix $\mathbf{A} \in \mathbb{R}^{m \times m}$ and vector $\boldsymbol{x} \in \mathbb{R}^{m \times 1}$, the scalar mapping $\boldsymbol{x}^\intercal\mathbf{A}\boldsymbol{x}$ is called a *quadratic form*, and has the following properties:

1) $\mathbf{A}$ is positive definite *iff* $\boldsymbol{x}^\intercal\mathbf{A}\boldsymbol{x} > 0, \ \forall \boldsymbol{x} \neq 0$.
2) $\mathbf{A}$ is positive semi-definite *iff* $\boldsymbol{x}^\intercal\mathbf{A}\boldsymbol{x} \geq 0, \ \forall \boldsymbol{x} \neq 0$.
3) $\mathbf{A}$ is negative definite *iff* $\boldsymbol{x}^\intercal\mathbf{A}\boldsymbol{x} < 0, \ \forall \boldsymbol{x} \neq 0$.
4) $\mathbf{A}$ is negative semi-definite *iff* $\boldsymbol{x}^\intercal\mathbf{A}\boldsymbol{x} \leq 0, \ \forall \boldsymbol{x} \neq 0$.

where *iff* is defined as "if and only if". The designer is often interested in the sign of the scalar output, $\boldsymbol{x}^\intercal\mathbf{A}\boldsymbol{x}$. When the matrix $\mathbf{A}$ fails to have any of the above properties, then the matrix $\mathbf{A}$ is sign indefinite.

H. Rank of Matrices

The rank of a matrix follows the following rules:
For $\mathbf{A} \in \mathbb{R}^{p \times q}$

$$\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}^\intercal) \tag{135}$$
$$\text{rank}(\mathbf{A}^\intercal\mathbf{A}) = \text{rank}(\mathbf{A}^\intercal\mathbf{A}^\intercal) = \text{rank}(\mathbf{A}) \tag{136}$$
$$\text{rank}(\mathbf{AB}) \leq \text{rank}(\mathbf{A}) \tag{137}$$
$$\text{rank}(\mathbf{AB}) \leq \text{rank}(\mathbf{B}) \tag{138}$$
$$\text{rank}(\mathbf{A}+\mathbf{B}) \leq \text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B}) \tag{139}$$
$$\text{rank}(\mathbf{A}) = \min\{p, q\} \tag{140}$$

Sylvester's inequality: for $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{B} \in \mathbb{R}^{n \times r}$

$$\text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B}) - n \leq \text{rank}(\mathbf{AB}) \tag{141}$$
$$\leq \min\{\text{rank}(\mathbf{A}), \text{rank}(\mathbf{B})\} \tag{142}$$

and

$$|\text{rank}(\mathbf{A}) - \text{rank}(\mathbf{B})| \leq \text{rank}(\mathbf{A}+\mathbf{B}) \tag{143}$$
$$\leq \text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B}) \tag{144}$$

For full-rank $\mathbf{A}$

$$\mathbf{A} \in \mathbb{R}^{n \times n}, \quad \text{rank}(\mathbf{A}) = n \tag{145}$$
$$\mathbf{A} \in \mathbb{R}^{n \times m}, \quad \text{rank}(\mathbf{A}) = m, \quad m > n \tag{146}$$
$$\mathbf{A} \in \mathbb{R}^{m \times n}, \quad \text{rank}(\mathbf{A}) = n, \quad m > n \tag{147}$$

Elementary transformations do not change rank, e.g. swap rows or columns, addition operations on rows, multiplication operations by non-zero values for rows or columns.

I. Independence and Determinants

Given vectors $\mathbf{u}_i \in \mathbb{R}^n$ for $i = 1, \cdots, m$, the vectors are linearly dependent if there exists $\alpha_i \in \mathbb{R}$, not all zero, such that $\sum_{i=1}^{m} \alpha_i \mathbf{u}_i = 0$. If the summation is only zero when all $\alpha_i = 0$, then the set of vectors is linearly independent. A set of $m > n$ vectors in $\mathbb{R}^n$ is always linearly dependent.

A set of n vectors $\mathbf{u}_i \in \mathbb{R}^n$, $i = 1, \cdots, n$ can be arranged as a square matrix $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n] \in \mathbb{R}^{n \times n}$. Note that a matrix can always be decomposed into its component column (or row) vectors. The rank of a matrix, denoted rank($\mathbf{U}$), is the number of independent column (or row) vectors in the matrix. If this set of vectors is linearly independent (i.e., rank($\mathbf{U}$) $= n$), then we say that the matrix $\mathbf{U}$ is nonsingular (or of full rank).

A convenient tool for checking whether a square matrix is nonsingular is the determinant. The determinant of $\mathbf{A} \in \mathbb{R}^{n \times n}$, denoted $|\mathbf{A}|$ or det($\mathbf{A}$), is a scalar real number that can be computed either as

$$|\mathbf{A}| = \sum_{j=1}^{n} a_{kj} c_{kj} \tag{148}$$

which uses row expansion; or,

$$|\mathbf{A}| = \sum_{k=1}^{n} a_{kj} c_{kj} \tag{149}$$

which uses column expansion. For a square matrix $\mathbf{A}$, the *cofactor* associated with a_{kj} is

$$c_{kj} = (1)^{k+j} M_{kj} \tag{150}$$

where M_{kj} is the *minor* associated with a_{kj}. M_{kj} is defined as the determinant of the matrix formed by dropping the k-th row and j-th column from $\mathbf{A}$. The determinant of $\mathbf{A} \in \mathbb{R}^{1 \times 1}$ is (the scalar) $\mathbf{A}$. The determinant of $\mathbf{A} \in \mathbb{R}^{n \times n}$ is written in terms of the summation of the determinants of matrices in $\mathbb{R}^{(n-1) \times (n-1)}$. This dimension reduction process continues until it involves only scalars, for which the computation is straightforward.

If $|\mathbf{A}| \neq 0$, then $\mathbf{A}$ is nonsingular and the vectors forming the rows (and columns) of $\mathbf{A}$ are linearly independent. If $|\mathbf{A}| = 0$, then $\mathbf{A}$ is singular and the vectors forming the rows (and columns) of $\mathbf{A}$ are linearly dependent.

Determinants have the following useful properties: For $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n\times n}$

1) $|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}|$
2) $|\mathbf{A}| = |\mathbf{A}^\intercal|$
3) If any row or column of $\mathbf{A}$ is entirely zero, then $|\mathbf{A}| = 0$.
4) If any two rows (or columns) of $\mathbf{A}$ are linearly dependent, then $|\mathbf{A}| = 0$.
5) Interchanging two rows (or two columns) of $\mathbf{A}$ reverses the sign of the determinant.
6) Multiplication of a row of $\mathbf{A}$ by $\alpha \in \mathbb{R}$, yields $\alpha|\mathbf{A}|$.
7) A scaled version of one row can be added to another row without changing the determinant.

A summary of the properties are

$$\det(\mathbf{A}) = \prod_i \lambda_i, \quad \lambda_i = \text{eig}(\mathbf{A}) \tag{151}$$
$$\det(c\mathbf{A}) = c^n \det(\mathbf{A}), \quad \text{if } \mathbf{A} \in \mathbb{R}^{n\times n} \tag{152}$$
$$\det(\mathbf{A}^\intercal) = \det(\mathbf{A}) \tag{153}$$
$$\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B}) \tag{154}$$
$$\det(\mathbf{A}^{-1}) = 1/\det(\mathbf{A}) \tag{155}$$
$$\det(\mathbf{A}^n) = \det(\mathbf{A})^n \tag{156}$$
$$\det(\mathbf{I} + \mathbf{u}\mathbf{v}^\intercal) = 1 + \mathbf{u}^\intercal\mathbf{v}, \quad \text{for } \mathbf{u}, \mathbf{v} \in \mathbb{R}^{n\times 1} \tag{157}$$

For $n = 2$:

$$\det(\mathbf{I} + \mathbf{A}) = 1 + \det(\mathbf{A}) + \text{Tr}(\mathbf{A}) \tag{158}$$

For $n = 3$:

$$\det(\mathbf{I} + \mathbf{A}) = 1 + \det(\mathbf{A}) + \text{Tr}(\mathbf{A}) + \frac{1}{2}\text{Tr}(\mathbf{A})^2 - \frac{1}{2}\text{Tr}(\mathbf{A}^2) \tag{159}$$

J. Matrix Inversion

For $\mathbf{A} \in \mathbb{R}^{n\times n}$, with $|\mathbf{A}| \neq 0$, we denote the inverse of $\mathbf{A}$ by $\mathbf{A}^{-1}$ which has the property that

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}. \tag{160}$$

When $\mathbf{A}$ is a square nonsingular matrix,

$$\mathbf{A}^{-1} = \frac{\mathbf{C}^\intercal}{|\mathbf{A}|} \tag{161}$$

where $\mathbf{C}$ is the cofactor matrix for $\mathbf{A}$ and $\mathbf{C}^\intercal$ is called the adjoint of $\mathbf{A}$. The matrix inverse has the following properties:

$$(\mathbf{A}^{-1})^{-1} = \mathbf{A} \tag{162}$$
$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1} \tag{163}$$
$$|\mathbf{A}^{-1}| = \frac{1}{|\mathbf{A}|} \tag{164}$$
$$(\mathbf{A}^\intercal)^{-1} = (\mathbf{A}^{-1})^\intercal = \mathbf{A}^{-\intercal} \tag{165}$$
$$(\mu\mathbf{A})^{-1} = \frac{1}{\mu}\mathbf{A}^{-1}. \tag{166}$$

The inverse of an orthonormal matrix is the same as its transpose.

K. Matrix Inversion Lemma

Two forms of the Matrix Inversion Lemma are presented. The Lemma is useful in least squares and Kalman filter derivations. Each lemma can be proved by direct multiplication.

Lemma 4.1: Given four matrices $\mathbf{P}_1$, $\mathbf{P}_2$, $\mathbf{H}$, and $\mathbf{R}$ of compatible dimensions, if $\mathbf{P}_1$, $\mathbf{P}_2$, $\mathbf{R}$, and $(\mathbf{H}^\intercal\mathbf{P}_1\mathbf{H} + \mathbf{R})$ are all invertible and

$$\mathbf{P}_2^{-1} = \mathbf{P}_1^{-1} + \mathbf{H}\mathbf{R}^{-1}\mathbf{H}^\intercal \tag{167}$$

then

$$\mathbf{P}_2 = \mathbf{P}_1 - \mathbf{P}_1\mathbf{H}(\mathbf{H}^\intercal\mathbf{P}_1\mathbf{H} + \mathbf{R})^{-1}\mathbf{H}^\intercal\mathbf{P}_1. \tag{168}$$

Lemma 4.2: Given four matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ of compatible dimensions, if $\mathbf{A}$, $\mathbf{C}$, and $\mathbf{A} + \mathbf{BCD}$ are invertible, then

$$(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{D}\mathbf{A}^{-1}\mathbf{B} + \mathbf{C}^{-1})\mathbf{D}\mathbf{A}^{-1} \tag{169}$$

The equivalence of the two forms is shown by defining:$\mathbf{A} = \mathbf{P}_1^{-1}$, $\mathbf{B} = \mathbf{H}$, $\mathbf{C} = \mathbf{R}^{-1}$, $\mathbf{D} = \mathbf{H}^\intercal$, and requiring $\mathbf{A} + \mathbf{BCD} = \mathbf{P}_2^{-1}$.

L. Eigenvalues and Eigenvectors

For $\mathbf{A} \in \mathbb{R}^{n\times n}$, the set of scalars $\lambda_i \in \mathrm{C}$ and (nonzero) vectors $\boldsymbol{x}_i \in \mathrm{C}^n$ satisfying $\mathbf{A}\boldsymbol{x}_i = \lambda_i\boldsymbol{x}_i$ or $(\lambda_i\mathbf{I}\mathbf{A})\boldsymbol{x}_i = \mathbf{0}_n$ are the eigenvalues and eigenvectors of $\mathbf{A}$.

We are only interested in nontrivial solutions (i.e., solution $\boldsymbol{x}_i = 0$ is not of interest). Nontrivial solutions exist only if $(\lambda\mathbf{I} - \mathbf{A})$ is a singular matrix. Therefore, the eigenvalues of $\mathbf{A}$ are the values of λ such that $|\lambda\mathbf{I} - \mathbf{A}| = 0$. This yields an n-th order polynomial in λ.

If $\mathbf{A}$ is a symmetric matrix, then all of its eigenvalues and eigenvectors are real. If $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are eigenvectors of symmetric matrix $\mathbf{A}$ and their eigenvalues are not equal (i.e., $\lambda_i \neq \lambda_j$), then the eigenvectors are orthogonal (i.e. $\boldsymbol{x}_i \cdot \boldsymbol{x}_j = 0$).

A square matrix $\mathbf{A}$ is *idempotent* if and only if $\mathbf{AA} = \mathbf{A}$. Idempotent matrices are sometimes also called *projection* matrices. Idempotent matrices have the following properties:

1) $\text{rank}(\mathbf{A}) = \text{Tr}(\mathbf{A})$;
2) the eigenvalues of $\mathbf{A}$ are all either 0 or 1;
3) the multiplicity of 1 as an eigenvalue is the $\text{rank}(\mathbf{A})$;
4) $\mathbf{A}(\mathbf{I} - \mathbf{A}) = (\mathbf{I} - \mathbf{A})\mathbf{A} = 0$; and,
5) $\mathbf{A}^\intercal$, $(\mathbf{I} - \mathbf{A})$ and $(\mathbf{I} - \mathbf{A}^\intercal)$ are idempotent.

M. The Special Case 2x2

Consider the matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}, \tag{170}$$

the determinant and trace are

$$\det(\mathbf{A}) = \mathbf{A}_{11}\mathbf{A}_{22} - \mathbf{A}_{12}\mathbf{A}_{21} \tag{171}$$
$$\text{Tr}(\mathbf{A}) = \mathbf{A}_{11} + \mathbf{A}_{22}, \tag{172}$$

the eigenvalues are

$$\lambda^2 - \lambda \cdot \mathrm{Tr}(\mathbf{A}) + \det(\mathbf{A}) = 0 \tag{173}$$
$$\lambda_1 = \frac{\mathrm{Tr}(\mathbf{A}) + \sqrt{\mathrm{Tr}(\mathbf{A})^2 - 4\det(\mathbf{A})}}{2} \tag{174}$$
$$\lambda_2 = \frac{\mathrm{Tr}(\mathbf{A}) + \sqrt{\mathrm{Tr}(\mathbf{A})^2 - 4\det(\mathbf{A})}}{2} \tag{175}$$
$$\lambda_1 + \lambda_2 = \mathrm{Tr}(\mathbf{A}) \tag{176}$$
$$\lambda_1\lambda_2 = \det(\mathbf{A}), \tag{177}$$

the eigenvectors are

$$\mathbf{v}_1 \propto \begin{bmatrix} \mathbf{A}_{12} \\ \lambda_1 \mathbf{A}_{11} \end{bmatrix} \tag{178}$$
$$\mathbf{v}_2 \propto \begin{bmatrix} \mathbf{A}_{12} \\ \lambda_2 \mathbf{A}_{11} \end{bmatrix}, \tag{179}$$

and the inverse is

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})} \begin{bmatrix} \mathbf{A}_{22} & -\mathbf{A}_{12} \\ -\mathbf{A}_{21} & \mathbf{A}_{11} \end{bmatrix}. \tag{180}$$

N. Matrix Exponential

The Matrix Exponential may be computed by many methods; provided here are two such examples.

1) Power Series: The power series expansion of the scalar exponential function is

$$e^{at} = \sum_{n=0}^{\infty} \frac{1}{n!}(at)^n = 1 + at + \frac{(at)^2}{2!} + \frac{(at)^3}{3!} + \cdots \tag{181}$$

for $a, t \in \mathbb{R}$. Extension of this power series to matrix arguments serves as a definition of the matrix exponential

$$e^{\mathbf{A}} \equiv \sum_{n=0}^{\infty} \frac{1}{n!}\mathbf{A}^n = \mathbf{I} + \mathbf{A} + \frac{\mathbf{A}^2}{2!} + \frac{\mathbf{A}^3}{3!} + \cdots \tag{182}$$
$$e^{-\mathbf{A}} \equiv \sum_{n=0}^{\infty} \frac{1}{n!}(-1)^n\mathbf{A}^n = \mathbf{I} - \mathbf{A} + \frac{\mathbf{A}^2}{2!} - \frac{\mathbf{A}^3}{3!} + \cdots \tag{183}$$
$$e^{\mathbf{A}t} \equiv \sum_{n=0}^{\infty} \frac{1}{n!}(\mathbf{A}t)^n = \mathbf{I} + \mathbf{A}t + \frac{(\mathbf{A}t)^2}{2!} + \frac{(\mathbf{A}t)^3}{3!} + \cdots \tag{184}$$
$$\ln(\mathbf{I} + \mathbf{A}) \equiv \sum_{n=0}^{\infty} \frac{(-1)^{n-1}}{n}\mathbf{A}^n = \mathbf{A} - \frac{\mathbf{A}^2}{2} + \frac{\mathbf{A}^3}{3} + \cdots \tag{185}$$

where $t \in \mathbb{R}$, $\mathbf{A} \in \mathbb{R}^{n\times n}$, and $\mathbf{I}$ is the identity matrix in $\mathbb{R}^{n\times n}$. Note that by the definition of the matrix exponential, it is always true that a matrix commutes with its exponential:

$$\mathbf{A}e^{\mathbf{A}t} = e^{\mathbf{A}t}\mathbf{A}. \tag{186}$$

Also, $e^{\mathbf{A}t}|_{t=0} = \mathbf{I}$.

