Mathematics for Robotics (ROB-GY 6013) **Section A**

Formula Sheet

$q \to p$ (converse of $p \to q$) $\quad \sim q \to \sim p$ (contrapositive of $p \to q$)

A **field** consists of a set, denoted by $\mathcal{F}$, of elements called **scalars** and two operations called addition "+" and multiplication "·"; the two operations are defined over $\mathcal{F}$ such that they satisfy the following conditions:

1. To every pair of elements α and β in $\mathcal{F}$, there correspond an element $\alpha + \beta$ in $\mathcal{F}$, called the sum of α and β, and an element $\alpha \cdot \beta$ (or simply $\alpha\beta$) in $\mathcal{F}$ called the product of α and β.
2. Addition and multiplication are respectively commutative: For any α and β in $\mathcal{F}$,

$$\alpha + \beta = \beta + \alpha \qquad\qquad \alpha \cdot \beta = \beta \cdot \alpha$$

3. Addition and multiplication are respectively associative: For any α, β, γ in $\mathcal{F}$,

$$(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma) \qquad\qquad (\alpha \cdot \beta) \cdot \gamma = \alpha \cdot (\beta \cdot \gamma)$$

4. Multiplication is distributive with respect to addition: For any α, β, γ in $\mathcal{F}$,

$$\alpha \cdot (\beta + \gamma) = (\alpha \cdot \beta) + (\alpha \cdot \gamma)$$

5. $\mathcal{F}$ contains an element, denoted by 0, and an element, denoted by 1, such that $\alpha + 0 = \alpha$ and $1 \cdot \alpha = \alpha$ for every α in $\mathcal{F}$.
6. To every α in $\mathcal{F}$, there is an element β in $\mathcal{F}$ such that $\alpha + \beta = 0$. The element β is called the *additive inverse*.
7. To every α in $\mathcal{F}$ which is not the element 0, there is an element γ in $\mathcal{F}$ such that $\alpha \cdot \gamma = 1$. The element γ is called the *multiplicative inverse*.

A **vector space** (or, **linear space**) over a field $\mathcal{F}$, denoted by $(\mathcal{X}, \mathcal{F})$, consists of a set, denoted by $\mathcal{X}$, of elements called **vectors**, a field $\mathcal{F}$, and two operations called **vector addition** and **scalar multiplication**. The two operations are defined over $\mathcal{X}$ and $\mathcal{F}$ such that they satisfy all the following conditions:

1. To every pair of vectors v^1 and v^2 in $\mathcal{X}$, there corresponds a vector $v^1 + v^2$ in $\mathcal{X}$, called the sum of v^1 and v^2.
2. Addition is commutative: For any v^1, v^2 in $\mathcal{X}$, $v^1 + v^2 = v^2 + v^1$.
3. Addition is associative: For any v^1, v^2, and v^3 in $\mathcal{X}$, $(v^1 + v^2) + v^3 = v^1 + (v^2 + v^3)$
4. $\mathcal{X}$ contains a vector, denoted by $\mathbf{0}$, such that $\mathbf{0} + v = v$ for every v in $\mathcal{X}$. The vector $\mathbf{0}$ is called the zero vector or the origin.
5. To every v in $\mathcal{X}$, there is a vector $\overline{v}$ in $\mathcal{X}$, such that $v + \overline{v} = 0$.
6. To every α in $\mathcal{F}$, and every v in $\mathcal{X}$, there corresponds a vector $\alpha \cdot v$ in $\mathcal{X}$ called the scalar product of α and v.
7. Scalar multiplication is associative: For any α, β in $\mathcal{F}$ and any v in $\mathcal{X}$, $\alpha \cdot (\beta \cdot v) = (\alpha \cdot \beta) \cdot v$.
8. Scalar multiplication is distributive with respect to vector addition: For any α in $\mathcal{F}$ and any v^1, v^2 in $\mathcal{X}$, $\alpha \cdot (v^1 + v^2) = \alpha \cdot v^1 + \alpha \cdot v^2$.
9. Scalar multiplication is distributive with respect to scalar addition: For any α, β in $\mathcal{F}$ and any v in $\mathcal{X}$, $(\alpha + \beta) \cdot v = \alpha \cdot v + \beta \cdot v$.
10. For any v in $\mathcal{X}$, $1 \cdot v = v$, where 1 is the element 1 in $\mathcal{F}$.

