

Kalman filter & Singular Value Decomposition

ROB 501

Necmiye Ozay

- Derivation of Kalman filter (and why it is a minimum variance estimator)
- (if time) Singular value decomposition

$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \sim \mathcal{N}_{\mu, \Sigma}$

Gaussian random variables (important properties)

Let $X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ be a Gaussian random vector with

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}.$$

$$\boxed{\mu_{1|2} = \mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (x - \mu_2)}$$

- Fact 1: Random vector $(X_1 | X_2 = x_2)$ is a Gaussian with

$$\underline{\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}}$$

Shur complement of Σ_{11} in Σ .

Fact 2: Let $X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}$ be a Gaussian with $\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}, \Sigma = \begin{bmatrix} \Sigma_{11} & 0 & \Sigma_{13} \\ 0 & \Sigma_{22} & 0 \\ \Sigma_{13}^T & 0 & \Sigma_{33} \end{bmatrix}$

$\rightarrow X_2$ is indep. of X_1 and X_3 .

Then X_1 and X_2 are conditionally independent given X_3 : $\text{cov}(X_1 | X_3, X_2 | X_3) = 0$.

Fact 3: Let X_1 and X_2 indep. Gaussians. $Y = A X_1 + B X_2$ is a Gaussian with

$$\mu_Y = A\mu_1 + B\mu_2 \quad \Sigma_Y = \underline{A \Sigma_{11} A^T + B \Sigma_{22} B^T}$$

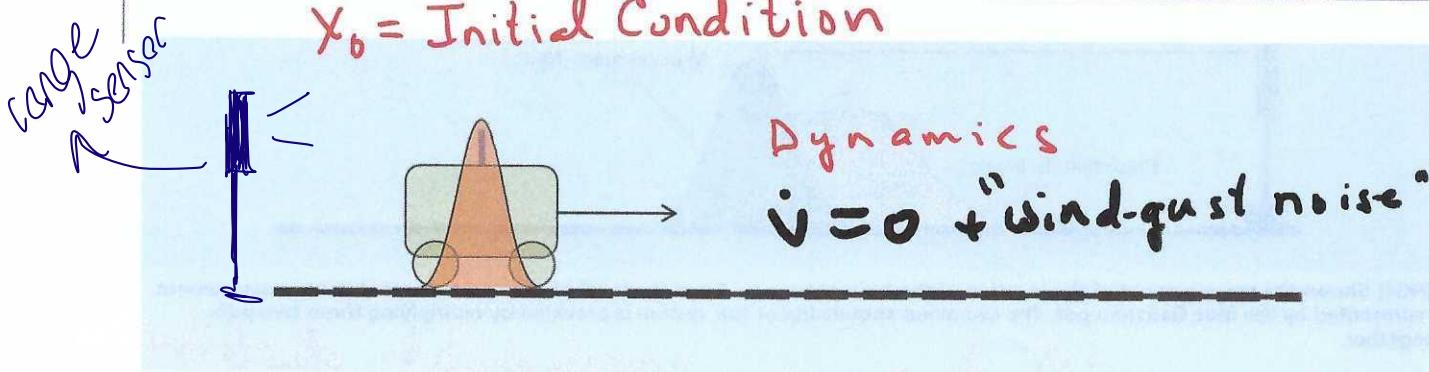
Fact 4: (recursive computation of conditionals) $\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} : (X | \begin{bmatrix} Y=y \\ Z=z \end{bmatrix}) = (X | z=z) \mid (Y | z=z)$

shortened notation: $X | \begin{bmatrix} Y \\ Z \end{bmatrix} = (X | Z) \mid (Y | Z)$

Kalman Filter Motivation

lecture NOTES continued

$x_0 = \text{Initial Condition}$

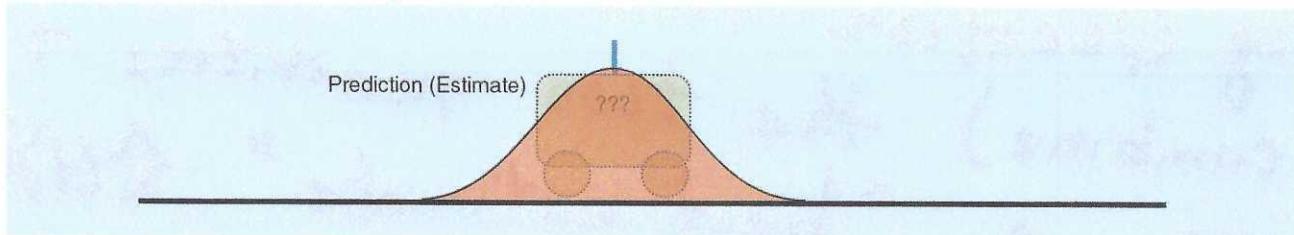


[FIG2] The initial uncertainty of the system at time $t = 0$. The initial distribution corresponds to the initial state.

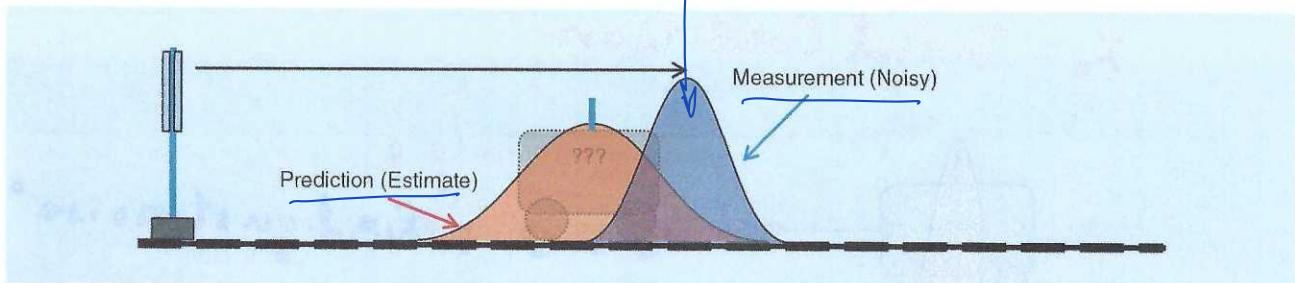
$$\underbrace{x(t+1)}_{\mathcal{N}} = \underbrace{Ax(t) + Gw(t)}_{\text{wind-gust noise}}$$

↳ Can compute this distribution using fact 3.