For $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n\times n}$, some useful properties of the exponential are

$$e^{\mathbf{A}}e^{\mathbf{B}} = e^{\mathbf{A}+\mathbf{B}}, \quad \text{if } \mathbf{AB} = \mathbf{BA} \tag{187}$$
$$(e^{\mathbf{A}})^{-1} = e^{-\mathbf{A}} \tag{188}$$
$$\frac{d}{dt}e^{t\mathbf{A}} = \mathbf{A}e^{t\mathbf{A}} = e^{t\mathbf{A}}\mathbf{A}, \; t \in \mathbb{R} \tag{189}$$
$$\frac{d}{dt}\mathrm{Tr}(e^{t\mathbf{A}}) = \mathrm{Tr}(\mathbf{A}e^{t\mathbf{A}}) \tag{190}$$
$$\det(e^{\mathbf{A}}) = e^{\mathrm{Tr}(\mathbf{A})} \tag{191}$$

Some useful trigonometric functions are

$$\sin(\mathbf{A}) \equiv \sum_{n=0}^{\infty} \frac{(-1)^n\mathbf{A}^{2n+1}}{(2n+1)!} = \mathbf{A} - \frac{\mathbf{A}^3}{3!} + \frac{\mathbf{A}^5}{5!} - \cdots \tag{192}$$
$$\cos(\mathbf{A}) \equiv \sum_{n=0}^{\infty} \frac{(-1)^n\mathbf{A}^{2n}}{(2n)!} = \mathbf{I} - \frac{\mathbf{A}^2}{2!} + \frac{\mathbf{A}^4}{4!} - \cdots \tag{193}$$

Power series expansion is usually not the best numeric technique for the computation of matrix exponentials; however when the structure of the $\mathbf{A}$ matrix is appropriate, power series methods are one approach for determining closed form solutions for the matrix exponential of $\mathbf{A}$.

2) Laplace Transform: The formula

$$\mathbf{F}e^{\mathbf{A}t} = \mathcal{L}^{-1}\{(s\mathbf{I} - \mathbf{A})^{-1}\} \tag{194}$$

is derived by taking the Laplace transform of both sides of eqn. (182)

$$\mathcal{L}\{e^{\mathbf{A}t}\} = \mathcal{L}\{\mathbf{I} + \mathbf{A}t + \frac{(\mathbf{A}t)^2}{2!} + \frac{(\mathbf{A}t)^3}{3!} + \cdots\} \tag{195}$$
$$= \frac{1}{s}\mathbf{I} + \frac{1}{s^2}\mathbf{A} + \frac{1}{s^3}\mathbf{A}^2 + \frac{1}{s^4}\mathbf{A}^3 + \cdots \tag{196}$$
$$= (s\mathbf{I} - \mathbf{A})^{-1} \tag{197}$$
$$\mathbf{A}e^{\mathbf{A}t} = \mathcal{L}^{-1}\{(s\mathbf{I} - \mathbf{A})^{-1}\}. \tag{198}$$

This derivation has used the fact that

$$(s\mathbf{I} - \mathbf{A})^{-1} = \left(\frac{1}{s}\mathbf{I} + \frac{1}{s^2}\mathbf{A} + \frac{1}{s^3}\mathbf{A}^2 + \frac{1}{s^4}\mathbf{A}^3 + \cdots\right) \tag{199}$$

which can be shown by direct multiplication of both sides of the equation by $(s\mathbf{I} - \mathbf{A})$.

O. Matrix Calculus

This section covers differentiation of a number of expressions with respect to a matrix $\mathbf{X}$. Note that it is always assumed that $\mathbf{X}$ has no special structure, i.e. that the elements of $\mathbf{X}$ are independent (e.g. not symmetric, Toeplitz, positive definite). See Section ?? for differentiation of structured matrices. The basic assumptions can be written as

$$\frac{\partial \mathbf{X}_{kl}}{\partial \mathbf{X}_{ij}} = \delta_{ik}\delta_{lk}$$

for vector forms,

$$\left[\frac{\partial \boldsymbol{x}}{\partial y}\right]_i = \frac{\partial x_i}{\partial y} \qquad \left[\frac{\partial x}{\partial \mathbf{y}}\right]_i = \frac{\partial x}{\partial y_i} \qquad \left[\frac{\partial \boldsymbol{x}}{\partial \mathbf{y}}\right]_{ij} = \frac{\partial x_i}{\partial y_j}$$

The following rules are general and very useful when deriving the differential of an expression:

$$\begin{aligned}
\partial \mathbf{A} &= 0 \qquad (\mathbf{A} \text{ is a constant}) \\
\partial(\alpha \mathbf{X}) &= \alpha \partial \mathbf{X} \\
\partial(\mathbf{X}+\mathbf{Y}) &= \partial \mathbf{X} + \partial \mathbf{Y} \\
\partial(\mathrm{Tr}(\mathbf{X})) &= \mathrm{Tr}(\partial \mathbf{X}) \\
\partial(\mathbf{X}\mathbf{Y}) &= (\partial \mathbf{X})\mathbf{Y} + \mathbf{X}(\partial \mathbf{Y}) \\
\partial(\mathbf{X} \circ \mathbf{Y}) &= (\partial \mathbf{X}) \circ \mathbf{Y} + \mathbf{X} \circ (\partial \mathbf{Y}) \\
\partial(\mathbf{X} \otimes \mathbf{Y}) &= (\partial \mathbf{X}) \otimes \mathbf{Y} + \mathbf{X} \otimes (\partial \mathbf{Y}) \\
\partial(\mathbf{X}^{-1}) &= -\mathbf{X}(\partial \mathbf{X})\mathbf{X}^{-1} \\
\partial(\det(\mathbf{X})) &= \mathrm{Tr}(\mathrm{adj}(\mathbf{X})\partial \mathbf{X}) \\
\partial(\det(\mathbf{X})) &= \det(\mathbf{X})\mathrm{Tr}(\mathbf{X}^{-1}\partial \mathbf{X}) \\
\partial(\ln(\det(\mathbf{X}))) &= \mathrm{Tr}(\mathbf{X}^{-1}\partial \mathbf{X}) \\
\partial \mathbf{X}^{\intercal} &= (\partial \mathbf{X})^{\intercal} \\
\partial \mathbf{X}^{\mathrm{H}} &= (\partial \mathbf{X})^{\mathrm{H}}
\end{aligned}$$

1) Derivatives of a Determinant:

General Form:

$$\begin{aligned}
\frac{\partial \det(\mathbf{Y})}{\partial x} &= \det(\mathbf{Y})\mathrm{Tr}\left[\mathbf{Y}^{-1}\frac{\partial \mathbf{Y}}{\partial x}\right] \\
\sum_k \frac{\partial \det(\mathbf{Y})}{\partial X_{ik}} X_{jk} &= \delta_{ij}\det(\mathbf{X}) \\
\frac{\partial^2 \det(\mathbf{Y})}{\partial x^2} &= \det(\mathbf{Y})\left\{ \mathrm{Tr}\left[\mathbf{Y}^{-1}\frac{\partial \frac{\partial \mathbf{Y}}{\partial x}}{\partial x}\right] \right. \\
&\quad + \mathrm{Tr}\left[\mathbf{Y}^{-1}\frac{\partial \mathbf{Y}}{\partial x}\right]\mathrm{Tr}\left[\mathbf{Y}^{-1}\frac{\partial \mathbf{Y}}{\partial x}\right] \\
&\quad \left. - \mathrm{Tr}\left[\left(\mathbf{Y}^{-1}\frac{\partial \mathbf{Y}}{\partial x}\right)\left(\mathbf{Y}^{-1}\frac{\partial \mathbf{Y}}{\partial x}\right)\right] \right\}
\end{aligned}$$

Linear Forms:

$$\begin{aligned}
\frac{\partial \det(\mathbf{X})}{\partial \mathbf{X}} &= \det(\mathbf{X})(\mathbf{X}^{-1})^{\intercal} \\
\sum_k \frac{\partial \det(\mathbf{X})}{\partial X_{ik}} X_{jk} &= \delta_{ij}\det(\mathbf{X}) \\
\frac{\partial \det(\mathbf{A}\mathbf{X}\mathbf{B})}{\partial \mathbf{X}} &= \det(\mathbf{A}\mathbf{X}\mathbf{B})(\mathbf{X}^{-1})^{\intercal} \\
&= \det(\mathbf{A}\mathbf{X}\mathbf{B})(\mathbf{X}^{\intercal})^{-1}
\end{aligned}$$

Square Forms: If $\mathbf{X}$ is square and invertible, then

$$\frac{\partial \det(\mathbf{X}^{\intercal}\mathbf{A}\mathbf{X})}{\partial \mathbf{X}} = 2\det(\mathbf{X}^{\intercal}\mathbf{A}\mathbf{X})\mathbf{X}^{-\intercal}$$

If $\mathbf{X}$ is not square but $\mathbf{A}$ is symmetric, then

$$\frac{\partial \det(\mathbf{X}^{\intercal}\mathbf{A}\mathbf{X})}{\partial \mathbf{X}} = 2\det(\mathbf{X}^{\intercal}\mathbf{A}\mathbf{X})\mathbf{A}\mathbf{X}(\mathbf{X}^{\intercal}\mathbf{A}\mathbf{X})^{-1}$$

If $\mathbf{X}$ is not square and $\mathbf{A}$ is not symmetric, then

$$\begin{aligned}
\frac{\partial \det(\mathbf{X}^{\intercal}\mathbf{A}\mathbf{X})}{\partial \mathbf{X}} &= 2\det(\mathbf{X}^{\intercal}\mathbf{A}\mathbf{X})\Big(\mathbf{A}\mathbf{X}(\mathbf{X}^{\intercal}\mathbf{A}\mathbf{X})^{-1} \\
&\quad + \mathbf{A}^{\intercal}\mathbf{X}(\mathbf{X}^{\intercal}\mathbf{A}^{\intercal}\mathbf{X})^{-1}\Big)
\end{aligned}$$

Other Nonlinear Forms: Some special cases are

$$\begin{aligned}
\frac{\partial \ln(\det(\mathbf{X}^{\intercal}\mathbf{X}))}{\partial \mathbf{X}} &= 2(\mathbf{X}^{+})^{\intercal} \\
\frac{\partial \ln(\det(\mathbf{X}^{\intercal}\mathbf{X}))}{\partial \mathbf{X}^{+}} &= -2\mathbf{X}^{\intercal} \\
\frac{\partial \ln(\det(\mathbf{X}))}{\partial \mathbf{X}} &= (\mathbf{X}^{-1})^{\intercal} = (\mathbf{X}^{\intercal})^{-1} \\
\frac{\partial \det(\mathbf{X}^{k})}{\partial \mathbf{X}} &= k \det(\mathbf{X}^{k})\mathbf{X}^{-\intercal}
\end{aligned}$$

2) Derivatives of an Inverse: The basic identity is

$$\frac{\partial \mathbf{Y}^{-1}}{\partial x} = \mathbf{Y}^{-1}\frac{\partial \mathbf{Y}}{\partial x}\mathbf{Y}^{-1}$$

from which it follows

$$\begin{aligned}
\frac{\partial (\mathbf{X}^{-1})_{kl}}{\partial X_{ij}} &= (\mathbf{X}^{-1})_{ki}(\mathbf{X}^{-1})_{jl} \\
\frac{\partial \mathbf{a}^{\intercal}\mathbf{X}^{-1}\mathbf{b}}{\partial \mathbf{X}} &= -\mathbf{X}^{-\intercal}\mathbf{a}\mathbf{b}^{\intercal}\mathbf{X}^{-\intercal} \\
\frac{\partial \det(\mathbf{X}^{-1})}{\partial \mathbf{X}} &= -\det(\mathbf{X}^{-1})(\mathbf{X}^{-1})^{\intercal} \\
\frac{\partial \mathrm{Tr}(\mathbf{A}\mathbf{X}^{-1}\mathbf{B})}{\partial \mathbf{X}} &= -(\mathbf{X}^{-1}\mathbf{B}\mathbf{A}\mathbf{X}^{-1})^{\intercal} \\
\frac{\partial \mathrm{Tr}((\mathbf{X}+\mathbf{A})^{-1})}{\partial \mathbf{X}} &= -((\mathbf{X}+\mathbf{A})^{-1}(\mathbf{X}+\mathbf{A})^{-1})^{\intercal}
\end{aligned}$$

We also have the following result: Let $\mathbf{A}$ be an $n \times n$ invertible square matrix, $\mathbf{W}$ be the inverse of $\mathbf{A}$, and $J(\mathbf{A})$ is an $n \times n$ -variate and differentiable function with respect to $\mathbf{A}$, then the partial differentials of J with respect to $\mathbf{A}$ and $\mathbf{W}$ satisfy

$$\frac{\partial J}{\partial \mathbf{A}} = -\mathbf{A}^{-\intercal}\frac{\partial J}{\partial \mathbf{W}}\mathbf{A}^{-\intercal}$$

3) Derivatives of Eigenvalues: The following hold,

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{x}}\sum \mathrm{eig}(\mathbf{X}) &= \frac{\partial}{\partial \boldsymbol{x}}\mathrm{Tr}(\mathbf{X}) = \mathbf{I} \\
\frac{\partial}{\partial \boldsymbol{x}}\prod \mathrm{eig}(\mathbf{X}) &= \frac{\partial}{\partial \boldsymbol{x}}\det(\mathbf{X}) = \det(\mathbf{X})\mathbf{X}^{-\intercal}
\end{aligned}$$

If $\mathbf{A}$ is real and symmetric, $\boldsymbol{\lambda}_i$ and $\mathbf{v}_i$ are distinct eigenvalues and eigenvectors of $\mathbf{A}$ with $\mathbf{v}_i^{\intercal}\mathbf{v}_i = 1$, then

$$\begin{aligned}
\partial \boldsymbol{\lambda}_i &= \mathbf{v}_i^{\intercal}\partial(\mathbf{A})\mathbf{v}_i \\
\partial \mathbf{v}_i &= (\boldsymbol{\lambda}_i\mathbf{I} - \mathbf{A})^{+}\partial(\mathbf{A})\mathbf{v}_i
\end{aligned}$$

4) Derivatives of Matrices, Vectors and Scalar Forms: Add text here...

First Order:

$$\begin{aligned}
\frac{\partial \boldsymbol{x}^{\intercal}\mathbf{a}}{\partial \boldsymbol{x}} &= \frac{\partial \mathbf{a}^{\intercal}\boldsymbol{x}}{\partial \boldsymbol{x}} = \mathbf{a} \\
\frac{\partial \mathbf{a}^{\intercal}\mathbf{X}\mathbf{b}}{\partial \mathbf{X}} &= \mathbf{a}\mathbf{b}^{\intercal} \\
\frac{\partial \mathbf{a}^{\intercal}\mathbf{X}^{\intercal}\mathbf{b}}{\partial \mathbf{X}} &= \mathbf{b}\mathbf{a}^{\intercal} \\
\frac{\partial \mathbf{a}^{\intercal}\mathbf{X}\mathbf{a}}{\partial \mathbf{X}} &= \frac{\partial \mathbf{a}^{\intercal}\mathbf{X}^{\intercal}\mathbf{a}}{\partial \boldsymbol{x}} = \mathbf{a}\mathbf{a}^{\intercal}
\end{aligned}$$

Second Order:

$$
\begin{aligned}
\frac{\partial}{\partial X_{ij}} \sum_{klmn} X_{kl} X_{mn} &= 2 \sum_{kl} X_{kl} \\
\frac{\partial \mathbf{b}^\intercal \mathbf{X}^\intercal \mathbf{X} \mathbf{c}}{\partial \mathbf{X}} &= \mathbf{X}(\mathbf{b}\mathbf{c}^\intercal + \mathbf{c}\mathbf{b}^\intercal) \\
\frac{\partial (\mathbf{B}\boldsymbol{x} + \mathbf{b})^\intercal \mathbf{C} (\mathbf{D}\boldsymbol{x} + \mathbf{d})}{\partial \boldsymbol{x}} &= \mathbf{B}^\intercal \mathbf{C} (\mathbf{D}\boldsymbol{x} + \mathbf{d}) \\
&\quad + \mathbf{D}^\intercal \mathbf{C}^\intercal (\mathbf{B}\boldsymbol{x} + \mathbf{b})
\end{aligned}
$$

Higher-order and Nonlinear:

Gradient and Hessian:

5) Derivatives of Traces: Add text here...