Let $(\mathcal{X},\mathcal{F})$ be a vector space, and let $\mathcal{Y}$ be a subset of $\mathcal{X}$. Then $\mathcal{Y}$ is a **subspace** if using the rules of vector addition and scalar multiplication defined in $(\mathcal{X},\mathcal{F})$, we have that $(\mathcal{Y},\mathcal{F})$ is a vector space.
The following are equivalent:

a) $(\mathcal{Y},\mathcal{F})$ is a subspace of $(\mathcal{X},\mathcal{F})$.
b) $\forall v^1, v^2 \in \mathcal{Y}$, $v^1 + v^2 \in \mathcal{Y}$ (closed under vector addition) and
$\forall y \in \mathcal{Y}$, $\forall \alpha \in \mathcal{F}$, $\alpha y \in \mathcal{Y}$ (closed under scalar multiplication)
c) $\forall v^1, v^2 \in \mathcal{Y}$, $\forall \alpha \in \mathcal{F}$, $\alpha \cdot v^1 + v^2 \in \mathcal{Y}$
d) $\forall v^1, v^2 \in \mathcal{Y}$, $\forall \alpha_1, \alpha_2 \in \mathcal{F}$, $\alpha_1 \cdot v^1 + \alpha_2 \cdot v^2 \in \mathcal{Y}$

A **linear combination** is a finite sum of the form:

$$\alpha_1 v^1 + \alpha_2 v^2 + \ldots + \alpha_n v^n \text{ where } n \geq 1,\ \alpha_1 \in \mathcal{F},\ v^i \in \mathcal{X},\ v^i \in \mathcal{X},\ 1 \leq i \leq n$$

$$[x]_v := \begin{bmatrix} \alpha_1 & \alpha_2 & \ldots & \alpha_n \end{bmatrix}^T \in \mathcal{F}^n$$

$$[x]_{\bar{u}} = P[x]_u$$

$$[\mathcal{L}(x)]_v = A[x]_u \qquad \begin{array}{ccc} \mathcal{X} & \xrightarrow{\mathcal{L}} & \mathcal{Y} \\ [x]_u \downarrow & & \downarrow [y]_v \\ \mathcal{F}^m & \xrightarrow{A} & \mathcal{F}^n \end{array}$$

$$Av = \lambda v \qquad A = M \Lambda M^{-1}$$

$$\det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = ad - bc$$

$$\det\left(\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}\right) = a\det\left(\begin{bmatrix} e & f \\ h & i \end{bmatrix}\right) - b\det\left(\begin{bmatrix} d & f \\ g & i \end{bmatrix}\right) + c\det\left(\begin{bmatrix} d & e \\ g & h \end{bmatrix}\right)$$

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}$$

Let $(\mathcal{X},\mathcal{F})$ be a vector space where the field $\mathcal{F}$ is either $\mathbb{R}$ or $\mathbb{C}$. A function $\|\cdot\|: \mathcal{X} \to \mathbb{R}$ is a **norm** if it satisfies:

a) $\|x\| \geq 0, \forall x \in \mathcal{X}$ and $\|x\| = 0 \Leftrightarrow x = 0$
b) Triangle inequality: $\|x + y\| \leq \|x\| + \|y\|, \forall x, y \in \mathcal{X}$
c) $\|\alpha x\| = |\alpha| \cdot \|x\|, \forall x \in \mathcal{X}, \alpha \in \mathcal{F}, \begin{cases} \text{if } \alpha \in \mathbb{R}, |\alpha| \text{ means the absolute value} \\ \text{if } \alpha \in \mathbb{C}, |\alpha| \text{ means the magnitude} \end{cases}$

$d(x, y) := \|x - y\|$ is called the **distance** from x to y.