Uncertainty typically grows
as time increases

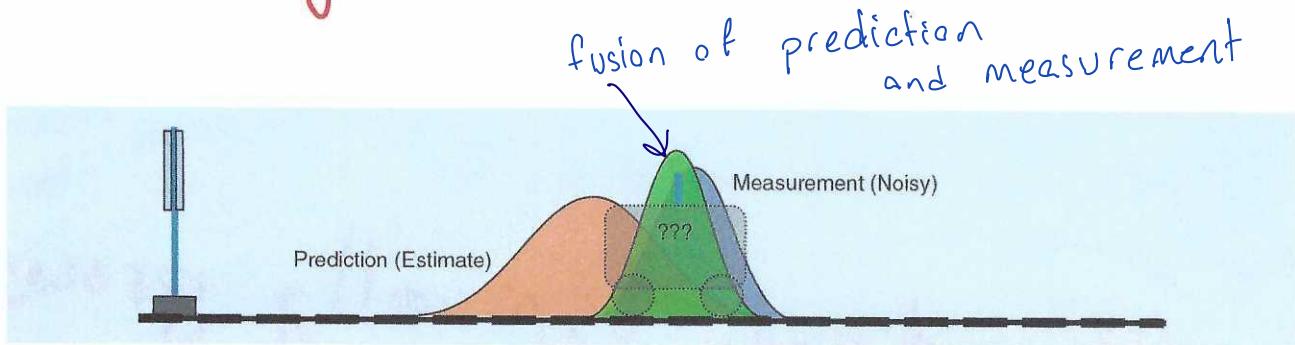


[FIG3] Here, the prediction of the location of the train at time $t = 1$ and the level of uncertainty in that prediction is shown. The confidence in the knowledge of the position of the train has decreased, as we are not certain if the train has undergone any accelerations or decelerations in the intervening period from $t = 0$ to $t = 1$.



[FIG4] Shows the measurement of the location of the train at time $t = 1$ and the level of uncertainty in that noisy measurement, represented by the blue Gaussian pdf. The combined knowledge of this system is provided by multiplying these two pdfs together.

$x(t)$ has a "probability distribution"
 $y(t)$ is not perfect = "probability distribution"



[FIG5] Shows the new pdf (green) generated by multiplying the pdfs associated with the prediction and measurement of the train's location at time $t = 1$. This new pdf provides the best estimate of the location of the train, by fusing the data from the prediction and the measurement.

Big Question: How to "fuse" (= combine) the two quantities to have a "best estimate" $\hat{x}(t)???$

$$\text{RLS: } A_k = I, G_k = 0$$

v_k = deterministic =
= no stochastic or random model.

Rob 501 Handout: Grizzle

Discrete-Time Kalman Filter and its Derivation

Model Linear time-varying discrete-time system with “white¹” Gaussian noise

want to track the state

$$x_{k+1} = A_k x_k + G_k w_k, \quad x_0 \text{ initial condition}$$

measurements

$$y_k = C_k x_k + v_k, \quad \text{measurement noise}$$

$x \in \mathbb{R}^n$, $w \in \mathbb{R}^p$, $y \in \mathbb{R}^m$, $v \in \mathbb{R}^m$. Moreover, the random vectors x_0 , and, for $k \geq 0$, w_k , v_k are all independent² Gaussian (normal) random vectors.

Notation: $\delta_{kl} = 1 \Leftrightarrow k = l$ (and $\delta_{kl} = 0$, $k \neq l$)

Precise Assumptions on the Random Variables

- For all $k \geq 0$, $l \geq 0$, x_0 , w_k , v_l are jointly Gaussian.
- w_k is a 0-mean white noise process: $\mathcal{E}\{w_k\} = 0$, and $\text{cov}(w_k, w_l) = R_k \delta_{kl}$
- v_k is a 0-mean white noise process: $\mathcal{E}\{v_k\} = 0$, and $\text{cov}(v_k, v_l) = Q_k \delta_{kl}$
- Uncorrelated noise processes: $\text{cov}(w_k, v_l) = 0 \quad \forall k, l$
- The initial condition x_0 is uncorrelated with all other noise sequences.
- We denote the mean and covariance of x_0 by

$$\bar{x}_0 = \mathcal{E}\{x_0\} \quad \text{and} \quad P_0 = \text{cov}(x_0) = \text{cov}(x_0, x_0) = \mathcal{E}\{(x_0 - \bar{x}_0)(x_0 - \bar{x}_0)^\top\}$$

¹Recall that in white light, all frequencies are present. When only certain frequency components are present, you get “colored” light, such as blue light or red light. The term “white” noise means that if you compute the power spectral density of the noise random process, it is a constant, meaning that all frequency components are equally represented, just as in white light.

²Recall that for normal random variables, uncorrelated and independent are the same thing. This is one of several special properties of Gaussian random variables.

$$x_1 = A_0 x_0 + G_0 \omega_0$$

$$x_2 = A_1 \underline{x}_1 + G_1 \omega_1$$

$$x_2 = A_1 A_0 x_0 + A_1 G_0 \omega_0 + G_1 \omega_1$$

Short-hand notation for the noise modeling assumptions:

$$\text{cov} \left(\begin{bmatrix} w_k \\ v_k \\ x_0 \end{bmatrix}, \begin{bmatrix} w_l \\ v_l \\ x_0 \end{bmatrix} \right) = \begin{bmatrix} R_k \delta_{kl} & 0 & 0 \\ 0 & Q_k \delta_{kl} & 0 \\ 0 & 0 & P_0 \end{bmatrix}, \quad \delta_{kl} = \begin{cases} 1, & k = l \\ 0, & k \neq l \end{cases}$$

Lemma (Properties of x_k and y_k Coming from the Model)

- For all $k \geq 1$, \underline{x}_k is a linear combination of x_0 and w_0, \dots, w_{k-1} . In particular, \underline{x}_k is uncorrelated with w_k .
 \underline{x}_k is a lin. com.
- For all $k \geq 1$, \underline{y}_k is a linear combination of x_0, w_0, \dots, w_{k-1} and v_k . In particular, \underline{y}_k is uncorrelated with w_k .
- For all $k \geq 0$, v_k is uncorrelated with x_k .

The proof is by induction using the recursive nature of the discrete-time model. We skip it. The reader can easily fill it in.

Remark: On the next page, we give (one form of) the discrete-time Kalman Filter. After that, we do the main elements of its derivation. There are many variations of the basic filter, all equivalent to the one we give, but some preferable over others for numerical reasons. Page 8 provides a version of the filter with the measurement update and prediction steps combined.