First Order:

Second Order:

Higher-order:

Other:

6) Derivatives of Vector Norms:

Two-norm:

$$
\begin{aligned}
\frac{\partial}{\partial \boldsymbol{x}} \|\boldsymbol{x} - \mathbf{a}\|_2 &= \frac{\boldsymbol{x} - \mathbf{a}}{\|\boldsymbol{x} - \mathbf{a}\|_2} \\
\frac{\partial}{\partial \boldsymbol{x}} \frac{\boldsymbol{x} - \mathbf{a}}{\|\boldsymbol{x} - \mathbf{a}\|_2} &= \frac{\mathbf{I}}{\|\boldsymbol{x} - \mathbf{a}\|_2} - \frac{(\boldsymbol{x} - \mathbf{a})(\boldsymbol{x} - \mathbf{a})^\intercal}{\|\boldsymbol{x} - \mathbf{a}\|_2^3} \\
\frac{\partial \|\boldsymbol{x}\|_2^2}{\partial \boldsymbol{x}} &= \frac{\partial \|\boldsymbol{x}^\intercal \boldsymbol{x}\|_2}{\partial \boldsymbol{x}} = 2\boldsymbol{x}
\end{aligned}
$$

7) Derivatives of Matrix Norms:

Frobenius norm:

$$
\frac{\partial}{\partial \mathbf{X}} \|\mathbf{X}\|_F^2 = \frac{\partial}{\partial \mathbf{X}} \mathrm{Tr}(\mathbf{X}\mathbf{X}^H) = 2\mathbf{X}
$$

Note that this is also a special case of the result in eqn. (**??**)

8) Derivatives of Structured Matrices: Add text here...

The Chain Rule:

Symmetric:

Diagonal:

Toeplitz:

9) Derivatives with Respect to Scalars: In the case where the vectors $\mathbf{u}$ and $\mathbf{v}$, the matrix $\mathbf{A}$, and the scalar μ are functions of a scalar quantity s, the derivative with respect to s has the following properties:

$$
\begin{aligned}
\frac{d}{ds}(\mathbf{u} + \mathbf{v}) &= \frac{d\mathbf{u}}{ds} + \frac{d\mathbf{v}}{ds} \\
\frac{d}{ds}(\mathbf{u} \cdot \mathbf{v}) &= \frac{d\mathbf{u}}{ds} \cdot \mathbf{v} + \mathbf{u} \cdot \frac{d\mathbf{v}}{ds} \\
\frac{d}{ds}(\mathbf{u} \times \mathbf{v}) &= \frac{d\mathbf{u}}{ds} \times \mathbf{v} + \mathbf{u} \times \frac{d\mathbf{v}}{ds} \\
\frac{d}{ds}(\mu \mathbf{v}) &= \frac{d\mu}{ds} \mathbf{v} + \mu \frac{d\mathbf{v}}{ds} \\
\frac{d}{ds} \mathbf{A}^{-1} &= -\mathbf{A}^{-1} \left(\frac{d}{ds} \mathbf{A} \right) \mathbf{A}^{-1}.
\end{aligned}
$$

10) Derivatives with Respect to Vectors: Using the convention that gradients of scalar functions are defined as row vectors,

$$
\begin{aligned}
\frac{d}{d\mathbf{v}}(\mathbf{u} \cdot \mathbf{v}) &= \frac{d}{d\mathbf{v}}(\mathbf{v} \cdot \mathbf{u}) = \mathbf{u}^\intercal \\
\frac{d}{d\mathbf{v}}(\mathbf{A}\mathbf{v}) &= \mathbf{A} \\
\frac{d}{d\mathbf{v}}(\mathbf{v}^\intercal \mathbf{A}) &= \mathbf{A}^\intercal \\
\frac{d}{d\mathbf{v}}(\mathbf{v}^\intercal \mathbf{A} \mathbf{v}) &= \mathbf{v}^\intercal (\mathbf{A} + \mathbf{A}^\intercal) \\
&= 2\mathbf{v}^\intercal \mathbf{A}, \text{if } \mathbf{A} \text{ is symmetric.}
\end{aligned}
$$

11) Derivatives with Respect to Matrices: The derivative of the scalar μ with respect to the matrix $\mathbf{A}$ is defined by

$$
\frac{d\mu}{d\mathbf{A}} = \begin{bmatrix} \frac{d\mu}{da_{11}} & \frac{d\mu}{da_{12}} & \cdots & \frac{d\mu}{da_{1n}} \\ \frac{d\mu}{da_{21}} & \frac{d\mu}{da_{22}} & \cdots & \frac{d\mu}{da_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{d\mu}{da_{m1}} & \frac{d\mu}{da_{m2}} & \cdots & \frac{d\mu}{da_{mn}} \end{bmatrix}. \tag{200}
$$

For the scalar operation defined by the trace, the following differentiation formulas are useful:

$$
\begin{aligned}
\frac{d(\mathrm{Tr}(\mathbf{A}\mathbf{B}))}{d\mathbf{A}} &= \mathbf{B}^\intercal && (\mathbf{A}\mathbf{B} \text{ must be square}) \\
\frac{d(\mathrm{Tr}(\mathbf{A}\mathbf{B}\mathbf{A}^\intercal))}{d\mathbf{A}} &= 2\mathbf{A}\mathbf{B}^\intercal && (\mathbf{B} \text{ must be symmetric}).
\end{aligned}
$$

V. Solutions to Linear Equations

A. Existence in Linear Systems

Consider the linear system $\mathbf{A}\boldsymbol{x} = \mathbf{b}$, and the augmented matrix $[\mathbf{A}\ \mathbf{b}]$, $\mathbf{A} \in \mathbb{R}^{p \times q}$, $\mathbf{b} \in \mathbb{R}^p$, $\boldsymbol{x} \in \mathbb{R}^q$, only one of the following is valid:

Condition	*Solution*
$\mathrm{rank}([\mathbf{A}\ \mathbf{b}]) \geq \mathrm{rank}(\mathbf{A})$	0 solutions exist
$\mathrm{rank}([\mathbf{A}\ \mathbf{b}]) = \mathrm{rank}(\mathbf{A}) = q$	1 solution exists
$\mathrm{rank}([\mathbf{A}\ \mathbf{b}]) = \mathrm{rank}(\mathbf{A}) < q$	∞ solutions exist

B. Standard Square

Assume $\mathbf{A}$ is square and invertible, then

$$
\mathbf{A}\boldsymbol{x} = \mathbf{b} \quad \Longrightarrow \quad \boldsymbol{x} = \mathbf{A}^{-1}\mathbf{b}
$$

C. Degenerated Square

Assume $\mathbf{A}$ is $n \times n$ but of rank $r < n$. In that case, the system $\mathbf{A}\boldsymbol{x} = \mathbf{b}$ is solved by

$$\boldsymbol{x} = \mathbf{A}^{+}\mathbf{b}$$

where $\mathbf{A}^{+}$ is the pseudo-inverse of the rank-deficient matrix.

D. Cramer's rule

The equation $\mathbf{A}\boldsymbol{x} = \mathbf{b}$ where $\mathbf{A}$ is square, has exactly one solution $\boldsymbol{x}$ if the i^{th} element in $\boldsymbol{x}$ can be found as

$$\boldsymbol{x}_i = \frac{|\mathbf{B}|}{|\mathbf{A}|}$$

where $\mathbf{B}$ equals $\mathbf{A}$, but the i^{th} column in $\mathbf{A}$ has been substituted by $\mathbf{b}$.

E. Over-determined Rectangular

Assume $\mathbf{A}$ to be $n \times m$, $n > m$ (tall) and $\text{rank}(\mathbf{A}) = m$, then

$$\mathbf{A}\boldsymbol{x} = \mathbf{b} \quad \Longrightarrow \quad \boldsymbol{x} = (\mathbf{A}^{\mathsf{T}}\mathbf{A})^{-1}A^{\mathsf{T}}\mathbf{b} = \mathbf{A}^{+}\mathbf{b}$$

that is if there exists a solution $\boldsymbol{x}$ at all! If there is no solution, the following can be useful:

$$\mathbf{A}\boldsymbol{x} = \mathbf{b} \quad \Longrightarrow \quad \boldsymbol{x}_{\text{min}} = \mathbf{A}^{+}\mathbf{b}$$

Now $\boldsymbol{x}_{\text{min}}$ is the vector $\boldsymbol{x}$ which minimizes $\|\mathbf{A}\boldsymbol{x} - \mathbf{b}\|_2^2$. The matrix $\mathbf{A}^{+}$ is the pseudo-inverse of $\mathbf{A}$.

F. Under-determined Rectangular

Assume $\mathbf{A}$ is $n \times m$ and $n < m$ (broad) and $\text{rank}(\mathbf{A}) = n$.

$$\mathbf{A}\boldsymbol{x} = \mathbf{b} \quad \Longrightarrow \quad \boldsymbol{x}_{\text{min}} = \mathbf{A}^{\mathsf{T}}(\mathbf{A}\mathbf{A}^{\mathsf{T}})^{-1}\mathbf{b}$$

The equation have many solutions $\boldsymbol{x}$. But $\boldsymbol{x}_{\text{min}}$ is the solution which minimizes $\|\mathbf{A}\boldsymbol{x} - \mathbf{b}\|_2^2$ and also the solution with the smallest norm $\|\boldsymbol{x}\|_2^2$. The same holds for a matrix version: Assume $\mathbf{A}$ is $n \times m$, $\mathbf{X}$ is $m \times n$ and $\mathbf{B}$ is $n \times n$, then

$$\mathbf{A}\mathbf{X} = \mathbf{B} \quad \Longrightarrow \quad \mathbf{X}_{\text{min}} = \mathbf{A}^{+}\mathbf{B}$$

The equation may have many solutions $\mathbf{X}$. But $\mathbf{X}_{\text{min}}$ is the solution which minimizes $\|\mathbf{A}\mathbf{X} - \mathbf{B}\|_2^2$ and also the solution with the smallest norm $\|\mathbf{X}\|_2^2$.

Similar but different: Assume $\mathbf{A}$ is square $n \times n$ and the matrices $\mathbf{B}_0$, $\mathbf{B}_1$ are $n \times N$, where $N > n$, then if $\mathbf{B}_0$ has maximal rank

$$\mathbf{A}\mathbf{B}_0 = \mathbf{B}1 \quad \Longrightarrow \quad \mathbf{A}_{\text{min}} = \mathbf{B}_1\mathbf{B}_0^{\mathsf{T}}(\mathbf{B}_0\mathbf{B}_0^{\mathsf{T}})^{-1}$$

where $\mathbf{A}_{\text{min}}$ denotes the matrix which is optimal in a least square sense. An interpretation is that $\mathbf{A}$ is the linear approximation which maps the columns vectors of $\mathbf{B}_0$ into the columns vectors of $\mathbf{B}_1$.

G. Linear form and zeros

$$\mathbf{A}\boldsymbol{x} = \mathbf{0}, \ \forall \boldsymbol{x} \quad \Longrightarrow \quad \mathbf{A} = \mathbf{0}$$

H. Square form and zeros

If $\mathbf{A}$ is symmetric, then

$$\boldsymbol{x}^{\mathsf{T}}\mathbf{A}\boldsymbol{x} = \mathbf{0}, \ \forall \boldsymbol{x} \quad \Longrightarrow \quad \mathbf{A} = \mathbf{0}$$

VI. Special Matrices

A. Block matrices

Let $\mathbf{A}_{ij}$ denote the $(i,j)^{th}$ block of the matrix $\mathbf{A}$.

1) Multiplication: Assuming the dimensions of the blocks are equal, we have

$$\left[\begin{array}{c|c} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \hline \mathbf{A}_{21} & \mathbf{A}_{22} \end{array}\right] \left[\begin{array}{c|c} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \hline \mathbf{B}_{21} & \mathbf{B}_{22} \end{array}\right] = \left[\begin{array}{c|c} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{21} & \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} \\ \hline \mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21} & \mathbf{A}_{21}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} \end{array}\right]$$

2) The Determinant: The determinant can be expressed as

$$\det\left(\left[\begin{array}{c|c} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \hline \mathbf{A}_{21} & \mathbf{A}_{22} \end{array}\right]\right) = \det(\mathbf{A}_{22}) \cdot \det(\mathbf{C}_1) = \det(\mathbf{A}_{11}) \cdot \det(\mathbf{C}_2)$$

where,

$$\mathbf{C}_1 = \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}$$
$$\mathbf{C}_2 = \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$$

3) The Inverse: The inverse can be expressed as

$$\left[\begin{array}{c|c} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \hline \mathbf{A}_{21} & \mathbf{A}_{22} \end{array}\right]^{-1} = \left[\begin{array}{c|c} \mathbf{C}_1^{-1} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{C}_2^{-1} \\ \hline -\mathbf{C}_2^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{C}_2^{-1} \end{array}\right] = \left[\begin{array}{c|c} \mathbf{B}_1 & \mathbf{B}_2 \\ \hline \mathbf{B}_3 & \mathbf{B}_4 \end{array}\right]$$

where,

$$\mathbf{B}_1 = \mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{C}_2^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1}$$
$$\mathbf{B}_2 = -\mathbf{C}_1^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1}$$
$$\mathbf{B}_3 = -\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{C}_1^{-1}$$
$$\mathbf{B}_4 = \mathbf{A}_{22}^{-1} + \mathbf{A}_{22}^{-1}\mathbf{A}_{21}\mathbf{C}_1^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1}$$
$$\mathbf{C}_1 = \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}$$
$$\mathbf{C}_2 = \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$$

4) Block diagonal: For block diagonal matrices we have

$$\left[\begin{array}{c|c} \mathbf{A}_{11} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{A}_{22} \end{array}\right]^{-1} = \left[\begin{array}{c|c} (\mathbf{A}_{11})^{-1} & \mathbf{0} \\ \hline \mathbf{0} & (\mathbf{A}_{22})^{-1} \end{array}\right]$$

$$\det\left(\left[\begin{array}{c|c} \mathbf{A}_{11} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{A}_{22} \end{array}\right]\right) = \det(\mathbf{A}_{11}) \cdot \det(\mathbf{A}_{22})$$

5) Schur complement: Consider the matrix

$$\left[\begin{array}{c|c} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \hline \mathbf{A}_{21} & \mathbf{A}_{22} \end{array}\right] \tag{201}$$

The Schur complement of block $\mathbf{A}_{11}$ of the matrix above is the matrix (denoted $\mathbf{C}_2$ in the text above)

$$\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$$

The Schur complement of block $\mathbf{A}_{22}$ of the matrix above is the matrix (denoted $\mathbf{C}_1$ in the text above)

$$\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}$$

Using the Schur complement, one can rewrite the inverse of a block matrix

$$\left[\begin{array}{c|c} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \hline \mathbf{A}_{21} & \mathbf{A}_{22} \end{array}\right]^{-1} = \left[\begin{array}{c|c} \mathbf{I} & \mathbf{0} \\ \hline -\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{I} \end{array}\right] \cdot \left[\begin{array}{c|c} (\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{A}_{22}^{-1} \end{array}\right] \cdot \left[\begin{array}{c|c} \mathbf{I} & -\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ \hline \mathbf{0} & \mathbf{I} \end{array}\right]$$

The Schur complement is useful when solving linear systems of the form

$$\left[\begin{array}{c|c} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \hline \mathbf{A}_{21} & \mathbf{A}_{22} \end{array}\right] \left[\begin{array}{c} \boldsymbol{x}_1 \\ \hline \boldsymbol{x}_2 \end{array}\right] = \left[\begin{array}{c} \mathbf{b}_1 \\ \hline \mathbf{b}_2 \end{array}\right]$$

which has the following equation for $\boldsymbol{x}_1$

$$(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})\boldsymbol{x}_1 = \mathbf{b}_1 - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{b}_2$$

When the appropriate inverses exists, this can be solved for $\boldsymbol{x}_1$ which can then be inserted in the equation for $\boldsymbol{x}_2$ to solve for $\boldsymbol{x}_2$.