Let $S \subset \mathcal{X}$ be a subset. $d(x,S) := \inf_{y \in S} \|x - y\|$

Let $(\mathcal{X},\mathbb{C})$ be a vector space. A function $\langle \cdot\,,\ \cdot\rangle : \mathcal{X}\times\mathcal{X}\to\mathbb{C}$ is an inner product if

a) $\langle a,b\rangle = \overline{\langle b,a\rangle}$
b) $\langle \alpha_1 x_1 + \alpha_2 x_2, y\rangle = \alpha_1\langle x_1,y\rangle + \alpha_2\langle x_2,y\rangle$
c) $\langle x,x\rangle \geq 0$ for any $x\in\mathcal{X}$, and $\langle x,x\rangle = 0 \Leftrightarrow x=0$

Let $(\mathcal{X},\mathcal{F},\langle \cdot\,,\ \cdot\rangle)$ be an inner product space, with $\mathcal{F}$ either $\mathbb{R}$ or $\mathbb{C}$. Then for all $x,y\in\mathcal{X}$

$$|\langle x,y\rangle| \leq \langle x,x\rangle^{1/2}\langle y,y\rangle^{1/2}$$

Let $(\mathcal{X},\mathcal{F},\langle \cdot\,,\ \cdot\rangle)$ be an inner product space, $\{y^1, \ldots, y^k\}$ a linearly independent set, and $\{v^1, \ldots, v^{k-1}\}$ an orthogonal set satisfying span$\{v^1, \ldots, v^{k-1}\}$ = span$\{y^1, \ldots, y^{k-1}\}$.

Define $v^k = y^k - \sum_{j=1}^{k-1}\frac{\langle y^k, v^j\rangle}{\|v^j\|^2}\cdot v^j$ where $\|v^j\|^2 = \langle v^j, v^j\rangle$.

Then $\{v^1, \ldots, v^{k-1}\}$ is orthogonal and span$\{v^1, \ldots, v^k\}$ = span$\{y^1, \ldots, y^k\}$.

Let $\mathcal{X}$ be a finite-dimensional (real) inner product space, M be a subspace of $\mathcal{X}$, and x be an arbitrary point in $\mathcal{X}$.

a) If $\exists m_0\in M$ such that $\|x-m_0\| \leq \|x-m\|\ \ \forall m\in M$, then m_0 is unique.
b) A necessary and sufficient condition for m_0 to be a minimizing vector in M is that the vector $x - m_0$ is orthogonal to M.

Remarks:

a) If $\exists m_0\in M$ such that $\|x-m_0\| = d(x,M) = \inf_{m\in M}\|x-m\|$, then m_0 is unique.
b) $\|x-m_0\| = d(x,M) \Leftrightarrow x-m_0 \perp M$

Let $(\mathcal{X},\mathbb{R})$ be a finite-dimensional real inner product space and M a subspace of $\mathcal{X}$. Then $\forall x\in\mathcal{X}$, there exists a unique $\hat{x}\in M$ such that $\|x-\hat{x}\| = d(x,M) := \inf_{m\in M}\|x-m\| = \min_{m\in M}\|x-m\|$. Moreover, $\hat{x}\in M$ is characterized by $x-\hat{x}\perp M$.

$$G^T\alpha = \beta \qquad G := \begin{bmatrix} \langle y^1,y^1\rangle & \langle y^1,y^2\rangle & \cdots & \langle y^1,y^k\rangle \\ \langle y^2,y^1\rangle & \langle y^2,y^2\rangle & \cdots & \langle y^2,y^k\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle y^k,y^1\rangle & \langle y^k,y^2\rangle & \ldots & \langle y^k,y^k\rangle \end{bmatrix} \quad \beta := \begin{bmatrix} \langle x,y^1\rangle \\ \langle x,y^2\rangle \\ \vdots \\ \langle x,y^k\rangle \end{bmatrix}$$

$M = \begin{bmatrix} A & B \\ B^T & C \end{bmatrix}$ The following are equivalent:

a) $M > 0$ b) $A > 0$ and $C - B^TA^{-1}B > 0$ c) $C > 0$ and $A - BC^{-1}B^T > 0$

$$\hat{\alpha} = \underset{\alpha\in\mathbb{R}^m}{\arg\min}\|A\alpha - b\|^2 \Leftrightarrow \hat{\alpha} = (A^TSA)^{-1}A^TSb$$

$$\hat{x} = \underset{Ax=b}{\arg\min}\|x\|^2 \Leftrightarrow \hat{x} = S^{-1}A^T(AS^{-1}A^T)^{-1}b$$

(Ω, $\mathcal{F}$, P) is called a **probability space**.