The Kalman Filter

Definition of Terms:

$$\widehat{x}_{k|k} := \mathcal{E}\{x_k | y_0, \dots, y_k\}$$

$$\underline{P}_{k|k} := \mathcal{E}\{(x_k - \widehat{x}_{k|k})(x_k - \widehat{x}_{k|k})^\top | y_0, \dots, y_k\}$$

$$(x_k | y_0, \dots, y_k) \sim \mathcal{N}(\widehat{x}_{k|k}, P_{k|k})$$

prediction
 of \widehat{x}_{k+1} before
 measuring
 y_{k+1}

$$\widehat{x}_{k+1|k} := \mathcal{E}\{x_{k+1} | y_0, \dots, y_k\}$$

$$P_{k+1|k} := \mathcal{E}\{(x_{k+1} - \widehat{x}_{k+1|k})(x_{k+1} - \widehat{x}_{k+1|k})^\top | y_0, \dots, y_k\}$$

Initial Conditions:

$$\widehat{x}_{0|-1} := \bar{x}_0 = \mathcal{E}\{x_0\}, \text{ and } P_{0|-1} := P_0 = \text{cov}(x_0)$$

For $k \geq 0$

Measurement Update Step:

$$\rightarrow K_k = P_{k|k-1} C_k^\top (C_k P_{k|k-1} C_k^\top + Q_k)^{-1}$$

(Kalman Gain)

$$\widehat{x}_{k|k} = \widehat{x}_{k|k-1} + K_k (y_k - C_k \widehat{x}_{k|k-1})$$

$$P_{k|k} = P_{k|k-1} - K_k C_k P_{k|k-1}$$

innovation term

Time Update or Prediction Step:

$$\widehat{x}_{k+1|k} = A_k \widehat{x}_{k|k}$$

$$P_{k+1|k} = A_k P_{k|k} A_k^\top + G_k R_k G_k^\top$$

Fact 3

$$x_{k+1} = A_k x_k + G_k w_k$$

$$\mathcal{E}(x_{k+1}) = \mathcal{E}(A_k x_k) + \mathcal{E}(G_k w_k)$$

End of For Loop (Just stated this way to emphasize the recursive nature of the filter)

Preliminaries

Measurements: We collect all of the measurements at time k as

$$\underline{Y_k} = (\underline{y_k}, \underline{y_{k-1}}, \dots, \underline{y_0}).$$

Strictly speaking, we should be stacking them up into a column vector as we have done for all of our estimation problems, but notationally, it is more convenient to write them in a row. Also, it is more convenient to put the most recent measurement at the head of the list. We note that

$$Y_k = (\underline{y_k}, \underline{\underline{Y_{k-1}}}).$$

Hence,

$$\rightarrow \hat{x}_{k|k} := \mathcal{E}\{x_k | Y_k\}$$

$$\rightarrow P_{k|k} := \mathcal{E}\{(x_k - \hat{x}_{k|k})(x_k - \hat{x}_{k|k})^\top | Y_k\}$$

mean and covariance of the conditional normal random vector $x_k | Y_k$

$$\rightarrow \hat{x}_{k+1|k} := \mathcal{E}\{x_{k+1} | Y_k\}$$

$$\rightarrow P_{k+1|k} := \mathcal{E}\{(x_{k+1} - \hat{x}_{k+1|k})(x_{k+1} - \hat{x}_{k+1|k})^\top | Y_k\}$$

mean and covariance of the conditional normal random vector $x_{k+1} | Y_k$

Important Remarks:

- The conditional random vector $\boxed{x_k | Y_k}$ is distributed $N(\hat{x}_{k|k}, P_{k|k})$.
- The conditional random vector $\boxed{x_{k+1} | Y_k}$ is distributed $N(\hat{x}_{k+1|k}, P_{k+1|k})$.

Filter Derivation Using Induction and Properties of Conditional Distributions of Gaussian Random Vectors

Base step: The initial conditions of the filter at time $k = 0$, namely

$$\widehat{x}_{0|-1} := \bar{x}_0, \quad \text{and} \quad P_{0|-1} := P_0$$

Induction step: At time $k \geq 0$, we suppose that $(\widehat{x}_{k|k-1}, P_{k|k-1})$ are known, and we derive $(\widehat{x}_{k|k}, P_{k|k})$ and $(\widehat{x}_{k+1|k}, P_{k+1|k})$.

Key idea of the development: We need to compute the distribution (or density) of the conditional random vector

$$x_k | Y_k = \boxed{x_k | (y_k, Y_{k-1})} \quad Y_k = \begin{bmatrix} y_k \\ Y_{k-1} \end{bmatrix}$$

From HW 8, we learned **Fact 4** $X|(Y, Z) = X|Z \mid Y|Z$. From this we obtain

$$\boxed{x_k | Y_k} = x_k | (y_k, Y_{k-1}) = x_k | Y_{k-1} \mid y_k | Y_{k-1} \quad (*)$$

where we have identified

$$x_k \leftrightarrow X, \quad y_k \leftrightarrow Y, \quad \text{and} \quad Y_{k-1} \leftrightarrow Z.$$

Hence, if we can compute the distribution (or density) of

$$\boxed{\begin{bmatrix} x_k \\ y_k \end{bmatrix} \mid Y_{k-1}} = \begin{bmatrix} x_k \mid Y_{k-1} \\ y_k \mid Y_{k-1} \end{bmatrix}$$

then we can apply **Fact 1** to obtain (*).

The following calculations are aimed at doing just this.

$$y_k = C_k x_k + v_k$$

Measurement Update: We seek to derive the filter equations on page 3. To begin, we have that $y_k = C_k x_k + v_k$. It follows by linearity that the conditional random variable $y_k | Y_{k-1}$ is equal to

$$y_k | Y_{k-1} = C_k x_k | Y_{k-1} + v_k | Y_{k-1}.$$

From our assumptions on the noise, v_k is independent of both x_k and Y_{k-1} , and hence by **Fact 2**, $v_k | Y_{k-1}$ is independent of the conditional random variable $x_k | Y_{k-1}$. Moreover, because v_k is independent of Y_{k-1} , $v_k | Y_{k-1} = v_k$. Putting this together, we have that

$$y_k | Y_{k-1} = \underbrace{C_k x_k | Y_{k-1}}_{\text{independent}} + v_k,$$

and $x_k | Y_{k-1}$ and v_k are independent. Hence

$$\begin{aligned}\hat{y}_{k|k-1} &:= \mathcal{E}\{y_k | Y_{k-1}\} \\ &= \mathcal{E}\{C_k x_k | Y_{k-1}\} + \mathcal{E}\{v_k\} \\ &= C_k \mathcal{E}\{x_k | Y_{k-1}\} + \mathcal{E}\{v_k\} \\ &= C_k \hat{x}_{k|k-1} + 0 \\ &= C_k \hat{x}_{k|k-1}.\end{aligned}$$

$$y = C_k X + I \cdot V$$

Moreover, the independence of $x_k | Y_{k-1}$ and v_k with **Fact 3** yields

$$\text{cov}(y_k | Y_{k-1}, y_k | Y_{k-1}) = C_k P_{k|k-1} C_k^\top + Q_k.$$

f

$$\begin{bmatrix} x_k | Y_{k-1} \\ y_k | Y_{k-1} \end{bmatrix}$$

We use independence again to obtain

$$\text{cov}(x_k | Y_{k-1}, y_k | Y_{k-1}) = \text{cov}(x_k | Y_{k-1}, C_k x_k | Y_{k-1}) = P_{k|k-1} C_k^\top.$$