B. Hermitian Matrices and skew-Hermitian

1) Hermitian Matrices: A matrix $\mathbf{A} \in \mathbb{C}^{m \times n}$ is called *Hermitian* if

$$\mathbf{A}^H = \mathbf{A}$$

For real valued matrices, Hermitian and symmetric matrices are equivalent.

$$\begin{aligned} \mathbf{A} \text{ is Hermitian} &\iff \boldsymbol{x}^H\mathbf{A}\boldsymbol{x} \in \mathbb{R}, \quad \forall \boldsymbol{x} \in \mathbb{C}^{n \times 1} \\ \mathbf{A} \text{ is Hermitian} &\iff \text{eig}(\mathbf{A}) \in \mathbb{R} \end{aligned}$$

Note that

$$\mathbf{A} = \mathbf{B} + i\mathbf{C}$$

where $\mathbf{B}$, $\mathbf{C}$ are Hermitian, then

$$\mathbf{B} = \frac{\mathbf{A} + \mathbf{A}^H}{2}$$
$$\mathbf{C} = \frac{\mathbf{A} - \mathbf{A}^H}{2i}$$

2) skew-Hermitian: A matrix $\mathbf{A}$ is called *skew-Hermitian* if

$$\mathbf{A} = -\mathbf{A}^H$$

For real valued matrices, skew-Hermitian and skew-symmetric matrices are equivalent.

$$\begin{aligned} \mathbf{A} \text{ is Hermitian} &\iff i\mathbf{A} \text{ is skew-Hermitian} \\ \mathbf{A} \text{ is skew-Hermitian} &\iff \boldsymbol{x}^H\mathbf{A}\mathbf{y} = -\boldsymbol{x}^H\mathbf{A}^H\mathbf{y}, \quad \forall \boldsymbol{x}, \mathbf{y} \\ \mathbf{A} \text{ is skew-Hermitian} &\iff \text{eig}(\mathbf{A}) = i\lambda, \quad \lambda \in \mathbb{R} \end{aligned}$$

C. Idempotent Matrices

A matrix $\mathbf{A}$ is idempotent if

$$\mathbf{A}\mathbf{A} = \mathbf{A}$$

Idempotent matrices $\mathbf{A}$ and $\mathbf{B}$, have the following properties

$$\begin{aligned} \mathbf{A}^n &= \mathbf{A}, \text{ for } n = 1, 2, 3, \ldots \\ \mathbf{I} - \mathbf{A} &\text{ is idempotent} \\ \mathbf{A}^H &\text{ is idempotent} \\ \mathbf{I} - \mathbf{A}^H &\text{ is idempotent} \\ \text{If } \mathbf{AB} = \mathbf{BA} \implies \mathbf{AB} &\text{ is idempotent} \\ \text{rank}(\mathbf{A}) &= \text{Tr}(\mathbf{A}) \\ \mathbf{A}(\mathbf{I} - \mathbf{A}) &= 0 \\ (\mathbf{I} - \mathbf{A})\mathbf{A} &= 0 \\ \mathbf{A}^+ &= \mathbf{A} \\ f(s\mathbf{I} + t\mathbf{A}) &= (\mathbf{I} - \mathbf{A})f(s) + \mathbf{A}f(s + t) \end{aligned}$$

Note that $\mathbf{A} - \mathbf{I}$ is not necessarily idempotent.

1) Nilpotent: A matrix $\mathbf{A}$ is nilpotent if

$$\mathbf{A}^2 = \mathbf{0}$$

A nilpotent matrix has the following property:

$$f(s\mathbf{I} + t\mathbf{A}) = \mathbf{I}f(s) + t\mathbf{A}f'(s)$$

2) Unipotent: A matrix $\mathbf{A}$ is unipotent if

$$\mathbf{A}\mathbf{A} = \mathbf{I}$$

A unipotent matrix has the following property:

$$f(s\mathbf{I} + t\mathbf{A}) = [(\mathbf{I} + \mathbf{A})f(s + t) + (\mathbf{I} - \mathbf{A})f(s - t)]/2$$

D. Orthogonal Matrices

If a square matrix $\mathbf{Q}$ is orthogonal, if and only if,

$$\mathbf{Q}^\mathsf{T}\mathbf{Q} = \mathbf{Q}\mathbf{Q}^\mathsf{T} = \mathbf{I}$$

and then $\mathbf{Q}$ has the following properties

- Its eigenvalues are placed on the unit circle.
- Its eigenvectors are unitary, i.e. have length one.
- The inverse of an orthogonal matrix is orthogonal too.

Basic properties for the orthogonal matrix $\mathbf{Q}$

$$\begin{aligned} \mathbf{Q}^{-1} &= \mathbf{Q}^\mathsf{T} \\ \mathbf{Q}^{\mathsf{T}} &= \mathbf{Q} \\ \mathbf{Q}\mathbf{Q}^\mathsf{T} &= \mathbf{I} \\ \mathbf{Q}^\mathsf{T}\mathbf{Q} &= \mathbf{I} \\ |\mathbf{Q}| &= \pm 1 \end{aligned}$$

1) Ortho-Symmetric: A matrix $\mathbf{Q}_+$ which simultaneously is orthogonal and symmetric is called an ortho-sym matrix. Therefore

$$\mathbf{Q}_+^\mathsf{T}\mathbf{Q}_+ = \mathbf{I} \tag{202}$$
$$\mathbf{Q}_+ = \mathbf{Q}_+^\mathsf{T} \tag{203}$$

The powers of an ortho-sym matrix are given by the following rule

$$\begin{aligned}\mathbf{Q}_+^k &= \frac{1+(-1)^k}{2}\mathbf{I} + \frac{1+(-1)^{k+1}}{2}\mathbf{Q}_+ \\ &= \frac{1+\cos(k\pi)}{2}\mathbf{I} + \frac{1-\cos(k\pi)}{2}\mathbf{Q}_+\end{aligned}$$

2) Ortho-Skew: A matrix which simultaneously is orthogonal and antisymmetric is called an ortho-skew matrix. Therefore

$$\mathbf{Q}_-^H\mathbf{Q}_- = \mathbf{I} \tag{204}$$
$$\mathbf{Q}_- = \mathbf{Q}_-^H \tag{205}$$

The powers of an ortho-skew matrix are given by the following rule

$$\begin{aligned}\mathbf{Q}_-^k &= \frac{i^k+(-i)^k}{2}\mathbf{I} - i\frac{i^k+(-i)^k}{2}\mathbf{Q}_- \\ &= \cos(k\frac{\pi}{2})\mathbf{I} + \sin(k\frac{\pi}{2})\mathbf{Q}_-\end{aligned}$$

E. Positive Definite and Semi-definite Matrices

1) Definition: A matrix $\mathbf{A}$ is positive definite if and only if

$$\boldsymbol{x}^\mathsf{T}\mathbf{A}\boldsymbol{x} > \mathbf{0}, \ \forall \boldsymbol{x} \neq \mathbf{0}$$

A matrix $\mathbf{A}$ is positive semi-definite if and only if

$$\boldsymbol{x}^\mathsf{T}\mathbf{A}\boldsymbol{x} \geq \mathbf{0}, \ \forall \boldsymbol{x}$$

Note that if $\mathbf{A}$ is positive definite, then $\mathbf{A}$ is also positive semi-definite.

2) Eigenvalues: The following holds with respect to the eigenvalues:

$$\begin{aligned}\mathbf{A} \text{ pos. def.} &\iff \text{eig}\left(\frac{\mathbf{A}+\mathbf{A}^H}{2}\right) > 0 \\ \mathbf{A} \text{ pos. semi-def.} &\iff \text{eig}\left(\frac{\mathbf{A}+\mathbf{A}^H}{2}\right) \geq 0\end{aligned}$$

3) Trace: The following holds with respect to the trace:

$$\begin{aligned}\mathbf{A} \text{ pos. def.} &\implies \text{Tr}(\mathbf{A}) > 0 \\ \mathbf{A} \text{ pos. semi-def.} &\implies \text{Tr}(\mathbf{A}) \geq 0\end{aligned}$$

4) Inverse: If $\mathbf{A}$ is positive definite, then $\mathbf{A}$ is invertible and $\mathbf{A}^{-1}$ is also positive definite.

5) Diagonal: If $\mathbf{A}$ is positive definite, then $A_{ii} > 0, \ \forall i$.

F. Symmetric, Skew-symmetric/Antisymmetric

1) Symmetric: The matrix $\mathbf{A}$ is said to be symmetric if

$$\mathbf{A} = \mathbf{A}^\mathsf{T}$$

Symmetric matrices have many important properties, e.g. that their eigenvalues are real and eigenvectors orthogonal.

2) Skew-symmetric/Antisymmetric: The antisymmetric matrix is also known as the skew symmetric matrix. It has the following property from which it is defined

$$\mathbf{A} = -\mathbf{A}^\mathsf{T}$$

Hereby, it can be seen that the antisymmetric matrices always have a zero diagonal. The $n \times n$ antisymmetric matrices also have the following properties.

$$\begin{aligned}|\mathbf{A}^\mathsf{T}| = |-\mathbf{A}| &= (-1)^n|\mathbf{A}| \\ -|\mathbf{A}| = |-\mathbf{A}| &= 0, \ \text{if n is odd}\end{aligned}$$

The eigenvalues of an antisymmetric matrix are placed on the imaginary axis and the eigenvectors are unitary.

G. Decomposition

(add text here... use EE230 summary)

1) LU: Assume $\mathbf{A}$ is a square matrix with non-zero leading principal minors, then

$$\mathbf{A} = \mathbf{LU}$$

where $\mathbf{L}$ is a unique unit lower triangular matrix and $\mathbf{U}$ is a unique upper triangular matrix.

2) LUP: Assume $\mathbf{A}$ is a square matrix with non-zero leading principal minors, then

$$\mathbf{A} = \mathbf{LUP}$$

where $\mathbf{L}$ is a unique unit lower triangular matrix, $\mathbf{U}$ is a unique upper triangular matrix, and $\mathbf{P}$ is a permutation matrix.

3) LDL: The LDL decomposition is a special case of the LU decomposition. Assume $\mathbf{A}$ is a non-singular symmetric definite square matrix, then

$$\mathbf{A} = \mathbf{LDL}^\mathsf{T} = \mathbf{L}^\mathsf{T}\mathbf{DL}$$

where $\mathbf{L}$ is a unit lower triangular matrix and $\mathbf{D}$ is a diagonal matrix. If $\mathbf{A}$ is also positive definite, then $\mathbf{D}$ has strictly positive diagonal entries.

4) QR: **(add text here...)**

5) QR with Givens Rotations: **(add text here...)**

6) Cholesky: Assume $\mathbf{A}$ is a symmetric positive definite square matrix, then

$$\mathbf{A} = \mathbf{U}^\mathsf{T}\mathbf{U} = \mathbf{LL}^\mathsf{T}$$

where $\mathbf{U}$ is a unique upper triangular matrix and $\mathbf{L}$ is a lower triangular matrix.

Consider the matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$

$$\begin{aligned}\mathbf{A} &= \mathbf{LL}^\mathsf{T} \\ &= \begin{bmatrix} L_{11} & 0 & \cdots & 0 \\ L_{21} & L_{22} & \cdots & \vdots \\ \vdots & & \ddots & 0 \\ L_{n1} & \cdots & \cdots & L_{nn}\end{bmatrix}\begin{bmatrix} L_{11} & L_{12} & \cdots & L_{1n} \\ 0 & L_{22} & \cdots & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & L_{nn}\end{bmatrix}\end{aligned}$$

Then,

$$L_{jj} = \sqrt{A_{jj} - \sum_{k=1}^{j-1} L_{jk}^2}$$
$$L_{ij} = \frac{1}{L_{jj}} \left(A_{ij} - \sum_{k=1}^{j-1} L_{ik} L_{jk} \right), \quad i > j$$

7) EVD: Two forms of the Eigenvalue Decomposition exist.

Symmetric Square decomposed into squares: Assume $\mathbf{A}$ to be $n \times n$ and symmetric. Then

$$\mathbf{A} = \mathbf{V}\mathbf{D}\mathbf{V}^\intercal$$

where $\mathbf{D}$ is diagonal with the eigenvalues of $\mathbf{A}$, and $\mathbf{V}$ is orthogonal and the eigenvectors of $\mathbf{A}$.

Square decomposed into squares: Assume $\mathbf{A} \in \mathbb{R}^{n \times n}$. Then

$$\mathbf{A} = \mathbf{V}\mathbf{D}\mathbf{U}^\intercal$$

where $\mathbf{D}$ is diagonal with the square root of the eigenvalues of $\mathbf{A}\mathbf{A}^\intercal$, $\mathbf{V}$ are the eigenvectors of $\mathbf{A}\mathbf{A}^\intercal$ and $\mathbf{U}^\intercal$ are the eigenvectors of $\mathbf{A}^\intercal\mathbf{A}$.

8) SVD: Any $n \times m$ matrix $\mathbf{A}$ can be written as

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\intercal$$

where

$$\mathbf{U} = \text{ eigenvectors of } \mathbf{A}\mathbf{A}^\intercal, \; n \times n$$
$$\mathbf{\Sigma} = \sqrt{\text{diag}(\text{eig}(\mathbf{A}\mathbf{A}^\intercal))}, \; n \times m$$
$$\mathbf{V} = \text{ eigenvectors of } \mathbf{A}^\intercal\mathbf{A}, \; m \times m$$

Consider the matrix $\mathbf{A} \in \mathbb{R}^{p \times q}$

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\intercal$$

where $\mathbf{U}\mathbf{U}^\intercal = \mathbf{I}$, and $\mathbf{V}\mathbf{V}^\intercal = \mathbf{I}$. The columns of $\mathbf{U}$ form an orthonormal basis for $\mathbf{A}$ in $\mathbb{R}^p$. The columns of $\mathbf{V}$ form an orthonormal basis for $\mathbf{A}$ in $\mathbb{R}^q$. The diagonal matrix $\mathbf{\Sigma}$, contains the singular values of $\mathbf{A}$.

Let $\mathbf{U} \triangleq [U_1, \ldots, U_r | U_{r+1}, \ldots, U_p]$ and $\mathbf{V} \triangleq [V_1, \ldots, V_r | V_{r+1}, \ldots, V_q]$, then

- $U_{1:r}$ are an orthonormal basis for the column space of $\mathbf{A}$, i.e. $CS(\mathbf{A})$.
- $U_{(r+1):p}$ are an orthonormal basis for the left-null space of $\mathbf{A}$, i.e. $LN(\mathbf{A})$.
- $V_{1:r}$ are an orthonormal basis for the row space of $\mathbf{A}$, i.e. $RS(\mathbf{A})$.
- $V_{(r+1):q}$ are an orthonormal basis for the null space of $\mathbf{A}$, i.e. $Null(\mathbf{A})$.
- any vector in $U_{1:r} \perp U_{(r+1):p}$, i.e. $CS(\mathbf{A}) \perp LN(\mathbf{A})$
- any vector in $V_{1:r} \perp V_{(r+1):q}$, i.e. $RS(\mathbf{A}) \perp Null(\mathbf{A})$

Finally, we have that

- $CS(\mathbf{A}) \equiv \text{range}(\mathbf{A}) \equiv \text{image}(\mathbf{A})$.
- $\dim(Null(\mathbf{A})) = \text{nullity}(\mathbf{A})$.
- $\text{rank}(\mathbf{A}) + \dim(Null(\mathbf{A})) = \#$ of columns.
- dim $=$ # of linearly independent vectors in a space, which equals a basis.

VII. Probability Theory

The following section contains some useful relations from Probability Theory.

A. Definitions

The *expected value*, or *mean*, of a discrete random variable (r.v. hereafter) $\mathbf{X}$:

$$\mathbf{m_X} = \mathrm{E}\langle \mathbf{X} \rangle = \sum_{k=1}^{\infty} \boldsymbol{x}_k \; P_{\mathbf{X}}(\boldsymbol{x}_k)$$

The *variance* of the r.v. $\mathbf{X}$:

$$\boldsymbol{\sigma}_{\mathbf{X}}^2 = \text{var}\langle \mathbf{X} \rangle = \mathrm{E}\left\langle (\mathbf{X} - \mathbf{m_X})^2 \right\rangle = \sum_{k=1}^{\infty} (\boldsymbol{x} - \mathbf{m_X})^2 \; P_{\mathbf{X}}(\boldsymbol{x}_k)$$

which can also be expressed as

$$\begin{aligned}\text{var}\langle \mathbf{X} \rangle &= \mathrm{E}\left\langle (\mathbf{X} - \mathbf{m_X})^2 \right\rangle \\ &= \mathrm{E}\left\langle \mathbf{X}^2 - 2\mathbf{m_X}\mathbf{X} + \mathbf{m_X^2} \right\rangle \\ &= \mathrm{E}\left\langle \mathbf{X}^2 \right\rangle - 2\mathbf{m_X}\mathrm{E}\langle \mathbf{X} \rangle + \mathbf{m_X^2} \\ &= \mathrm{E}\left\langle \mathbf{X}^2 \right\rangle - \mathbf{m_X^2}\end{aligned}$$

where $\mathrm{E}\left\langle \mathbf{X}^2 \right\rangle$ is the called the second moment of $\mathbf{X}$.