- Ω is the sample space. Think of it as the set of all possible outcomes of an experiment.
- $E \subset \Omega$ is an event.
- $\mathcal{F}$ is the collection of allowed events. It must at least contain $\emptyset$ and Ω. It is closed with respect to set complement, countable unions, and countable intersections.

$P : \mathcal{F} \to [0, 1]$ is a probability measure. It has to satisfy a few basic operations

1. $P(\emptyset) = 0$ and $P(\Omega) = 1$.
2. For each $E \in \mathcal{F}$, $0 \le P(E) \le 1$
3. If the sets $E_1, E_2, \ldots$ are disjoint (i.e., $E_i \cap E_j = \emptyset$ for $i \neq j$), then

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$

These are typically called the **Axioms of Probability**.

A function $X : \Omega \to \mathbb{R}$ is a **random variable** if $\forall\, x \in \mathbb{R}$, the set $\{\omega \in \Omega \,|X(\omega) \le x\} \in \mathcal{F}$, that is $P(\{\omega \in \Omega \,|X(\omega) \le x\})$ is defined.

A (piecewise continuous) function $f : \mathbb{R} \to [0,\infty)$ is a **probability density** if $\int_{-\infty}^{\infty} f(x)dx = 1$

A function $X : \Omega \to \mathbb{R}$ is a **continuous random variable** with density $f : \mathbb{R} \to [0,\infty)$ if

a) it is a random variable, and
b) $\forall\, x \in \mathbb{R}$, $P(\{\omega \in \Omega \,|X(\omega) \le x\}) = \int_{-\infty}^{x} f(\bar{x})d\bar{x} = 1$

$$E\{g(X)\} := \int_{-\infty}^{\infty} g(x) f_X(x)dx \qquad \mu = E\{X\}$$

For a random variable: $\quad \mathrm{Var}(X) := \sigma^2 = E\{(X-\mu)^2\}$

For a random vector: $\quad \Sigma := \mathrm{cov}(X) = \mathrm{cov}(X,X) = E\{(X-\mu)(X-\mu)^T\} \quad \mathrm{Var}(X) = \mathrm{trace}\big(\mathrm{cov}(X,X)\big)$

Suppose the random vector $X : \Omega \to \mathbb{R}^p$ is partitioned into two components $X_1 : \Omega \to \mathbb{R}^n$ and $X_2 : \Omega \to \mathbb{R}^m$, with $p = n + m$, so that, $X = [X_1 \quad X_2]^T$. We denote the density of X by $f_X(x) = f_{\left[\begin{smallmatrix} X_1 \\ X_2 \end{smallmatrix}\right]}(x_1, x_2) = f_{X_1 X_2}(x_1, x_2)$ and it is called the **joint density** of X_1 and X_2.

The mean and covariance are: $\quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} E\{X_1\} \\ E\{X_2\} \end{bmatrix}$

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = E\left\{ \begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \end{bmatrix} \begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \end{bmatrix}^T \right\} = E\left\{ \begin{bmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \end{bmatrix} \begin{bmatrix} (X_1 - \mu_1)^T & (X_2 - \mu_2)^T \end{bmatrix} \right\}$$

$$= E\left\{ \begin{bmatrix} (X_1 - \mu_1)(X_1 - \mu_1)^T & (X_1 - \mu_1)(X_2 - \mu_2)^T \\ (X_2 - \mu_2)(X_1 - \mu_1)^T & (X_2 - \mu_2)(X_2 - \mu_2)^T \end{bmatrix} \right\}$$

where $\Sigma_{12} = \Sigma_{12}{}^T = \mathrm{cov}(X_1, X_2) = E\{(X_1 - \mu_1)(X_2 - \mu_2)^T\}$ is also called the correlation of X_1 and X_2.