With this information, we conclude that the vector

$$\begin{bmatrix} x_k \\ y_k \end{bmatrix} | Y_{k-1}$$

is jointly normally distributed, with mean and covariance

$$\begin{bmatrix} \hat{x}_{k|k-1} \\ C_k \hat{x}_{k|k-1} \end{bmatrix}, \begin{bmatrix} P_{k|k-1} & P_{k|k-1} C_k^\top \\ C_k P_{k|k-1} & C_k P_{k|k-1} C_k^\top + Q_k \end{bmatrix} \quad (*)$$

As discussed on the previous page, to compute the distribution of $(x_k | Y_k)$, we have from **Fact 4**

$$x_k | Y_k = x_k \Big| (y_k, Y_{k-1}) = x_k | Y_{k-1} \Big| y_k | Y_{k-1},$$

and thus applying **Fact 1 to (*)** we compute the mean and covariance of $x_k | Y_k = x_k | Y_{k-1} \Big| y_k | Y_{k-1}$ to be

$$\begin{aligned} \hat{x}_{k|k} &= \hat{x}_{k|k-1} + \underbrace{P_{k|k-1} C_k^\top}_{\text{Kalman gain}} \underbrace{[C_k P_{k|k-1} C_k^\top + Q_k]^{-1}}_{\text{Kalman gain}} (y_k - C_k \hat{x}_{k|k-1}) \\ P_{k|k} &= P_{k|k-1} - \underbrace{P_{k|k-1} C_k^\top}_{\text{Kalman gain}} \underbrace{[C_k P_{k|k-1} C_k^\top + Q_k]^{-1}}_{\text{Kalman gain}} C_k P_{k|k-1} \end{aligned}$$

Remark: We note that $P_{k|k}$ is the Schur complement $C_k P_{k|k-1} C_k^\top + Q_k$ in the covariance of

$$\begin{bmatrix} x_k \\ y_k \end{bmatrix} | Y_{k-1}$$

$$\begin{aligned} K_k &= P_{k|k-1} C_k^\top (C_k P_{k|k-1} C_k^\top + Q_k)^{-1} \\ (\text{Kalman Gain}) \\ \hat{x}_{k|k} &= \hat{x}_{k|k-1} + K_k (y_k - C_k \hat{x}_{k|k-1}) \\ P_{k|k} &= P_{k|k-1} - K_k C_k P_{k|k-1} \end{aligned}$$

Prediction or Time Update: We seek to derive the filter equations on page 3. This time we use the state-variable model instead of the output model, namely

$$x_{k+1} = A_k x_k + G_k w_k,$$

and we are interested in the random vector

$$x_{k+1}|Y_k = A_k x_k|Y_k + G_k w_k|Y_k.$$

Because x_k and Y_k are both independent of w_k , by **Fact 2**, $x_k|Y_k$ and $w_k|Y_k$ are also independent. It follows that

$$\begin{aligned}\widehat{x}_{k+1|k} &= \mathcal{E}\{x_{k+1}|Y_k\} \\ &= \mathcal{E}\{A_k x_k + G_k w_k|Y_k\} \\ &= A_k \mathcal{E}\{x_k|Y_k\} + G_k \mathcal{E}\{w_k|Y_k\} \\ &= A_k \widehat{x}_{k|k} + G_k \mathcal{E}\{w_k\} \\ &= A_k \widehat{x}_{k|k},\end{aligned}$$

where we have used $w_k|Y_k = w_k$, and $\mathcal{E}\{w_k\} = 0$.

Next, we use **Fact 3** and the conditional independence of the random vectors $x_k|Y_k$ and $w_k|Y_k$ to evaluate the covariance of $x_{k+1}|Y_k$ as

$$P_{k+1|k} = A_k P_{k|k} A_k^\top + G_k R_k G_k^\top.$$

That's the Proof Folks!

Combined Update Version of the Filter (often easier to implement)

Here, we will assume the model also has a *deterministic* input u_k , and thus

$$x_{k+1} = A_k x_k + \underbrace{B_k u_k}_{\text{Actuation model}} + G_k w_k \\ y_k = C_k x_k + v_k,$$

with the assumptions on the random vectors x_0 , w_k and v_k the same as before.

Combined Filter: The measurement-update step and time-update step of the Kalman Filter can be combined into a single step. The algorithm becomes:

Initial Conditions:

$$\hat{x}_{0|-1} := \bar{x}_0 = \mathcal{E}\{x_0\}, \quad \text{and} \quad P_{0|-1} := P_0 = \text{cov}(x_0)$$

For $k \geq 0$

$$K_k = (P_{k|k-1} C_k^\top) [C_k P_{k|k-1} C_k^\top + Q_k]^{-1}$$

$$\hookrightarrow \hat{x}_{k+1|k} = \underbrace{A_k \hat{x}_{k|k-1} + B_k u_k}_{\text{state update}} + A_k K_k (y_k - C_k \hat{x}_{k|k-1})$$

$$\hookrightarrow P_{k+1|k} = A_k [P_{k|k-1} - K_k C_k P_{k|k-1}] A_k^\top + G_k R_k G_k^\top$$

End of For Loop

Remark: You do not have to start at $k = 0$. In MATLAB, it is often easier to begin with $k = 1$. In that case, the initial conditions are

$$\hat{x}_{1|0} := \bar{x}_0 = \mathcal{E}\{x_0\}, \quad \text{and} \quad P_{1|0} := P_0 = \text{cov}(x_0)$$

Remark: $K_k C_k P_{k|k-1} = (P_{k|k-1} C_k^\top) [C_k P_{k|k-1} C_k^\top + Q_k]^{-1} C_k P_{k|k-1}$ is symmetric positive semi-definite and represents the value of the measurement in reducing the covariance of the state estimate, just as in the MVE.