The *standard deviation* of the r.v. $\mathbf{X}$:

$$\boldsymbol{\sigma}_{\mathbf{X}} = \text{std}\langle \mathbf{X} \rangle = \sqrt{\text{var}\langle \mathbf{X} \rangle}$$

For the discrete r.v. $\boldsymbol{x} \in \mathbb{R}^{m \times 1}$ and scalar variable c, the variance has the following properties, each resulting in a *scalar random variable*:

1) Adding a constant c, does not affect the variance of $\mathbf{X}$

$$\begin{aligned}\text{var}\langle \boldsymbol{x} + c \rangle &= \mathrm{E}\left\langle (\boldsymbol{x} + c - (\mathrm{E}\langle \boldsymbol{x} \rangle + c))^2 \right\rangle \\ &= \mathrm{E}\left\langle (\boldsymbol{x} - \mathrm{E}\langle \boldsymbol{x} \rangle)^2 \right\rangle \\ &= \text{var}\langle \boldsymbol{x} \rangle\end{aligned}$$

2) Multiplying by a constant c, scales the variance of $\boldsymbol{x}$ by c^2

$$\begin{aligned}\text{var}\langle c\boldsymbol{x} \rangle &= \mathrm{E}\left\langle (c\boldsymbol{x} - c\mathrm{E}\langle \boldsymbol{x} \rangle)^2 \right\rangle \\ &= \mathrm{E}\left\langle c^2(\boldsymbol{x} - \mathrm{E}\langle \boldsymbol{x} \rangle)^2 \right\rangle \\ &= c^2\text{var}\langle \boldsymbol{x} \rangle\end{aligned}$$

3) Let $\boldsymbol{x} = c$, then the constant r.v. has zero variance

$$\begin{aligned}\text{var}\langle \boldsymbol{x} \rangle &= \mathrm{E}\left\langle (\boldsymbol{x} - c)^2 \right\rangle \\ &= \mathrm{E}\langle 0 \rangle \\ &= 0\end{aligned}$$

For the discrete r.v. $\boldsymbol{x}, \mathbf{y} \in \mathbb{R}^{m \times 1}$, the covariance has the following property, resulting in a *matrix of random variables*:

$$\begin{aligned}\text{cov}\langle \boldsymbol{x}, \mathbf{y} \rangle &= \mathrm{E}\langle \boldsymbol{x}\mathbf{y} - \boldsymbol{x}\mathrm{E}\langle \mathbf{y} \rangle - \mathbf{y}\mathrm{E}\langle \boldsymbol{x} \rangle + \mathrm{E}\langle \boldsymbol{x} \rangle \mathrm{E}\langle \mathbf{y} \rangle \rangle \\ &= \mathrm{E}\langle \boldsymbol{x}\mathbf{y} \rangle - 2\mathrm{E}\langle \boldsymbol{x} \rangle \mathrm{E}\langle \mathbf{y} \rangle + \mathrm{E}\langle \boldsymbol{x} \rangle \mathrm{E}\langle \mathbf{y} \rangle \\ &= \mathrm{E}\langle \boldsymbol{x}\mathbf{y} \rangle - \mathrm{E}\langle \boldsymbol{x} \rangle \mathrm{E}\langle \mathbf{y} \rangle\end{aligned}$$

If either $\boldsymbol{x}$ or $\mathbf{y}$ have zero mean, i.e. $\mathrm{E}\langle\boldsymbol{x}\rangle = 0$ or $\mathrm{E}\langle\mathbf{y}\rangle = 0$, then $\mathrm{cov}\langle\boldsymbol{x},\mathbf{y}\rangle = \mathrm{E}\langle\boldsymbol{x}\mathbf{y}\rangle$. If the correlation of $\boldsymbol{x}$ and $\mathbf{y}$ is zero, i.e. $\mathrm{E}\langle\boldsymbol{x}\mathbf{y}\rangle = 0$ then $\boldsymbol{x}$ and $\mathbf{y}$ are orthogonal.

For the discrete r.v. $\boldsymbol{x},\mathbf{y} \in \mathbb{R}^{m\times 1}$, the correlation coefficient has the following property

$$\begin{aligned}\rho_{\boldsymbol{x},\mathbf{y}} &= \frac{\mathrm{cov}\langle\boldsymbol{x},\mathbf{y}\rangle}{\boldsymbol{\sigma}_{\boldsymbol{x}}\boldsymbol{\sigma}_{\mathbf{y}}}\\ &= \frac{\mathrm{E}\langle\boldsymbol{x}\mathbf{y}\rangle - \mathrm{E}\langle\boldsymbol{x}\rangle\,\mathrm{E}\langle\mathbf{y}\rangle}{\boldsymbol{\sigma}_{\boldsymbol{x}}\boldsymbol{\sigma}_{\mathbf{y}}}\end{aligned}$$

where $\boldsymbol{\sigma}_{\boldsymbol{x}} = \sqrt{\mathrm{var}\langle\boldsymbol{x}\rangle}$, $\boldsymbol{\sigma}_{\mathbf{y}} = \sqrt{\mathrm{var}\langle\mathbf{y}\rangle}$, and $-1 \le \rho_{\boldsymbol{x},\mathbf{y}} \le 1$. If $\rho_{\boldsymbol{x},\mathbf{y}} = 0$ then $\boldsymbol{x}$ and $\mathbf{y}$ are said to be *uncorrelated.*

B. Linear Combinations

1) Linear Forms: Assume $\mathbf{X}$ and $\boldsymbol{x}$ to be a matrix and a vector of random variables. Then

$$\begin{aligned}\mathrm{E}\langle\mathbf{AXB}+\mathbf{C}\rangle &= \mathbf{A}\ \mathrm{E}\langle\mathbf{X}\rangle\ \ \mathbf{B}+\mathbf{C}\\ \mathrm{var}\langle\mathbf{A}\boldsymbol{x}\rangle &= \mathbf{A}\ \mathrm{var}\langle\boldsymbol{x}\rangle\ \ \mathbf{A}^\mathsf{T}\\ \mathrm{cov}\langle\mathbf{A}\boldsymbol{x},\mathbf{B}\mathbf{y}\rangle &= \mathbf{A}\ \mathrm{cov}\langle\boldsymbol{x},\mathbf{y}\rangle\ \ \mathbf{B}^\mathsf{T}\end{aligned}$$

Assume $\boldsymbol{x}$ to be a stochastic vector with mean $\mathbf{m}$, then

$$\begin{aligned}\mathrm{E}\langle\mathbf{A}\boldsymbol{x}+\mathbf{b}\rangle &= \mathbf{Am}+\mathbf{b}\\ \mathrm{E}\langle\mathbf{A}\boldsymbol{x}\rangle &= \mathbf{Am}\\ \mathrm{E}\langle\boldsymbol{x}+\mathbf{b}\rangle &= \mathbf{m}+\mathbf{b}\end{aligned}$$

2) Quadratic Forms: Assume $\mathbf{A}$ is symmetric, $\mathbf{c} = \mathrm{E}\langle\boldsymbol{x}\rangle$ and $\boldsymbol{\Sigma} = \mathrm{var}\langle\boldsymbol{x}\rangle$. Assume also that all coordinates x_i are independent, have the same central moments μ_1,μ_2,μ_3,μ_4 and denote $\mathbf{a} = \mathrm{diag}(\mathbf{A})$. Then

$$\begin{aligned}\mathrm{E}\langle\boldsymbol{x}^\mathsf{T}\mathbf{A}\boldsymbol{x}\rangle &= \mathrm{Tr}(\mathbf{A}\boldsymbol{\Sigma})+\mathbf{c}^\mathsf{T}\mathbf{Ac}\\ \mathrm{var}\langle\boldsymbol{x}^\mathsf{T}\mathbf{A}\boldsymbol{x}\rangle &= 2\mu_2^2\mathrm{Tr}(\mathbf{A}^2)+4\mu_2\mathbf{c}^\mathsf{T}\mathbf{A}^2\mathbf{c}\\ &\quad+4\mu_3\mathbf{c}^\mathsf{T}\mathbf{Aa}+(\mu_4-3\mu_2^2)\mathbf{a}^\mathsf{T}\mathbf{a}\end{aligned}$$

Also, assume $\boldsymbol{x}$ to be a stochastic vector with mean $\mathbf{m}$, and covariance $\mathbf{M}$. Then

$$\begin{aligned}\mathrm{E}\langle(\mathbf{A}\boldsymbol{x}+\mathbf{a})(\mathbf{B}\boldsymbol{x}+\mathbf{b})^\mathsf{T}\rangle &= \mathbf{AMB}^\mathsf{T}+(\mathbf{Am}+\mathbf{a})(\mathbf{Bm}+\mathbf{b})^\mathsf{T}\\ \mathrm{E}\langle\boldsymbol{x}\boldsymbol{x}^\mathsf{T}\rangle &= \mathbf{M}+\mathbf{mm}^\mathsf{T}\\ \mathrm{E}\langle\boldsymbol{x}\mathbf{a}^\mathsf{T}\boldsymbol{x}\rangle &= (\mathbf{M}+\mathbf{mm}^\mathsf{T})\mathbf{a}\\ \mathrm{E}\langle\boldsymbol{x}^\mathsf{T}\mathbf{a}\boldsymbol{x}^\mathsf{T}\rangle &= \mathbf{a}^\mathsf{T}(\mathbf{M}+\mathbf{mm}^\mathsf{T})\\ \mathrm{E}\langle(\mathbf{A}\boldsymbol{x})(\mathbf{A}\boldsymbol{x})^\mathsf{T}\rangle &= \mathbf{A}(\mathbf{M}+\mathbf{mm}^\mathsf{T})\mathbf{A}^\mathsf{T}\\ \mathrm{E}\langle(\boldsymbol{x}+\mathbf{a})(\boldsymbol{x}+\mathbf{a})^\mathsf{T}\rangle &= \mathbf{M}+(\mathbf{m}+\mathbf{a})(\mathbf{m}+\mathbf{a})^\mathsf{T}\\ \mathrm{E}\langle(\mathbf{A}\boldsymbol{x}+\mathbf{a})^\mathsf{T}(\mathbf{B}\boldsymbol{x}+\mathbf{b})\rangle &= \mathrm{Tr}(\mathbf{AMB}^\mathsf{T})+(\mathbf{Am}+\mathbf{a})^\mathsf{T}(\mathbf{Bm}+\mathbf{b})\\ \mathrm{E}\langle\boldsymbol{x}^\mathsf{T}\boldsymbol{x}\rangle &= \mathrm{Tr}(\mathbf{M})+\mathbf{m}^\mathsf{T}\mathbf{m}\\ \mathrm{E}\langle\boldsymbol{x}^\mathsf{T}\mathbf{A}\boldsymbol{x}\rangle &= \mathrm{Tr}(\mathbf{AM})+\mathbf{m}^\mathsf{T}\mathbf{Am}\\ \mathrm{E}\langle(\mathbf{A}\boldsymbol{x})^\mathsf{T}(\mathbf{A}\boldsymbol{x})\rangle &= \mathrm{Tr}(\mathbf{AMA}^\mathsf{T})+(\mathbf{Am})^\mathsf{T}(\mathbf{Am})\\ \mathrm{E}\langle(\boldsymbol{x}+\mathbf{a})^\mathsf{T}(\boldsymbol{x}+\mathbf{a})\rangle &= \mathrm{Tr}(\mathbf{M})+(\mathbf{m}+\mathbf{a})^\mathsf{T}(\mathbf{m}+\mathbf{a})\end{aligned}$$

3) Cubic Forms: Assume $\boldsymbol{x}$ to be a stochastic vector with independent coordinates, mean $\mathbf{m}$, covariance $\mathbf{M}$ and central moments $\mathbf{v}_3 = \mathrm{E}\left\langle(\boldsymbol{x}-\mathbf{m})^3\right\rangle$. Then

$$\begin{aligned}&\mathrm{E}\langle(\mathbf{A}\boldsymbol{x}+\mathbf{a})(\mathbf{B}\boldsymbol{x}+\mathbf{b})^\mathsf{T}(\mathbf{C}\boldsymbol{x}+\mathbf{c})\rangle\\ &\quad= \mathbf{A}\ \mathrm{diag}(\mathbf{B}^\mathsf{T}\mathbf{C})\mathbf{v}_3\\ &\qquad+\mathrm{Tr}(\mathbf{BMC}^\mathsf{T})(\mathbf{Am}+\mathbf{a})\\ &\qquad+\mathbf{AMC}^\mathsf{T}(\mathbf{Bm}+\mathbf{b})\\ &\qquad+(\mathbf{AMB}^\mathsf{T}+(\mathbf{Am}+\mathbf{a})(\mathbf{Bm}+\mathbf{b})^\mathsf{T})(\mathbf{Cm}+\mathbf{c})\\ &\mathrm{E}\langle\boldsymbol{x}\boldsymbol{x}^\mathsf{T}\boldsymbol{x}\rangle\\ &\quad= \mathbf{v}_3+2\mathbf{Mm}+(\mathrm{Tr}(\mathbf{M})+\mathbf{m}^\mathsf{T}\mathbf{m})\mathbf{m}\\ &\mathrm{E}\langle(\mathbf{A}\boldsymbol{x}+\mathbf{a})(\mathbf{A}\boldsymbol{x}+\mathbf{a})^\mathsf{T}(\mathbf{A}\boldsymbol{x}+\mathbf{a})\rangle\\ &\quad= \mathbf{A}\ \mathrm{diag}(\mathbf{A}^\mathsf{T}\mathbf{A})\mathbf{v}_3\\ &\qquad+[2\mathbf{AMA}^\mathsf{T}+(\mathbf{A}\boldsymbol{x}+\mathbf{a})(\mathbf{A}\boldsymbol{x}+\mathbf{a})^\mathsf{T}](\mathbf{Am}+\mathbf{a})\\ &\qquad+\mathrm{Tr}(\mathbf{AMA}^\mathsf{T})(\mathbf{Am}+\mathbf{a})\\ &\mathrm{E}\langle(\mathbf{A}\boldsymbol{x}+\mathbf{a})\mathbf{b}^\mathsf{T}(\mathbf{C}\boldsymbol{x}+\mathbf{c})(\mathbf{D}\boldsymbol{x}+\mathbf{d})^\mathsf{T}\rangle\\ &\quad= (\mathbf{A}\boldsymbol{x}+\mathbf{a})\mathbf{b}^\mathsf{T}(\mathbf{CMD}^\mathsf{T}+(\mathbf{Cm}+\mathbf{c})(\mathbf{Dm}+\mathbf{d})^\mathsf{T})\\ &\qquad+(\mathbf{AMC}^\mathsf{T}+(\mathbf{Am}+\mathbf{a})(\mathbf{Cm}+\mathbf{c})^\mathsf{T})\mathbf{b}(\mathbf{Dm}+\mathbf{d})^\mathsf{T}\\ &\qquad+\mathbf{b}^\mathsf{T}(\mathbf{Cm}+\mathbf{c})(\mathbf{AMD}^\mathsf{T}-(\mathbf{Am}+\mathbf{a})(\mathbf{Dm}+\mathbf{d})^\mathsf{T})\end{aligned}$$

VIII. Gaussians

The following section contains some useful relations from statistics theory.

A. Basics

1) Density and normalization: The density of $\boldsymbol{x} \sim \mathcal{N}(\mathbf{m},\boldsymbol{\Sigma})$ is

$$p(\boldsymbol{x}) = \frac{1}{\sqrt{\det(2\pi\boldsymbol{\Sigma})}}\exp\left[-\frac{1}{2}(\boldsymbol{x}-\mathbf{m})^\mathsf{T}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\mathbf{m})\right]$$

Note that if $\boldsymbol{x}$ is d-dimensional, then $\det(2\pi\boldsymbol{\Sigma}) = (2\pi)^d\det(\boldsymbol{\Sigma})$.