If $X = [X_1 \quad X_2]^T : \Omega \to \mathbb{R}^{n+m}$ is a continuous random vector, then its components X_1: $\Omega \to \mathbb{R}^n$ and X_2: $\Omega \to \mathbb{R}^m$, are also continuous random vectors and have densities, $f_{X_1}(x_1)$ and $f_{X_2}(x_2)$. $f_{X_1}(x_1)$ and $f_{X_2}(x_2)$ are the **marginal densities** of X_1 and X_2.

$$\begin{aligned} f_{X_1}(x_1) &:= \int_{-\infty}^{\infty} f_{X_1X_2}(x_1,x_2)dx_2 \\ &:= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1X_2}(\underbrace{\bar{x}_1,\dots,\bar{x}_n}_{x_1},\underbrace{\bar{x}_{n+1},\cdots,\bar{x}_{n+m}}_{x_2})\underbrace{d\bar{x}_{n+1}\cdots d\bar{x}_{n+m}}_{dx_2} \end{aligned}$$

$$\begin{aligned} f_{X_2}(x_2) &:= \int_{-\infty}^{\infty} f_{X_1X_2}(x_1,x_2)dx_1 \\ &:= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1X_2}(\underbrace{\bar{x}_1,\dots,\bar{x}_n}_{x_1},\underbrace{\bar{x}_{n+1},\cdots,\bar{x}_{n+m}}_{x_2})\underbrace{d\bar{x}_1\cdots d\bar{x}_n}_{dx_1} \end{aligned}$$

Random vectors X_1 and X_2 are **independent** if their joint density factors

$$f_X(x) = f_{X_1X_2}(x_1,x_2) = f_{X_1}(x_1)f_{X_2}(x_2)$$

X_1 and X_2 are **uncorrelated** if their "cross covariance" or "correlation" is zero, that is,

$$\text{cov}(X_1,X_2) := E\{(X_1-\mu_1)(X_2-\mu_2)^T\} = 0_{n\times m}$$

The **conditional probability** of A **given** B is $P(A\,|\,B) := \dfrac{P(A\cap B)}{P(B)}$

X_1 given $X_2 = x_2$ is a random vector with density $f_{X_1|X_2}(x_1\,|\,x_2)$ with mean $\mu_{X_1|X_2=x_2}$ and covariance $\Sigma_{X_1|X_2=x_2}$

$$\mu_{X_1|X_2=x_2} := E\{X_1\,|\,X_2=x_2\} = \int_{-\infty}^{\infty} x_1 f_{X_1|X_2}(x_1\,|\,x_2)dx_1$$

$$\Sigma_{X_1|X_2=x_2} := E\{(X_1-\mu_{X_1|X_2=x_2})(X_1-\mu_{X_1|X_2=x_2})^T\,|\,X_2=x_2\} = \int_{-\infty}^{\infty}(X_1-\mu_{X_1|X_2=x_2})(X_1-\mu_{X_1|X_2=x_2})^T f_{X_1|X_2}(x_1\,|\,x_2)dx_1$$

A random variable X is **normally distributed** with mean μ and variance $\sigma^2 > 0$ if it has density

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

A finite collection of random variables $X_1, X_2, \ldots, X_p$, or equivalently, the random vector $X = [X_1 \quad X_2 \quad \ldots \quad X_p]^T$ has a (non-degenerate) **multivariate normal distribution** with mean μ and covariance $\Sigma > 0$ if the joint density is given by

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^p\,|\Sigma|}}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

Let X be a Gaussian random vector. Define a new random vector by $Y = AX + b$, with the rows of A linearly independent. Then Y is a Gaussian (normal) random vector with

$\mu_Y := E\{Y\} = A\mu + b$

$\Sigma_{YY} := \text{cov}(Y,Y) = E\{(Y-\mu_Y)(Y-\mu_Y)^T\} = A\Sigma A^T$