Rob 501 Handout: Grizzle

The SVD and Numerical Rank of a Matrix (Based on a handout of Prof. Freudenberg)

Motivation: In abstract linear algebra, a set of vectors is either linearly independent or not. There is nothing in between. For example, the set of vectors

$$\left\{ v^1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, v^2 = \begin{bmatrix} 0.999 \\ 1 \end{bmatrix} \right\}$$

is linearly independent. In this case, you look at it and say, yes, BUT, the vectors are “almost” dependent because when I take the determinant

$$\det \begin{bmatrix} 1 & 0.999 \\ 1 & 1 \end{bmatrix} = 0.001,$$

I get something pretty small, so I am OK with calling them dependent. Well, what about the set

$$\left\{ v^1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, v^2 = \begin{bmatrix} 10^4 \\ 1 \end{bmatrix} \right\}?$$

When you form the matrix and check the determinant, you get

$$\det \begin{bmatrix} 1 & 10^4 \\ 0 & 1 \end{bmatrix} = 1,$$

which seems pretty far from zero. So are these vectors “adequately” linearly independent?

Maybe not! Let’s note that

$$\begin{bmatrix} 1 & 10^4 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 10^{-4} & 0 \end{bmatrix} = \begin{bmatrix} 1 & 10^4 \\ 10^{-4} & 1 \end{bmatrix},$$

is clearly singular! Hence, we can add a very small perturbation to our vectors and make them dependent! This cannot be good! :(

Question: How to quantify the statement, “the rank is *nearly* 1” or more generally, how to quantify that a set of vectors is *nearly* linearly dependent?

Answer: The Singular Value Decomposition (**SVD**).

A good reference on numerical linear algebra is G. H. Golub and C. F. van Loan, *Matrix Computations*, The Johns Hopkins University Press, 1983.

Remark: In practice, you may have a need to deal with matrices that have complex entries, so the end of the handout also does things for $\mathbb{C}^{m \times n}$. The generalization of a real *symmetric* matrix is called a *Hermitian* matrix. And the generalization of a real *orthogonal* matrix is called a *unitary matrix*. These will not be on any ROB 501 exam.

Def. An $m \times n$ matrix Σ is rectangular diagonal if $\Sigma_{ij} = 0$ for $i \neq j$. The diagonal of Σ is

$$\text{diag}(\Sigma) = (\Sigma_{11}, \Sigma_{22}, \dots, \Sigma_{pp})$$

where $p := \min(m, n)$.

Examples Consider rectangular matrices

$$\Sigma_1 = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & -1 & 0 \end{bmatrix} \quad \text{and} \quad \Sigma_2 = \begin{bmatrix} 1 & 0 \\ 0 & -6 \\ 0 & 0 \end{bmatrix}$$

Then,

$$\text{diag}(\Sigma_1) = [3 \ 4 \ -1] \quad \text{and} \quad \text{diag}(\Sigma_2) = [1 \ -6]$$

SVD Theorem: Any $m \times n$ real matrix A can be factored as

$$A = U\Sigma V^\top$$

where

$U = m \times m$ orthogonal matrix

$V = n \times n$ orthogonal matrix

$\Sigma = m \times n$ rectangular diagonal matrix

and $\text{diag}(\Sigma) = [\sigma_1, \sigma_2, \dots, \sigma_p]$ satisfies $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p \geq 0$ where $p = \min(m, n)$. Moreover, the columns of U are eigenvectors of AA^\top , the columns of V are eigenvectors of $A^\top A$, and the $(\sigma_i)^2$ are eigenvalues of both AA^\top and $A^\top A$.

Remark: The entries of $\text{diag}(\Sigma)$ are called singular values of A . We refer to σ_i as the i 'th singular value, to u_i as the i 'th left singular vector, and to v_i as the i 'th right singular vector. The proof of the theorem is on page 12.

SVD Singular value decomposition.

[U,S,V] = SVD(X) produces a diagonal matrix S, of the same dimension as X and with non-negative diagonal elements in decreasing order, and unitary matrices U and V so that $X = U*S*V'$.

By itself, SVD(X) returns a vector containing diag(S).

```
A =[  
    0.8038    0.1788    0.0960  
    0.8576    0.6365    0.6991  
    0.1107    0.6680    0.8653  
    0.9522    0.6690    0.7041  
    0.6551    0.7961    0.9283];
```

```
>> [U,S,V]=svd(A)
```

```
U =  
   -0.2439   -0.6486    0.1591   -0.6821   -0.1711  
   -0.4940   -0.1655    0.7006    0.4508    0.1859  
   -0.3688    0.6649    0.2370   -0.3156   -0.5159  
   -0.5236   -0.2418   -0.5645    0.3955   -0.4385  
   -0.5350    0.2268   -0.3303   -0.2747    0.6912
```

```
S =  
   2.5686         0         0  
      0    0.8370         0  
      0         0    0.0100  
      0         0         0  
      0         0         0
```

```
V =
-0.5877 -0.8020 0.1065
-0.5375 0.2886 -0.7923
-0.6047 0.5229 0.6007
```

```
>> U*U'
```

```
ans =
1.0000 0.0000 0.0000 -0.0000 0.0000
0.0000 1.0000 -0.0000 0.0000 -0.0000
0.0000 -0.0000 1.0000 -0.0000 -0.0000
-0.0000 0.0000 -0.0000 1.0000 -0.0000
0.0000 -0.0000 -0.0000 -0.0000 1.0000
```

```
>> A-U*S*V'
```

```
ans =
1.0e-15 *
0 0.0278 -0.8327
-0.3331 0 0.1110
-0.2220 0 0.2220
-0.2220 0.2220 -0.2220
-0.1110 0.2220 0
```

→ **Theorem:** $\text{rank}(A) = \text{number of nonzero singular values.}$

Fact: The numerical rank of A is the number of singular values that are larger than a given threshold. Often the threshold is chosen as a percentage of the largest singular value.