Integration and normalization is given by:

$$\begin{aligned}&\int\exp\left[-\frac{1}{2}(\boldsymbol{x}-\mathbf{m})^\mathsf{T}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\mathbf{m})\right]d\boldsymbol{x}\\ &\quad= \sqrt{\det(2\pi\boldsymbol{\Sigma})}\\ &\int\exp\left[-\frac{1}{2}\boldsymbol{x}^\mathsf{T}\mathbf{A}\boldsymbol{x}+\mathbf{b}^\mathsf{T}\boldsymbol{x}\right]d\boldsymbol{x}\\ &\quad= \sqrt{\det(2\pi\mathbf{A}^{-1})}\exp\left[\frac{1}{2}\mathbf{b}^\mathsf{T}\mathbf{A}^{-1}\mathbf{b}\right]\\ &\int\exp\left[-\frac{1}{2}\mathrm{Tr}(\mathbf{S}^\mathsf{T}\mathbf{AS})+\mathrm{Tr}(\mathbf{B}^\mathsf{T}\mathbf{S})\right]d\mathbf{S}\\ &\quad= \sqrt{\det(2\pi\mathbf{A}^{-1})}\exp\left[\frac{1}{2}\mathrm{Tr}(\mathbf{B}^\mathsf{T}\mathbf{A}^{-1}\mathbf{B})\right]\end{aligned}$$

The derivatives of the density are

$$\begin{aligned}\frac{\partial p(\boldsymbol{x})}{\partial\boldsymbol{x}} &= -p(\boldsymbol{x})\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\mathbf{m})\\ \frac{\partial^2 p}{\partial\boldsymbol{x}\partial\boldsymbol{x}^\mathsf{T}} &= p(\boldsymbol{x})\left(\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\mathbf{m})(\boldsymbol{x}-\mathbf{m})^\mathsf{T}\boldsymbol{\Sigma}^{-1}-\boldsymbol{\Sigma}^{-1}\right)\end{aligned}$$

2) Marginal Distribution: Assume $\boldsymbol{x} \sim \mathcal{N}_{\boldsymbol{x}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_a \\ \boldsymbol{x}_b \end{bmatrix}$$
$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{bmatrix}$$
$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_a & \boldsymbol{\Sigma}_c \\ \boldsymbol{\Sigma}_c^\intercal & \boldsymbol{\Sigma}_b \end{bmatrix}$$

then

$$p(\boldsymbol{x}_a) = \mathcal{N}_{\boldsymbol{x}_a}(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a)$$
$$p(\boldsymbol{x}_b) = \mathcal{N}_{\boldsymbol{x}_b}(\boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b)$$

3) Conditional Distribution: Assume $\boldsymbol{x} \sim \mathcal{N}_{\boldsymbol{x}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{x}_a \\ \boldsymbol{x}_b \end{bmatrix}$$
$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{bmatrix}$$
$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_a & \boldsymbol{\Sigma}_c \\ \boldsymbol{\Sigma}_c^\intercal & \boldsymbol{\Sigma}_b \end{bmatrix}$$

then

$$p(\boldsymbol{x}_a|\boldsymbol{x}_b) = \mathcal{N}_{\boldsymbol{x}_a}(\hat{\boldsymbol{\mu}}_a, \hat{\boldsymbol{\Sigma}}_a)$$
$$p(\boldsymbol{x}_b|\boldsymbol{x}_a) = \mathcal{N}_{\boldsymbol{x}_b}(\hat{\boldsymbol{\mu}}_b, \hat{\boldsymbol{\Sigma}}_b)$$

where

$$\hat{\boldsymbol{\mu}}_a = \boldsymbol{\mu}_a + \boldsymbol{\Sigma}_c \boldsymbol{\Sigma}_b^{-1}(\boldsymbol{x}_b - \boldsymbol{\mu}_b)$$
$$\hat{\boldsymbol{\Sigma}}_a = \boldsymbol{\Sigma}_a - \boldsymbol{\Sigma}_c \boldsymbol{\Sigma}_b^{-1} \boldsymbol{\Sigma}_c^\intercal$$
$$\hat{\boldsymbol{\mu}}_b = \boldsymbol{\mu}_b + \boldsymbol{\Sigma}_c \boldsymbol{\Sigma}_a^{-1}(\boldsymbol{x}_a - \boldsymbol{\mu}_a)$$
$$\hat{\boldsymbol{\Sigma}}_b = \boldsymbol{\Sigma}_b - \boldsymbol{\Sigma}_c \boldsymbol{\Sigma}_a^{-1} \boldsymbol{\Sigma}_c^\intercal$$

4) Linear combination: Assume $\boldsymbol{x} \sim \mathcal{N}(\mathbf{m}_x, \boldsymbol{\Sigma}_x)$ and $\mathbf{y} \sim \mathcal{N}(\mathbf{m}_y, \boldsymbol{\Sigma}_y)$

$$\mathbf{A}\boldsymbol{x} + \mathbf{B}\mathbf{y} + \mathbf{c} \sim \mathcal{N}(\mathbf{A}\mathbf{m}_x + \mathbf{B}\mathbf{m}_y + \mathbf{c},\ \mathbf{A}\boldsymbol{\Sigma}_x\mathbf{A}^\intercal + \mathbf{B}\boldsymbol{\Sigma}_y\mathbf{B}^\intercal)$$

5) Rearranging Means:

$$\mathcal{N}_{\mathbf{A}\boldsymbol{x}}\left[\mathbf{m}, \boldsymbol{\Sigma}\right] = \frac{\sqrt{\det(2\pi(\mathbf{A}^\intercal\boldsymbol{\Sigma}^{-1}\mathbf{A})^{-1})}}{\sqrt{\det(2\pi\boldsymbol{\Sigma})}} \mathcal{N}_{\boldsymbol{x}}\left[\mathbf{A}^{-1}\mathbf{m},\ (\mathbf{A}^\intercal\boldsymbol{\Sigma}^{-1}\mathbf{A})^{-1}\right]$$

6) Rearranging into squared form: If $\mathbf{A}$ is symmetric, then

$$-\frac{1}{2}\boldsymbol{x}^\intercal\mathbf{A}\boldsymbol{x} + \mathbf{b}^\intercal\boldsymbol{x} = -\frac{1}{2}(\boldsymbol{x} - \mathbf{A}^{-1}\mathbf{b})^\intercal\mathbf{A}(\boldsymbol{x} - \mathbf{A}^{-1}\mathbf{b})^\intercal + \frac{1}{2}\mathbf{b}^\intercal\mathbf{A}^{-1}\mathbf{b}$$

$$-\frac{1}{2}\mathrm{Tr}(\mathbf{X}^\intercal\mathbf{A}\mathbf{X}) + Tr(\mathbf{B}^\intercal\mathbf{X}) = -\frac{1}{2}[(\mathbf{X} - \mathbf{A}^{-1}\mathbf{B})^\intercal\mathbf{A}(\mathbf{X} - \mathbf{A}^{-1}\mathbf{B})] + \frac{1}{2}\mathrm{Tr}(\mathbf{B}^\intercal\mathbf{A}^{-1}\mathbf{B})$$

7) Sum of two squared forms: In vector form (assuming $\boldsymbol{\Sigma}_1$, $\boldsymbol{\Sigma}_2$ are symmetric)

$$\begin{aligned}
&-\frac{1}{2}(\boldsymbol{x} - \mathbf{m}_1)^\intercal\boldsymbol{\Sigma}_1^{-1}(\boldsymbol{x} - \mathbf{m}_1) \\
&-\frac{1}{2}(\boldsymbol{x} - \mathbf{m}_2)^\intercal\boldsymbol{\Sigma}_2^{-1}(\boldsymbol{x} - \mathbf{m}_2) \\
&= -\frac{1}{2}(\boldsymbol{x} - \mathbf{m}_c)^\intercal\boldsymbol{\Sigma}_c^{-1}(\boldsymbol{x} - \mathbf{m}_c) + C
\end{aligned}$$
$$\boldsymbol{\Sigma}_c^{-1} = \boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1}$$
$$\mathbf{m}_c = (\boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1})^{-1}(\boldsymbol{\Sigma}_1^{-1}\mathbf{m}_1 + \boldsymbol{\Sigma}_2^{-1}\mathbf{m}_2)$$
$$C = \frac{1}{2}(\mathbf{m}_1^\intercal\boldsymbol{\Sigma}_1^{-1} + \mathbf{m}_2^\intercal\boldsymbol{\Sigma}_2^{-1})(\boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1})^{-1} (\boldsymbol{\Sigma}_1^{-1}\mathbf{m}_1 + \boldsymbol{\Sigma}_2^{-1}\mathbf{m}_2) - \frac{1}{2}(\mathbf{m}_1^\intercal\boldsymbol{\Sigma}_1^{-1}\mathbf{m}_1 + \mathbf{m}_2^\intercal\boldsymbol{\Sigma}_2^{-1}\mathbf{m}_2)$$

In a trace form (assuming $\boldsymbol{\Sigma}_1$, $\boldsymbol{\Sigma}_2$ are symmetric)

$$\begin{aligned}
&-\frac{1}{2}\mathrm{Tr}\left[(\mathbf{X} - \mathbf{M}_1)^\intercal\boldsymbol{\Sigma}_1^{-1}(\mathbf{X} - \mathbf{M}_1)\right] \\
&-\frac{1}{2}\mathrm{Tr}\left[(\mathbf{X} - \mathbf{M}_2)^\intercal\boldsymbol{\Sigma}_2^{-1}(\mathbf{X} - \mathbf{M}_2)\right] \\
&= -\frac{1}{2}\mathrm{Tr}\left[(\mathbf{X} - \mathbf{M}_c)^\intercal\boldsymbol{\Sigma}_c^{-1}(\mathbf{X} - \mathbf{M}_c)\right] + C
\end{aligned}$$
$$\boldsymbol{\Sigma}_c^{-1} = \boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1}$$
$$\mathbf{M}_c = (\boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1})^{-1}(\boldsymbol{\Sigma}_1^{-1}\mathbf{M}_1 + \boldsymbol{\Sigma}_2^{-1}\mathbf{M}_2)$$
$$C = \frac{1}{2}\mathrm{Tr}[(\mathbf{M}_1^\intercal\boldsymbol{\Sigma}_1^{-1} + \mathbf{M}_2^\intercal\boldsymbol{\Sigma}_2^{-1})(\boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1})^{-1} (\boldsymbol{\Sigma}_1^{-1}\mathbf{M}_1 + \boldsymbol{\Sigma}_2^{-1}\mathbf{M}_2)] - \frac{1}{2}(\mathbf{M}_1^\intercal\boldsymbol{\Sigma}_1^{-1}\mathbf{M}_1 + \mathbf{M}_2^\intercal\boldsymbol{\Sigma}_2^{-1}\mathbf{M}_2)$$

8) Product of gaussian densities: Let $\mathcal{N}_{\boldsymbol{x}}(\mathbf{m}, \boldsymbol{\Sigma})$ denote a density of $\boldsymbol{x}$, then

$$\mathcal{N}_{\boldsymbol{x}}(\mathbf{m}_1, \boldsymbol{\Sigma}_1) \cdot \mathcal{N}_{\boldsymbol{x}}(\mathbf{m}_2, \boldsymbol{\Sigma}_2) = c_c\mathcal{N}_{\boldsymbol{x}}(\mathbf{m}_c, \boldsymbol{\Sigma}_c)$$

$$\begin{aligned}
c_c &= \mathcal{N}_{\mathbf{m}_1}(\mathbf{m}_1, (\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)) \\
&= \frac{1}{\sqrt{\det(2\pi(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2))}} \exp\left[-\frac{1}{2}(\mathbf{m}_1 - \mathbf{m}_2)^\intercal(\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)^{-1}(\mathbf{m}_1 - \mathbf{m}_2)\right]
\end{aligned}$$
$$\mathbf{m}_c = (\boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1})^{-1}(\boldsymbol{\Sigma}_1^{-1}\mathbf{m}_1 + \boldsymbol{\Sigma}_2^{-1}\mathbf{m}_2)$$
$$\boldsymbol{\Sigma}_c = (\boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1})^{-1}$$

B. Moments

1) Mean and covariance of linear forms: First and second moments. Assume $\boldsymbol{x} \sim \mathcal{N}(\mathbf{m}, \boldsymbol{\Sigma})$

$$\mathrm{E}\left\langle\boldsymbol{x}\right\rangle = \mathbf{m} \tag{206}$$

then

$$\begin{aligned}
\mathrm{Cov}\left\langle\boldsymbol{x}, \boldsymbol{x}\right\rangle &= \mathrm{Var}\left\langle\boldsymbol{x}\right\rangle \\
&= \boldsymbol{\Sigma} \\
&= \mathrm{E}\left\langle\boldsymbol{x}\boldsymbol{x}^\intercal\right\rangle - \mathrm{E}\left\langle\boldsymbol{x}\right\rangle\mathrm{E}\left\langle\boldsymbol{x}^\intercal\right\rangle \\
&= \mathrm{E}\left\langle\boldsymbol{x}\boldsymbol{x}^\intercal\right\rangle - \mathbf{m}\mathbf{m}^\intercal
\end{aligned}$$

As for any other distribution is holds for gaussians that

$$\begin{aligned}
\mathrm{E}\langle \mathbf{A}\boldsymbol{x}\rangle &= \mathbf{A}\mathrm{E}\langle \boldsymbol{x}\rangle \\
\mathrm{Var}\langle \mathbf{A}\boldsymbol{x}\rangle &= \mathbf{A}\mathrm{Var}\langle \boldsymbol{x}\rangle \mathbf{A}^\intercal \\
\mathrm{Cov}\langle \mathbf{A}\boldsymbol{x}, \mathbf{B}\mathbf{y}\rangle &= \mathbf{A}\mathrm{Cov}\langle \boldsymbol{x}, \mathbf{y}\rangle \mathbf{B}^\intercal
\end{aligned}$$

2) Mean and variance of square forms: Mean and variance of square forms. Assume $\boldsymbol{x} \sim \mathcal{N}(\mathbf{m}, \boldsymbol{\Sigma})$

$$\begin{aligned}
\mathrm{E}\langle \boldsymbol{x}\boldsymbol{x}^\intercal\rangle &= \boldsymbol{\Sigma} + \mathbf{m}\mathbf{m}^\intercal \\
\mathrm{E}\langle \boldsymbol{x}^\intercal\mathbf{A}\boldsymbol{x}\rangle &= \mathrm{Tr}(\mathbf{A}\boldsymbol{\Sigma}) + \mathbf{m}^\intercal\mathbf{A}\mathbf{m} \\
\mathrm{Var}\langle \boldsymbol{x}^\intercal\mathbf{A}\boldsymbol{x}\rangle &= 2\sigma^4\mathrm{Tr}(\mathbf{A}^2) + 4\sigma^2\mathbf{m}^\intercal\mathbf{A}^2\mathbf{m} \\
\mathrm{Cov}\langle (\boldsymbol{x}-\mathbf{m}')^\intercal\mathbf{A}(\boldsymbol{x}-\mathbf{m}')\rangle &= (\boldsymbol{x}-\mathbf{m}')^\intercal\mathbf{A}(\boldsymbol{x}-\mathbf{m}') + \mathrm{Tr}(\mathbf{A}\boldsymbol{\Sigma})
\end{aligned}$$

Assume $\boldsymbol{x} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ and $\mathbf{A}$ and $\mathbf{B}$ to be symmetric, then

$$\mathrm{Cov}\langle \boldsymbol{x}^\intercal\mathbf{A}\boldsymbol{x}, \boldsymbol{x}^\intercal\mathbf{B}\boldsymbol{x}\rangle = 2\sigma^4\mathrm{Tr}(\mathbf{A}\mathbf{B})$$

3) Cubic forms:

$$\begin{aligned}
\mathrm{E}\langle \boldsymbol{x}\mathbf{b}^\intercal\boldsymbol{x}\boldsymbol{x}^\intercal\rangle = \mathbf{m}\mathbf{b}^\intercal(\mathbf{M}+\mathbf{m}\mathbf{m}^\intercal) \\
+ (\mathbf{M}+\mathbf{m}\mathbf{m}^\intercal)\mathbf{b}\mathbf{m}^\intercal \\
+ \mathbf{b}^\intercal\mathbf{m}(\mathbf{M}-\mathbf{m}\mathbf{m}^\intercal)
\end{aligned}$$

4) Moments:

$$\begin{aligned}
\mathrm{E}\langle \boldsymbol{x}\rangle &= \sum_k \rho_k \mathbf{m}_k \\
\mathrm{Cov}\langle \boldsymbol{x}\rangle &= \sum_k\sum_{k'} \rho_k\rho_{k'}(\boldsymbol{\Sigma}_k + \mathbf{m}_k\mathbf{m}_k^\intercal + \mathbf{m}_{k'}\mathbf{m}_{k'}^\intercal)
\end{aligned}$$

C. Whitening

Assume $\boldsymbol{x} \sim \mathcal{N}(\mathbf{m}, \boldsymbol{\Sigma})$, then

$$\mathbf{z} = \boldsymbol{\Sigma}^{-1/2}(\boldsymbol{x}-\mathbf{m}) \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

Conversely having $\boldsymbol{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ one can generate data $\boldsymbol{x} \sim \mathcal{N}(\mathbf{m}, \boldsymbol{\Sigma})$ by setting

$$\boldsymbol{x} = \boldsymbol{\Sigma}^{1/2}\mathbf{z} + \mathbf{m} \sim \mathcal{N}(\mathbf{m}, \boldsymbol{\Sigma})$$

Note that $\boldsymbol{\Sigma}^{1/2}$ means the matrix which fulfills $\boldsymbol{\Sigma}^{1/2}\boldsymbol{\Sigma}^{1/2} = \boldsymbol{\Sigma}$, and that it exists and is unique since $\boldsymbol{\Sigma}$ is positive definite.

D. Chi-Square

Assume $\boldsymbol{x} \sim \mathcal{N}(\mathbf{m}, \boldsymbol{\Sigma})$ and $\boldsymbol{x}$ to be n dimensional, then

$$z = (\boldsymbol{x}-\mathbf{m})^\intercal\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\mathbf{m}) \sim \boldsymbol{\chi}_n^2$$

where $\boldsymbol{\chi}_n^2$ denotes the Chi square distribution with n degrees of freedom.