Example: 5×5 matrix

$$A = \begin{bmatrix} -32.57514 & -3.89996 & -6.30185 & -5.67305 & -26.21851 \\ -36.21632 & -11.13521 & -38.80726 & -16.86330 & -1.42786 \\ -5.07732 & -21.86599 & -38.27045 & -36.61390 & -33.95078 \\ -36.51955 & -38.28404 & -19.40680 & -31.67486 & -37.34390 \\ -25.28365 & -38.57919 & -31.99765 & -38.36343 & -27.13790 \end{bmatrix}$$

`[U,Sigma,V]=svd(A);`

$$\Sigma = \begin{bmatrix} 132.459 & 0.00000 & 0.00000 & 0.00000 & 0.00000 \\ 0.00000 & 37.70811 & 0.00000 & 0.00000 & 0.00000 \\ 0.00000 & 0.00000 & 33.41836 & 0.00000 & 0.00000 \\ 0.00000 & 0.00000 & 0.00000 & 19.34060 & 0.00000 \\ 0.00000 & 0.00000 & 0.00000 & 0.00000 & 0.79164 \end{bmatrix}$$

Because the **smallest singular value** $\sigma_5 = 0.79164$ is less than 1% of the **largest singular value** $\sigma_1 = 132.459$, in many cases, one might say that the numerical rank of A was 4 instead of 5.

This notion of numerical rank can be formalized by asking the following question: Suppose rank(A) = r . How far away is A from a matrix of rank strictly less than r ?

The numerical rank of a matrix is based on the expansion

$$A = U\Sigma V^\top = \sum_{i=1}^p \sigma_i u_i v_i^\top = \sigma_1 u_1 v_1^\top + \sigma_2 u_2 v_2^\top + \cdots + \sigma_p u_p v_p^\top$$

where $p = \min\{m, n\}$, and once again, the singular values are ordered such that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p$. Each term $u_i v_i^\top$ is a rank-one matrix. The following exercises will help you understand the expansion.

Exercises: Suppose that A is $m \times n$, B is $n \times m$, and that $A = U\Sigma V^\top$ is the singular value decomposition of A .

- Partition A by columns, that is, $A = [A_1 \mid A_2 \mid \cdots \mid A_n]$ and B by rows, that is, $B^\top = [B_1^\top \mid B_2^\top \mid \cdots \mid B_n^\top]$. Show that

$$AB = \sum_{k=1}^n A_k B_k.$$

Hint: Show that $[A_k B_k]_{ij} = a_{ik} b_{kj}$ and recall the formula for $[AB]_{ij}$

- For here and the following, let $m = n$. $U\Sigma = [\sigma_1 u_1 \mid \sigma_2 u_2 \mid \cdots \mid \sigma_m u_m]$
- $A = U\Sigma V^\top = \sum_{i=1}^m \sigma_i u_i v_i^\top$
- $\forall 1 \leq j \leq m$, $[u_i v_i^\top] v_j = \begin{cases} u_i & j = i \\ 0 & j \neq i \end{cases}$
- $A^\top A = V\Sigma^2 V^\top$
- $A^\top A = V\Sigma^2 V^\top = \sum_{i=1}^m (\sigma_i)^2 v_i v_i^\top$
- The e-values of $v_i v_i^\top$ are $\lambda_1 = 1$ and the rest are zero. **Hint:** Show that

$$[v_i v_i^\top] v_j = \begin{cases} v_i & j = i \\ 0 & j \neq i \end{cases}$$

Recall from HW For a symmetric real matrix M ,

$$\max_{x^\top x=1} x^\top M x = \lambda_{\max}(M)$$

Def. (Induced matrix norm) Given $A \in \mathbb{R}^{m \times n}$. Then the *matrix norm induced by the Euclidean vector norm* is given by:

$$\|A\|_2 := \max_{x^\top x=1} \|Ax\| \quad (1)$$

$$= \max_{x^\top x=1} \sqrt{x^\top A^\top A x} \quad (2)$$

$$= \sqrt{\max_{x^\top x=1} x^\top A^\top A x} \quad (3)$$

$$= \sqrt{\lambda_{\max}(A^\top A)} \quad (4)$$

where $\lambda_{\max}(A^\top A)$ denotes the largest eigenvalue of the matrix $A^\top A$. (**Recall that we proved in lecture that all the eigenvalues of a matrix having the form $A^\top A$ are real and non-negative.**) (Also, recall HW 2)

Fact: Suppose that $\text{rank}(A) = r$, so that σ_r is the smallest non-zero singular value. Then

- (i) if an $n \times m$ matrix E satisfies $\|E\| < \sigma_r$, then $\text{rank}(A + E) \geq r$.
- (ii) there exists E with $\|E\| = \sigma_r$ and $\text{rank}(A + E) < r$.
- (iii) In fact, for $E = -\sigma_r u_r v_r^\top$, $\text{rank}(A + E) = r - 1$.
- (iv) Moreover, for $E = -\sigma_r u_r v_r^\top - \sigma_{r-1} u_{r-1} v_{r-1}^\top$, $\text{rank}(A + E) = r - 2$.

Corollary: Suppose A is square and invertible. Then σ_r measures the distance from A to the nearest singular matrix.

Example: Using A above

```
>> d=diag(Sigma);
>> d(end)=0;
>> D=diag(d);
>> B=U*D*V';
>> E=A-B;
```

$$E = \begin{bmatrix} -0.04169 & 0.12122 & 0.09818 & -0.21886 & 0.05458 \\ 0.02031 & -0.05906 & -0.04784 & 0.10663 & -0.02659 \\ 0.01966 & -0.05716 & -0.04629 & 0.10320 & -0.02574 \\ 0.07041 & -0.20476 & -0.16584 & 0.36968 & -0.09220 \\ -0.08160 & 0.23728 & 0.19218 & -0.42839 & 0.10684 \end{bmatrix}$$

```
>> max(sqrt(eig(E'*E)))
```

0.7916

```
>> [U,Sigma,V]=svd(A-E);
```

$$\Sigma = \begin{bmatrix} 132.45977 & 0.00000 & 0.00000 & 0.00000 & 0.00000 \\ 0.00000 & 37.70811 & 0.00000 & 0.00000 & 0.00000 \\ 0.00000 & 0.00000 & 33.41836 & 0.00000 & 0.00000 \\ 0.00000 & 0.00000 & 0.00000 & 19.34060 & 0.00000 \\ 0.00000 & 0.00000 & 0.00000 & 0.00000 & 0.00000 \end{bmatrix}$$

I added a matrix with norm 0.7916 and made the (exact) rank drop from 4 to 5! How cool is that? It really shows that the matrix was close to a singular matrix.

Another Example:

```
>> N=100;A=[1,N;0,1]; [U,S,V]=svd(A); A,S
```

A =

```
1    100  
0      1
```

S =

```
100.0100      0  
0      0.0100
```

Hence, yeah, the SVD captures the fact that A is nearly singular.

Interesting and Useful Facts:

- (a) We have not had the time to do anything with the nullspace and range of an $m \times n$ matrix A ; they are important subspaces.

Nullspace: $\mathbf{N}(A) := \{x \in \mathbb{R}^n \mid Ax = 0\}$

Range: $\mathbf{R}(A) := \{y \in \mathbb{R}^m \mid \exists x \in \mathbb{C}^n \text{ such that } y = Ax\}$

- (b) **Fact:** Let $[U, \Sigma, V] = \text{svd}(A)$; Then the columns of U corresponding to non-zero singular values are a basis for $\mathbf{R}(A)$ and the columns of V corresponding to zero singular values are a basis for $\mathbf{N}(A)$.
- (c) The SVD can also be used to compute an "effective" range and an "effective" nullspace of a matrix.
- (d) Suppose that $\sigma_1 \geq \dots \geq \sigma_r > \epsilon \geq \sigma_{r+1} \geq \dots \geq \sigma_n \geq 0$, so that r is the "effective" or numerical rank of A . (Note the ϵ inserted between σ_r and σ_{r+1} to denote the break point.)
- (e) Let $\mathbf{R}_{\text{eff}}(A)$ and $\mathbf{N}_{\text{eff}}(A)$ denote the effective range and effective nullspace of A , respectively. Then we can calculate bases for these subspaces by choosing appropriate singular vectors:

$$\mathbf{R}_{\text{eff}}(A) := \text{span}\{u_1, \dots, u_r\} \text{ and } \mathbf{N}_{\text{eff}}(A) := \text{span}\{v_{r+1}, \dots, v_n\}.$$

The SVD for Real Matrices

Def. An $m \times n$ matrix Σ is rectangular diagonal if $\Sigma_{ij} = 0$ for $i \neq j$. The diagonal of Σ is

$$\text{diag}(\Sigma) = (\Sigma_{11}, \Sigma_{22}, \dots, \Sigma_{kk})$$

where $k = \min(m, n)$.

Examples Consider rectangular matrices

$$\Sigma_1 = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & -1 & 0 \end{bmatrix} \quad \text{and} \quad \Sigma_2 = \begin{bmatrix} 1 & 0 \\ 0 & -6 \\ 0 & 0 \end{bmatrix}$$

Then,

$$\text{diag}(\Sigma_1) = [3 \ 4 \ -1] \quad \text{and} \quad \text{diag}(\Sigma_2) = [1 \ -6]$$

SVD Theorem: Any $m \times n$ real matrix A can be factored as

$$A = U\Sigma V^\top$$

where

$U = m \times m$ orthogonal matrix

$V = n \times n$ orthogonal matrix

$\Sigma = m \times n$ rectangular diagonal matrix

and $\text{diag}(\Sigma) = [\sigma_1, \sigma_2, \dots, \sigma_p]$ satisfies $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p \geq 0$ where $p = \min(m, n)$. Moreover, the columns of U are eigenvectors of AA^\top , the columns of V are eigenvectors of $A^\top A$, and the $(\sigma_i)^2$ are eigenvalues of both AA^\top and $A^\top A$.

Remark: The entries of $\text{diag}(\Sigma)$ are called singular values of A .

Proof of the theorem: $A^\top A$ is $n \times n$, real, and symmetric. Hence, there exist orthonormal eigenvectors $\{v^1, \dots, v^n\}$ such that $A^\top A v^j = \lambda_j v^j$. Without loss of generality, we can assume that

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$$

If not, we simply re-order the v^i 's to make it so.

For $\lambda_j > 0$, say $1 \leq j \leq r$, we define

$$\sigma_j = \sqrt{\lambda_j}$$

and

$$q^j = \frac{1}{\sigma_j} A v^j \in \mathbb{R}^m$$

Claim: $(q^i)^\top q^j = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$ for $1 \leq i, j \leq r$.

Proof of Claim:

$$\begin{aligned} (q^i)^\top q^j &= \frac{1}{\sigma_i} \frac{1}{\sigma_j} (v^i)^\top A^\top A v^j \\ &= \frac{\lambda_j}{\sigma_i \sigma_j} (v^i)^\top v^j \\ &= \begin{cases} \frac{\lambda_j}{(\sigma_i)^2} & i = j \\ 0 & i \neq j \end{cases} \\ &= \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} \end{aligned}$$

End of proof of Claim.

If $r < m$, we can extend the q^i 's to an orthonormal basis for \mathbb{R}^m . Define

$$\begin{aligned} U &= [q^1 \mid q^2 \mid \dots \mid q^m] \\ V &= [v^1 \mid v^2 \mid \dots \mid v^n] \end{aligned}$$

and define $\Sigma = m \times n$ by

$$\Sigma_{ij} = \begin{cases} \sigma_i \delta_{ij} & 1 \leq i, j \leq r \\ 0 & \text{otherwise} \end{cases}$$

Then, Σ is rectangular diagonal with

$$\text{diag}(\Sigma) = [\sigma_1, \sigma_2, \dots, \sigma_r, 0, \dots, 0]$$

To complete the proof of the theorem, it is enough to show that $U^\top AV = \Sigma$. We note that the ij element of this matrix is

$$(U^\top AV)_{ij} = q_i^\top Av^j$$

If $j > r$, then $Av^j = 0$, and thus $q_i^\top Av^j = 0$, as required. If $i > r$, then q^i was selected to be orthogonal to

$$\{q^1, \dots, q^r\} = \left\{ \frac{1}{\sigma_1} Av^1, \frac{1}{\sigma_2} Av^2, \dots, \frac{1}{\sigma_r} Av^r \right\}$$

and thus $(q^i)^\top Av^j = 0$.

Hence we now consider $1 \leq i, j \leq r$ and compute that

$$\begin{aligned} (U^\top AV)_{ij} &= \frac{1}{\sigma_i} (v^i)^\top A^\top Av^j \\ &= \frac{\lambda_j}{\sigma_i} (v^i)^\top v^j \\ &= \sigma_i \delta_{ij} \end{aligned}$$

as required. **End of Proof.**

SVD for Complex Matrices (not on ROB 501 Final Exam)

Hermitian: Consider $x \in \mathbb{C}^n$. Then we define the vector " x Hermitian" by $x^H := \bar{x}^\top$. That is, x^H is the complex conjugate transpose of x . Similarly, for a matrix $A \in \mathbb{C}^{m \times n}$, we define $A^H \in \mathbb{C}^{n \times m}$ by \bar{A}^\top . We say that a square matrix $A \in \mathbb{C}^{n \times n}$ is a *Hermitian matrix* if $A = A^H$.