E. Mixture of Gaussians

1) Density: The variable $\boldsymbol{x}$ is distributed as a mixture of gaussians if it has the density

$$p(\boldsymbol{x}) = \sum_{k=1}^K \rho_k \frac{1}{\sqrt{\det(2\pi\boldsymbol{\Sigma}_k)}} \exp\left[-\frac{1}{2}(\boldsymbol{x}-\mathbf{m}_k)^\intercal\boldsymbol{\Sigma}_k^{-1}(\boldsymbol{x}-\mathbf{m}_k)\right]$$

where ρ_k sum to 1 and the $\boldsymbol{\Sigma}_k$ all are positive definite.

2) Derivatives: Defining $p(s) = \sum_k \rho_k\mathcal{N}_s(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$, then

$$\begin{aligned}
\frac{\partial \ln p(s)}{\partial \rho_j} &= \frac{\rho_j\mathcal{N}_s(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}{\sum_k \rho_k\mathcal{N}_s(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)} \frac{\partial}{\partial \rho_j} \ln[\rho_j\mathcal{N}_s(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)] \\
&= \frac{\rho_j\mathcal{N}_s(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}{\sum_k \rho_k\mathcal{N}_s(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)} \frac{1}{\rho_j} \\
\frac{\partial \ln p(s)}{\partial \boldsymbol{\mu}_j} &= \frac{\rho_j\mathcal{N}_s(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}{\sum_k \rho_k\mathcal{N}_s(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)} \frac{\partial}{\partial \boldsymbol{\mu}_j} \ln[\rho_j\mathcal{N}_s(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)] \\
&= \frac{\rho_j\mathcal{N}_s(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}{\sum_k \rho_k\mathcal{N}_s(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)} [-\boldsymbol{\Sigma}_k^{-1}(s-\boldsymbol{\mu}_k)] \\
\frac{\partial \ln p(s)}{\partial \boldsymbol{\Sigma}_j} &= \frac{\rho_j\mathcal{N}_s(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}{\sum_k \rho_k\mathcal{N}_s(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)} \frac{\partial}{\partial \boldsymbol{\Sigma}_j} \ln[\rho_j\mathcal{N}_s(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)] \\
&= \frac{\rho_j\mathcal{N}_s(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}{\sum_k \rho_k\mathcal{N}_s(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)} \\
&\quad \frac{1}{2}[\boldsymbol{\Sigma}_j^{-\intercal} + \boldsymbol{\Sigma}_j^{-\intercal}(s-\boldsymbol{\mu}_j)(s-\boldsymbol{\mu}_j)^\intercal\boldsymbol{\Sigma}_j^{-\intercal}]
\end{aligned}$$

Note, ρ_k and $\boldsymbol{\Sigma}_k$ must be constrained.

IX. Least Squares

Consider the general measurement equation

$$\mathbf{y} = \mathbf{H}\boldsymbol{x} + \boldsymbol{\eta} + \mathbf{e} \tag{207}$$

where $\mathbf{y} \in \mathbb{R}^{m\times 1}$, $\mathbf{H} \in \mathbb{R}^{m\times n}$ where $m > n$ and $\mathrm{rank}(\mathbf{H}) = n$, $\boldsymbol{x} \in \mathbb{R}^{n\times 1}$, with Gaussian noise $\boldsymbol{\eta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\sigma}^2\mathbf{I}) \in \mathbb{R}^{n\times 1}$, and deterministic errors $\mathbf{e} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2\mathbf{I}) \in \mathbb{R}^{n\times 1}$.

Ignoring the noise and error vectors, the estimate of $\boldsymbol{x}$ is found by

$$\begin{aligned}
\mathbf{J}_{LS}(\hat{\boldsymbol{x}}) &= \frac{1}{2}(\mathbf{y}-\mathbf{H}\hat{\boldsymbol{x}})^\intercal(\mathbf{y}-\mathbf{H}\hat{\boldsymbol{x}}) && (208) \\
&= \frac{1}{2}(\mathbf{y}^\intercal\mathbf{y} - 2\mathbf{y}^\intercal\mathbf{H}\hat{\boldsymbol{x}} + \boldsymbol{x}^\intercal\mathbf{H}^\intercal\mathbf{H}\hat{\boldsymbol{x}}) && (209) \\
\frac{\partial \mathbf{J}_{LS}(\hat{\boldsymbol{x}})}{\partial \hat{\boldsymbol{x}}} &= -\mathbf{H}^\intercal\mathbf{y} + \mathbf{H}^\intercal\mathbf{H}\hat{\boldsymbol{x}} = 0 && (210) \\
\hat{\boldsymbol{x}} &= (\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal\mathbf{y} && (211) \\
&= \bar{\mathbf{H}}\mathbf{y} && (212)
\end{aligned}$$

where $\bar{\mathbf{H}} \triangleq (\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal$ is the generalized inverse, also known as the "Moore-Penrose pseudo-inverse". Note that $\bar{\mathbf{H}}$ transforms the measurement space to the state space. If $\mathbf{H}$ is full column-rank, then $\mathbf{H}$ has the following property

$$\bar{\mathbf{H}}\mathbf{H} = (\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal\mathbf{H} = \mathbf{I}_m \tag{213}$$

This is because $\mathbf{H}^\intercal\mathbf{H} \in \mathbb{R}^{m\times m}$ with $\mathrm{rank}(\mathbf{H}^\intercal\mathbf{H}) = m$, and therefore nonsingular. Then by the linear algebra property for the general matrix $\mathbf{A} \in \mathbb{R}^{m\times m}$ with $\mathrm{rank}(\mathbf{A}) = m$, the property $\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}_m$ is applied in eqn. (213).

By analysis, the estimate $\hat{\boldsymbol{x}}$ is

$$\begin{aligned}
\hat{\boldsymbol{x}} &= \bar{\mathbf{H}}\mathbf{y} && (214) \\
&= \bar{\mathbf{H}}(\mathbf{H}\boldsymbol{x} + \boldsymbol{\eta} + \mathbf{e}) && (215)
\end{aligned}$$

The estimation error is

$$\begin{aligned}
\delta\boldsymbol{x} &= \boldsymbol{x} - \hat{\boldsymbol{x}} && (216) \\
&= \boldsymbol{x} - \bar{\mathbf{H}}(\mathbf{H}\boldsymbol{x} + \boldsymbol{\eta} + \mathbf{e}) && (217) \\
&= -\bar{\mathbf{H}}(\boldsymbol{\eta} + \mathbf{e}) && (218)
\end{aligned}$$

The measurement estimate is

$$\begin{aligned}
\hat{\mathbf{y}} &= \mathbf{H}\hat{\boldsymbol{x}} && (219)\\
&= \mathbf{H}(\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal\mathbf{y} && (220)\\
&= \mathbf{H}(\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal(\mathbf{H}\boldsymbol{x}+\boldsymbol{\eta}+\mathbf{e}) && (221)\\
&= \mathbf{P}\mathbf{H}\boldsymbol{x}+\mathbf{P}(\boldsymbol{\eta}+\mathbf{e}) && (222)\\
&= \mathbf{H}\boldsymbol{x}+\mathbf{P}(\boldsymbol{\eta}+\mathbf{e}) && (223)
\end{aligned}$$

where the projection matrix $\mathbf{P} \triangleq \mathbf{H}(\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal$.
The measurement residual is

$$\begin{aligned}
\mathbf{r} &= \mathbf{y}-\hat{\mathbf{y}} && (224)\\
&= (\mathbf{H}\boldsymbol{x}+\boldsymbol{\eta}+\mathbf{e})-\mathbf{H}\boldsymbol{x}-\mathbf{H}(\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal(\boldsymbol{\eta}+\mathbf{e}) && (225)\\
&= (\mathbf{I}_m-\mathbf{H}(\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal)(\boldsymbol{\eta}+\mathbf{e}) && (226)\\
&= (\mathbf{I}_m-\mathbf{P})(\boldsymbol{\eta}+\mathbf{e}) && (227)\\
&= \mathbf{Q}(\boldsymbol{\eta}+\mathbf{e}) && (228)\\
&= \mathbf{Q}\boldsymbol{\eta}+\mathbf{Q}\mathbf{e} && (229)
\end{aligned}$$

where the orthogonal projection matrix $\mathbf{Q} \triangleq (\mathbf{I}_m-\mathbf{P})$.

Projection matrices $\mathbf{P}$ and $\mathbf{Q}$ are both idempotent, and have rank n and $m-n$ respectively. The proofs for idempotent and rank are presented in Section XI.

X. Weighted Least Squares

Consider the general measurement equation

$$\mathbf{y} = \mathbf{H}\boldsymbol{x}+\boldsymbol{\nu} \tag{230}$$

where $\mathbf{y} \in \mathbb{R}^{m\times 1}$, $\mathbf{H} \in \mathbb{R}^{m\times n}$ where $m > n$ and $\text{rank}(\mathbf{H}) = n$, $\boldsymbol{x} \in \mathbb{R}^{n\times 1}$, with Gaussian noise $\boldsymbol{\nu} \sim \mathcal{N}(\mathbf{0},\boldsymbol{\sigma}^2\mathbf{I}) \in \mathbb{R}^{n\times 1}$.

Ignoring the noise, the estimate of $\boldsymbol{x}$ is found by

$$\begin{aligned}
\mathbf{J}_{WLS}(\hat{\boldsymbol{x}}) &= \frac{1}{2}(\mathbf{y}-\mathbf{H}\hat{\boldsymbol{x}})^\intercal\mathbf{W}(\mathbf{y}-\mathbf{H}\hat{\boldsymbol{x}}) && (231)\\
&= \frac{1}{2}(\mathbf{y}^\intercal\mathbf{W}\mathbf{y}-2\mathbf{y}^\intercal\mathbf{W}\mathbf{H}\hat{\boldsymbol{x}}+\boldsymbol{x}^\intercal\mathbf{H}^\intercal\mathbf{W}\mathbf{H}\hat{\boldsymbol{x}}) && (232)\\
\frac{\partial\mathbf{J}_{WLS}(\hat{\boldsymbol{x}})}{\partial\hat{\boldsymbol{x}}} &= -\mathbf{H}^\intercal\mathbf{W}\mathbf{y}+\mathbf{H}^\intercal\mathbf{W}\mathbf{H}\hat{\boldsymbol{x}} = 0 && (233)\\
\hat{\boldsymbol{x}} &= (\mathbf{H}^\intercal\mathbf{W}\mathbf{H})^{-1}\mathbf{H}^\intercal\mathbf{W}\mathbf{y} && (234)
\end{aligned}$$

where $\mathbf{W} \in \mathbb{R}^{m\times m}$ is the weighting matrix.

The estimation error is

$$\begin{aligned}
\delta\boldsymbol{x} &= \boldsymbol{x}-\hat{\boldsymbol{x}} && (235)\\
&= \boldsymbol{x}-(\mathbf{H}^\intercal\mathbf{W}\mathbf{H})^{-1}\mathbf{H}^\intercal\mathbf{W}\mathbf{y} && (236)\\
&= \boldsymbol{x}-(\mathbf{H}^\intercal\mathbf{W}\mathbf{H})^{-1}\mathbf{H}^\intercal\mathbf{W}(\mathbf{H}\boldsymbol{x}+\boldsymbol{\nu}) && (237)\\
&= \left(\mathbf{I}-(\mathbf{H}^\intercal\mathbf{W}\mathbf{H})^{-1}\mathbf{H}^\intercal\mathbf{W}\mathbf{H}\right)\boldsymbol{x}\\
&\quad -(\mathbf{H}^\intercal\mathbf{W}\mathbf{H})^{-1}\mathbf{H}^\intercal\mathbf{W}\boldsymbol{\nu} && (238)\\
&= -(\mathbf{H}^\intercal\mathbf{W}\mathbf{H})^{-1}\mathbf{H}^\intercal\mathbf{W}\boldsymbol{\nu} && (239)
\end{aligned}$$

For $\boldsymbol{\nu} \sim \mathcal{N}(\mathbf{0},\boldsymbol{\sigma}^2\mathbf{I})$, $\mathbf{W} = \mathbf{R}^{-1}$

$$\begin{aligned}
\text{E}\langle\delta\boldsymbol{x}\rangle &= \mathbf{0} && (240)\\
\text{var}\langle\delta\boldsymbol{x}\rangle &= (\mathbf{H}^\intercal\mathbf{W}\mathbf{H})^{-1}\mathbf{H}^\intercal\mathbf{W}\mathbf{R}\mathbf{W}\mathbf{H}(\mathbf{H}^\intercal\mathbf{W}\mathbf{H})^{-1} && (241)
\end{aligned}$$

For $\mathbf{W} = \mathbf{I}_m$, the Least Squares (LS) estimate results

$$\begin{aligned}
\hat{\boldsymbol{x}} &= (\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal\mathbf{y} && (242)\\
\text{E}\langle\delta\boldsymbol{x}\rangle &= \mathbf{0} && (243)\\
\text{var}\langle\delta\boldsymbol{x}\rangle &= (\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal\mathbf{R}\mathbf{H}(\mathbf{H}^\intercal\mathbf{H})^{-1} && (244)
\end{aligned}$$

For $\mathbf{W} = \mathbf{R}^{-1}$, the Maximum Likelihood Estimate (MLE) results

$$\begin{aligned}
\hat{\boldsymbol{x}} &= (\mathbf{H}^\intercal\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^\intercal\mathbf{R}^{-1}\mathbf{y} && (245)\\
\text{E}\langle\delta\boldsymbol{x}\rangle &= \mathbf{0} && (246)\\
\text{var}\langle\delta\boldsymbol{x}\rangle &= (\mathbf{H}^\intercal\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^\intercal\mathbf{R}^{-1}\mathbf{R}\mathbf{R}^{-1}\mathbf{H}(\mathbf{H}^\intercal\mathbf{R}^{-1}\mathbf{H})^{-1} && (247)\\
&= (\mathbf{H}^\intercal\mathbf{R}^{-1}\mathbf{H})^{-1} && (248)\\
&= \mathbf{C} && (249)
\end{aligned}$$

where $\mathbf{C}$ is the covariance matrix, and $\mathbf{C}^{-1} = \mathbf{H}^\intercal\mathbf{R}^{-1}\mathbf{H}$ is the information matrix.

XI. Proof of Matrix Rank Using the SVD

A. Proof of idempotent P

For the matrix $\mathbf{P}$ to be idempotent, it must be the case that $\mathbf{P} = \mathbf{P}^\intercal\mathbf{P} = \mathbf{P}\mathbf{P}$, where $\mathbf{P} \triangleq \mathbf{H}(\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal$, and $\mathbf{H} \in \mathbb{R}^{m\times n}$, with $m > n$. Thus we can show:

$$\begin{aligned}
\mathbf{P}^\intercal &= (\mathbf{H}(\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal)^\intercal && (250)\\
&= \mathbf{H}(\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal && (251)\\
&= \mathbf{P} && (252)\\
\mathbf{P}\mathbf{P} &= \mathbf{H}(\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal\mathbf{H}(\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal && (253)\\
&= \mathbf{H}(\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal && (254)\\
&= \mathbf{P} && (255)\\
\therefore \mathbf{P}^\intercal\mathbf{P} &= \mathbf{P}\mathbf{P} = \mathbf{P}. && (256)
\end{aligned}$$

■

B. Proof of rank P

We can prove that $\text{rank}(\mathbf{P}) = n$. First recall that $\mathbf{H} \in \mathbb{R}^{m\times n}$, with $m > n$ and full column rank, i.e. $\text{rank}(\mathbf{H}) = n$. Let the SVD of $\mathbf{H}$ be defined as

$$\begin{aligned}
\mathbf{H} &= \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\intercal && (257)\\
&= [\mathbf{U}_1,\mathbf{U}_2]\left[\begin{pmatrix}\boldsymbol{\Sigma}_1\\ \boldsymbol{\Sigma}_0\end{pmatrix}\right]\mathbf{V}^\intercal && (258)
\end{aligned}$$

where $\boldsymbol{\Sigma} \in \mathbb{R}^{m\times m}$, $\boldsymbol{\Sigma}_1 = \text{diag}(\sigma_1,\ldots,\sigma_n) \in \mathbb{R}^{n\times n}$, and $\boldsymbol{\Sigma}_0 = \mathbf{0} \in \mathbb{R}^{(m-n)\times n}$, where σ_i for $i = 1,\ldots,n$ are the singular values of $\mathbf{H}$. Both $\mathbf{U} \in \mathbb{R}^{m\times m}$ and $\mathbf{V} \in \mathbb{R}^{n\times n}$ are unitary matrices, therefore $\mathbf{U}\mathbf{U}^\intercal = \mathbf{U}^\intercal\mathbf{U} = \mathbf{I} \in \mathbb{R}^{m\times m}$ and $\mathbf{V}\mathbf{V}^\intercal = \mathbf{V}^\intercal\mathbf{V} = \mathbf{I} \in \mathbb{R}^{m\times m}$. The columns of $\mathbf{U}_1 \in \mathbb{R}^{m\times n}$ form an orthonormal basis for the range-space of $\mathbf{H}$, and the columns of $\mathbf{U}_2 \in \mathbb{R}^{m\times(m-n)}$ form the null-space of $\mathbf{H}^\intercal$. Similarly the first n columns of $\mathbf{V}$ form an orthonormal basis for the range of $\mathbf{H}^\intercal$, and the $m-n$ columns of $\mathbf{V}$ form an orthonormal basis for the null-space of $\mathbf{H}$. Finally, the eigenvectors $\mathbf{V}$ of the matrix $\mathbf{H}^\intercal\mathbf{H}$ are the right singular values of $\mathbf{H}$, and the singular values of $\mathbf{H}$ squared are the corresponding nonzero eigenvalues: $\sigma_i = \sqrt{\lambda_i(\mathbf{H}^\intercal\mathbf{H})}$.