Important things to note:

- Similar to $A^\top A$ for real matrices, when A is complex, $A^H A$ has e-values that are real and non-negative. The proof is similar to things we have done in lecture; if you care to see it, you can find it online.
- In MATLAB, $A' = A^H$. Yikes! It is not the ordinary transpose? No, it is the complex conjugate transpose. If you want the ordinary transpose, use `transpose(A)`.

```
>> A=[j ,0 ;0 ,-j]  
  
A =  
  
    0 + 1.0000i      0  
    0                  0 - 1.0000i  
  
>> A'  
  
ans =  
  
    0 - 1.0000i      0  
    0                  0 + 1.0000i
```

Inner product on \mathbb{C}^n : Given $x, y \in \mathbb{C}^n$. Let the elements x and y be noted

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \text{and} \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

Then the Euclidean inner product is defined as

$$\langle x, y \rangle := x^H y \tag{5}$$

$$= \bar{x}_1 y_1 + \bar{x}_2 y_2 + \cdots + \bar{x}_n y_n \tag{6}$$

We note that this puts the linearity on the right side of the “bracket”, but as we have noted in HW, both definitions are common.

Euclidean vector norm: As in class, the vector norm associated with this inner product is given by

$$\|x\|_2 := \sqrt{\langle x, x \rangle} \tag{7}$$

$$= \sqrt{\sum_{i=1}^n |x_i|^2} \tag{8}$$

We often omit the subscript "2" when we are discussing the Euclidean norm (or "2-norm") exclusively.

Euclidean matrix norm: Given $A \in \mathbb{C}^{m \times n}$. Then the *matrix norm induced by*

the Euclidean vector norm is given by:

$$\|A\|_2 := \max_{x^H x = 1} \|Ax\| \quad (9)$$

$$= \max_{x^H x = 1} \sqrt{x^H A^H A x} \quad (10)$$

$$= \sqrt{\max_{x^H x = 1} x^H A^H A x} \quad (11)$$

$$= \sqrt{\lambda_{\max}(A^H A)} \quad (12)$$

where $\lambda_{\max}(A^H A)$ denotes the largest eigenvalue of the matrix $A^H A$. (As noted above, all the eigenvalues of a matrix having the form $A^H A$ are real and non-negative.) (Also, recall HW 2)

Orthogonality: Two vectors $x, y \in \mathbb{C}^n$ are *orthogonal* if $\langle x, y \rangle = 0$.

Orthonormal Set: A collection of vectors $\{x_1, x_2, \dots, x_m\} \in \mathbb{C}^n$ is said to be an *orthonormal set* if

$$\langle x_i, x_j \rangle = \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases} \quad (\text{Hence } \|x_i\| = 1, \forall i.)$$

Unitary Matrix: A matrix $U \in \mathbb{C}^{n \times n}$ is *unitary* if $U^H U = U U^H = I_n$.

Fact: If U is a unitary matrix, then the columns of U form an orthonormal basis (ONB) for \mathbb{C}^n .

Proof of Fact: Denote the columns of U as $U = [u_1 \ u_2 \ \cdots \ u_n]$. Then

$$U^H U = \begin{bmatrix} u_1^H \\ u_2^H \\ \vdots \\ u_n^H \end{bmatrix} \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} = \begin{bmatrix} u_1^H u_1 & u_1^H u_2 & \cdots & u_1^H u_n \\ u_2^H u_1 & u_2^H u_2 & \cdots & u_2^H u_n \\ \vdots & \vdots & \ddots & \vdots \\ u_n^H u_1 & u_n^H u_2 & \cdots & u_n^H u_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

For real matrices, unitary is the same thing as orthogonal.

Example:

$U =$

$$\begin{array}{cc} 0.1259 & 0.9920 \\ 0.9920 & -0.1259 \end{array}$$

$>> U * U'$

$ans =$

$$\begin{array}{cc} 1.0000 & 0.0000 \\ 0.0000 & 1.0000 \end{array}$$

Unitary matrices are effectively rotation matrices: they do not change the length of a vector, nor the angle between two vectors. Indeed,

1) From $U^H U = U U^H = I_n$, it follows that $U^{-1} = U^H$

2) Let's compute the inner product of Ux and Uy :

$$\langle Ux, Uy \rangle := (Ux)^H Uy = x^H U^H Uy = x^H y =: \langle x, y \rangle$$

3) It follows that

(a) norm of Ux equals the norm of x :

$$\|Ux\|^2 := \langle Ux, Ux \rangle = \langle x, x \rangle =: \|x\|^2$$

(b) angle between x and y is the same as the angle between Ux and Uy :

$$\cos(\angle(x, y)) := \frac{\langle x, y \rangle}{\|x\| \|y\|} = \frac{\langle Ux, Uy \rangle}{\|Ux\| \|Uy\|} =: \cos(\angle(Ux, Uy))$$

4) All of the e-values of U have magnitude 1. Indeed, suppose that λ is an e-value with e-vector v : $Uv = \lambda v$

Applying norms to both sides of the above yields: $\|Uv\| = \|\lambda v\|$

But, by item (3) above and properties of norms:

$$\|Uv\| = \|v\| \text{ and } \|\lambda v\| = |\lambda| \|v\|$$

which, with the above, implies $|\lambda| = 1$.

Theorem (SVD for Complex matrices): Consider $A \in \mathbb{C}^{m \times n}$. Then there exist unitary matrices

$$U = [u_1 \ u_2 \ \cdots \ u_m]$$

$$V = [v_1 \ v_2 \ \cdots \ v_n]$$

such that

$$A = \begin{cases} U \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} V^H, & m \geq n \\ U \begin{bmatrix} \Sigma & 0 \end{bmatrix} V^H, & m \leq n \end{cases}$$

where

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_p \end{bmatrix}, \quad p = \min(m, n)$$

and

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$$

Terminology: We refer to σ_i as the i 'th singular value, to u_i as the i 'th left singular vector, and to v_i as the i 'th right singular vector.

$$x^\top \begin{bmatrix} y_1^\top \\ \vdots \\ y_p^\top \end{bmatrix} = \begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix} \quad (x^\top + \tilde{x}^\top) y^\top = \begin{bmatrix} c_1 \\ \vdots \\ c_p \end{bmatrix}$$

$$\tilde{x} \in \mathcal{N}\left(\begin{bmatrix} y_1 & \cdots & y_p \end{bmatrix}\right) \quad y \tilde{x} = Q$$

$$\tilde{x}^\top y^\top = Q$$