Similarly, the eigenvectors of $\mathbf{H}\mathbf{H}^\intercal$ are the left singular vectors $\mathbf{U}$ of matrix $\mathbf{H}$, and the singular values of $\mathbf{H}$ squared are the nonzero eigenvalues of $\mathbf{H}\mathbf{H}^\intercal$: $\sigma_i = \sqrt{\lambda_i(\mathbf{H}\mathbf{H}^\intercal)}$.

Define $\mathbf{P}$ in terms of the SVD of $\mathbf{H}$:

$$\begin{aligned}
\mathbf{P} &= \mathbf{H}(\mathbf{H}^\intercal\mathbf{H})^{-1}\mathbf{H}^\intercal && (259)\\
&= (\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\intercal)(\mathbf{V}\boldsymbol{\Sigma}^\intercal\mathbf{U}^\intercal\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^t)^{-1}(\mathbf{V}\boldsymbol{\Sigma}^\intercal\mathbf{U}^\intercal) && (260)\\
&= (\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\intercal)(\mathbf{V}\boldsymbol{\Sigma}^\intercal\boldsymbol{\Sigma}\mathbf{V}^\intercal)^{-1}(\mathbf{V}\boldsymbol{\Sigma}^\intercal\mathbf{U}^\intercal) && (261)\\
&= (\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\intercal)(\mathbf{V}\boldsymbol{\Sigma}_1^2\mathbf{V}^\intercal)^{-1}(\mathbf{V}\boldsymbol{\Sigma}^\intercal\mathbf{U}^\intercal) && (262)\\
&= (\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\intercal)(\mathbf{V})^{-1}(\boldsymbol{\Sigma}_1^2)^{-1}(\mathbf{V}^\intercal)^{-1}(\mathbf{V}\boldsymbol{\Sigma}^\intercal\mathbf{U}^\intercal) && (263)\\
&= \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\intercal\mathbf{V}\boldsymbol{\Sigma}_1^{-2}\mathbf{V}^\intercal\mathbf{V}\boldsymbol{\Sigma}^\intercal\mathbf{U}^\intercal && (264)\\
&= \mathbf{U}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_1^{-2}\boldsymbol{\Sigma}_1^\intercal\mathbf{U}^\intercal && (265)\\
&= \mathbf{U}\boldsymbol{\Sigma}_1\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\Sigma}_1^\intercal\mathbf{U}^\intercal && (266)\\
&= \mathbf{U}\mathbf{I}_{n\times n}\mathbf{U}^\intercal && (267)\\
&= [\mathbf{U}_1\mathbf{U}_2]\begin{bmatrix}\mathbf{I}_{n\times n} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{U}_1^\intercal\\ \mathbf{U}_2^\intercal\end{bmatrix} && (268)\\
&= \mathbf{U}_1\mathbf{U}_1^\intercal. && (269)
\end{aligned}$$

The middle product in eqn. (262) can be separated because it is an $n \times n$ matrix with rank n, and it is non-singular. In eqns. (262)-(266), we need only consider $\boldsymbol{\Sigma}_1$ as $\boldsymbol{\Sigma}_0$ drops out.

The rank of matrix $\mathbf{P}$ is defined as the number of non-zero singular values of $\mathbf{P}$. Thus, rank$(\mathbf{P}) = n$. Similarly, because $\mathbf{P}$ is idempotent, rank$(\mathbf{P}) = tr(\mathbf{P})$, then rank$(\mathbf{P}) = n$.

■

C. Proof of idempotent Q

For the matrix $\mathbf{Q}$ to be idempotent, it must be the case that $\mathbf{Q} = \mathbf{Q}^\intercal\mathbf{Q} = \mathbf{Q}\mathbf{Q}$, where $\mathbf{Q} \triangleq (\mathbf{I} - \mathbf{P})$, and $\mathbf{P} \in \mathbb{R}^{m\times m}$. Thus we can show:

$$\begin{aligned}
\mathbf{Q}\mathbf{Q} &= (\mathbf{I}-\mathbf{P})(\mathbf{I}-\mathbf{P}) && (270)\\
&= \mathbf{I}-\mathbf{P} && (271)\\
&= \mathbf{Q} && (272)\\
\mathbf{Q}^\intercal\mathbf{Q} &= (\mathbf{I}-\mathbf{P})^\intercal(\mathbf{I}-\mathbf{P}) && (273)\\
&= (\mathbf{I}-\mathbf{P}^\intercal)(\mathbf{I}-\mathbf{P}) && (274)\\
&= \mathbf{I}-\mathbf{P}-\mathbf{P}^\intercal+\mathbf{P}^\intercal\mathbf{P}, \quad \mathbf{P}=\mathbf{P}^\intercal\mathbf{P} && (275)\\
&= \mathbf{I}-\mathbf{P}-\mathbf{P}^\intercal+\mathbf{P}, \quad \mathbf{P}^\intercal=\mathbf{P} && (276)\\
&= \mathbf{I}-\mathbf{P} && (277)\\
&= \mathbf{Q} && (278)\\
\therefore \mathbf{Q}^\intercal\mathbf{Q} &= \mathbf{Q}\mathbf{Q} = \mathbf{Q} && (279)
\end{aligned}$$

■

D. Proof of rank Q

We can prove that rank$(\mathbf{Q}) = m - n$ by the SVD of $\mathbf{H}$. Apply the result from the proof for the rank of $\mathbf{P}$, where $\mathbf{P} \in \mathbb{R}^{m\times m}$ and $\mathbf{I} \in \mathbb{R}^{m\times m}$. Using the inner product we can define $\mathbf{I}$ in terms of $\mathbf{U}$

$$\begin{aligned}
\mathbf{I} &= \mathbf{U}\mathbf{U}^\intercal && (280)\\
&= [\mathbf{U}_1\mathbf{U}_2]\begin{bmatrix}\mathbf{U}_1^\intercal\\ \mathbf{U}_2^\intercal\end{bmatrix} && (281)\\
&= \mathbf{U}_1\mathbf{U}_1^\intercal + \mathbf{U}_2\mathbf{U}_2^\intercal. && (282)
\end{aligned}$$

Alternatively, by the outer product we can define

$$\begin{aligned}
\mathbf{I} &= \mathbf{U}^\intercal\mathbf{U} && (283)\\
&= \begin{bmatrix}\mathbf{U}_1^\intercal\\ \mathbf{U}_2^\intercal\end{bmatrix}[\mathbf{U}_1\mathbf{U}_2] && (284)\\
&= \begin{bmatrix}\mathbf{U}_1^\intercal\mathbf{U}_1 & \mathbf{U}_1^\intercal\mathbf{U}_2\\ \mathbf{U}_2^\intercal\mathbf{U}_2 & \mathbf{U}_2^\intercal\mathbf{U}_2\end{bmatrix} && (285)
\end{aligned}$$

where $\mathbf{U}_1^\intercal\mathbf{U}_1 = \mathbf{I} \in \mathbb{R}^{n\times n}$, $\mathbf{U}_2^\intercal\mathbf{U}_2 = \mathbf{I} \in \mathbb{R}^{(m-n)\times(m-n)}$. Finally, $\mathbf{U}_2\mathbf{U}_2^\intercal = \mathbf{P} \in \mathbb{R}^{m\times m}$ as proved above, and $\mathbf{U}_1\mathbf{U}_1^\intercal = \mathbf{Q} \in \mathbb{R}^{m\times m}$ which is proven below.

Now define $\mathbf{Q}$ as

$$\begin{aligned}
\mathbf{Q} &= \mathbf{I}-\mathbf{P} && (286)\\
&= (\mathbf{U}_1\mathbf{U}_1^\intercal + \mathbf{U}_2\mathbf{U}_2^\intercal) - \mathbf{U}_1\mathbf{U}_1^\intercal && (287)\\
&= \mathbf{U}_2\mathbf{U}_2^\intercal. && (288)
\end{aligned}$$

The rank of matrix $\mathbf{Q}$ is defined as the number of non-zero singular values of $\mathbf{Q}$. Thus, for $\mathbf{Q} \triangleq (\mathbf{I}-\mathbf{P})$, and rank$(\mathbf{P}) = n$, the number of non-zero singular values of $\mathbf{Q}$ is at most $m - n$, and therefore the rank$(\mathbf{Q}) = m - n$.

■

E. Physical Interpretation of P & Q

The physical interpretation for $\mathbf{P}$ and $\mathbf{Q}$ is a mapping of the measurement and the residual, as shown in Fig. 5. $\mathbf{P}\mathbf{y}$ projects $\mathbf{y}$ onto the range$(\mathbf{P})$ along the direction of $\mathbf{y}$. The complementary projector is $\mathbf{Q}$, where $\mathbf{Q}\mathbf{y}$ projects $\mathbf{y}$ onto the range$(\mathbf{Q})$ which is orthogonal to the range$(\mathbf{P})$.

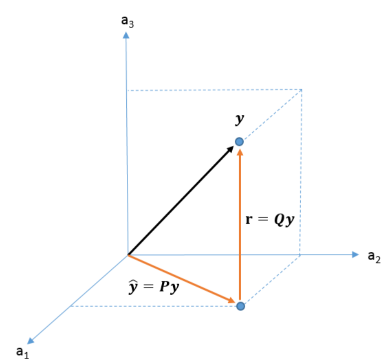

Fig. 5. For a general space in $\mathbb{R}^3$, the mapping $\mathbf{P}\mathbf{y} = \hat{\mathbf{y}}$ is the estimate for $\mathbf{y}$, and $\mathbf{Q}\mathbf{y} = \mathbf{r}$ is the estimation residual for $\mathbf{y}$.

From the SVD of $\mathbf{H}$ we have the relations:

1) $\mathbf{V}_1\mathbf{V}_1^\intercal$ is the orthogonal projector onto $[N(\mathbf{H})]^\perp = R(\mathbf{H}^\intercal)$.
2) $\mathbf{V}_2\mathbf{V}_2^\intercal$ is the orthogonal projector onto $N(\mathbf{H})$.
3) $\mathbf{U}_1\mathbf{U}_1^\intercal$ is the orthogonal projector onto $R(\mathbf{H})$.
4) $\mathbf{U}_2\mathbf{U}_2^\intercal$ is the orthogonal projector onto $[R(\mathbf{H})]^\perp = N(\mathbf{H}^\intercal)$.

XII. State Estimation

Let the state space model for a system with input $\mathbf{u}$ and output $\mathbf{y}$ be described by

$$\begin{aligned}
\dot{\boldsymbol{x}} &= \mathbf{f}(\boldsymbol{x}, \mathbf{u}) && (289)\\
\mathbf{y} &= \mathbf{h}(\boldsymbol{x}). && (290)
\end{aligned}$$

Assume that for a nominal input $\mathbf{u}_o(t)$ a nominal state trajectory $\boldsymbol{x}_o(t)$ is known which satisfies

$$\dot{\boldsymbol{x}}_o = \mathbf{f}(\boldsymbol{x}_o, \mathbf{u}_o) \tag{291}$$

$$\mathbf{y}_o = \mathbf{h}(\boldsymbol{x}_o). \tag{292}$$

Define the error state vector as $\boldsymbol{\delta x}(t) = \boldsymbol{x}(t) - \boldsymbol{x}_o(t)$. Then,

$$\boldsymbol{\delta}\dot{\boldsymbol{x}} = \dot{\boldsymbol{x}} - \dot{\boldsymbol{x}}_o = \mathbf{f}(\boldsymbol{x}, \mathbf{u}) - \mathbf{f}(\boldsymbol{x}_o, \mathbf{u}_o). \tag{293}$$

We can approximate $\mathbf{f}(\boldsymbol{x}, \mathbf{u})$ using a Taylor series expansion to yield

$$\begin{aligned}\boldsymbol{\delta}\dot{\boldsymbol{x}} = \mathbf{f}(\boldsymbol{x}_o, \mathbf{u}_o) &+ \left.\frac{\partial \mathbf{f}(\boldsymbol{x}, \mathbf{u})}{\partial \boldsymbol{x}}\right|_{\boldsymbol{x}_o(t), \mathbf{u}_o(t)} \boldsymbol{\delta x} \\ &+ \left.\frac{\partial \mathbf{f}(\boldsymbol{x}, \mathbf{u})}{\partial \mathbf{u}}\right|_{\boldsymbol{x}_o(t), \mathbf{u}_o(t)} \boldsymbol{\delta}\mathbf{u} \\ &+ h.o.t's - \mathbf{f}(\boldsymbol{x}_o, \mathbf{u}_o) \end{aligned} \tag{294}$$

$$\boldsymbol{\delta}\dot{\boldsymbol{x}} = \mathbf{F}(t)\boldsymbol{\delta x}(t) + \mathbf{G}(t)\boldsymbol{\delta}\mathbf{u}(t) + h.o.t's \tag{295}$$

where $\boldsymbol{\delta}\mathbf{u} = \mathbf{u} - \mathbf{u}_o$, $\mathbf{F}(t) = \frac{\partial \mathbf{f}(\boldsymbol{x}, \mathbf{u})}{\partial \boldsymbol{x}}\big|_{\boldsymbol{x}_o(t), \mathbf{u}_o(t)}$, and $\mathbf{G}(t) = \frac{\partial \mathbf{f}(\boldsymbol{x}, \mathbf{u})}{\partial \mathbf{u}}\big|_{\boldsymbol{x}_o(t), \mathbf{u}_o(t)}$. The resultant perturbation to the system output $\boldsymbol{\delta}\mathbf{y} = \mathbf{y} - \mathbf{y}_o$ is

$$\boldsymbol{\delta}\mathbf{y}(t) = \mathbf{h}(\boldsymbol{x}(t)) - \mathbf{h}(\boldsymbol{x}_o(t)) \tag{296}$$

$$= \left.\frac{\partial \mathbf{h}(\boldsymbol{x})}{\partial \boldsymbol{x}}\right|_{\boldsymbol{x}_o(t), \mathbf{u}_o(t)} \boldsymbol{\delta x}(t) + h.o.t's \tag{297}$$

$$\boldsymbol{\delta}\mathbf{y}(t) = \mathbf{H}(t)\boldsymbol{\delta x}(t) + h.o.t's \tag{298}$$

where $\mathbf{H}(t) = \frac{\partial \mathbf{h}(\boldsymbol{x})}{\partial \boldsymbol{x}}\big|_{\boldsymbol{x}_o(t), \mathbf{u}_o(t)}$. By dropping the higher-order terms ($h.o.t's$), eqns. (295) and (298) provide the time-varying linearization of the nonlinear system:

$$\boldsymbol{\delta}\dot{\boldsymbol{x}} = \mathbf{F}(t)\boldsymbol{\delta x}(t) + \mathbf{G}(t)\boldsymbol{\delta}\mathbf{u}(t) \tag{299}$$

$$\boldsymbol{\delta}\mathbf{y}(t) = \mathbf{H}(t)\boldsymbol{\delta x}(t) \tag{300}$$

which is accurate near the nominal trajectory (i.e., for small $\|\boldsymbol{\delta x}\|$ and $\|\boldsymbol{\delta}\mathbf{u}\|$).

References

[1] Anon., "Global Positioning System Interface Specification," U.S. Air Force Global Positioning Systems Directorate Systems Engineering and Integration, Report IS-GPS-200-H, 21 March 2014.

[2] J. A. Farrell, *Aided Navigation: GPS with High Rate Sensors*. McGraw Hill, 2008.

[3] N. Trawny and S. I. Roumeliotis, "Indirect Kalman filter for 3D attitude estimation," University of Minnesota, Dept. of Comp. Sci. & Eng., Tech. Rep. 2005-002, mar 2005.

[4] M. Shuster, "The nature of the quaternion," *Journal of the Astronautical Sciences*, vol. 56, no. 3, p. 359, 2010.