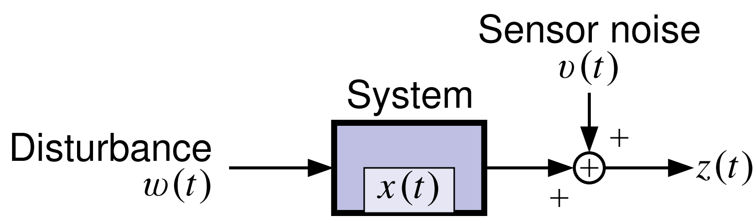
DYNAMIC SYSTEMS WITH NOISY INPUTS

3.1: Scalar random variables

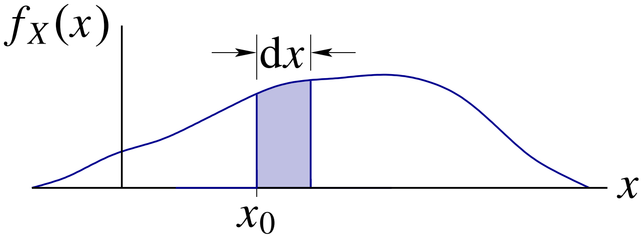
- The purpose of Kalman filters is to estimate the hidden internal state of some system where that state is affected by noise and where our measurements of system output are also corrupted by noise.



- By definition, noise is not deterministic—it is random in some sense.
- So, to discuss the impact of noise on the system dynamics, we must review the concept of a "random variable," (RV) X.
  - Cannot predict exactly what we will get each time we measure or sample the random variable, but
  - We can characterize the probability of each sample value by the "probability density function" (pdf).

Probability density functions (pdf)

- We denote probability density function (pdf) of RV X as $f_X(x)$.



- $f_X(x_0)\,\mathrm{d}x$ is the probability that random variable X is between $[x_0, x_0 + \mathrm{d}x]$.

- Properties that are true of all pdfs:

1. $f_X(x) \geq 0 \quad \forall \quad x$.
2. $\int_{-\infty}^{\infty} f_X(x)\,\mathrm{d}x = 1$.
3. $\Pr(X \leq x_0) = \int_{-\infty}^{x_0} f_X(x)\,\mathrm{d}x \triangleq F_X(x_0)$, which is the RV's cumulative distribution function (cdf).[1]

- Problem: Apart from simple examples it is often difficult to determine $f_X(x)$ accurately. ➟ Use approximations to capture the key behavior.
- Need to define key characteristics of $f_X(x)$.

EXPECTATION: Describes the expected outcome of a random trial.

$$\bar{x} = \mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x.$$

- Expectation is a linear operator (very important for working with it).
- So, for example, the first moment about the mean: $\mathbb{E}[X - \bar{x}] = 0$.

STATISTICAL AVERAGE: Different from expectation.

- Consider a (discrete) RV X that can assume n values $x_1, x_2, \ldots x_n$.
- Define the average by making many measurements $N \to \infty$. Then, m_i is the number of times the value of the measurement is i.

$$\bar{x} = \frac{1}{N}(m_1 x_1 + m_2 x_2 + \cdots + m_n x_n) = \frac{m_1}{N}x_1 + \frac{m_2}{N}x_2 + \cdots + \frac{m_n}{N}x_n.$$

- In the limit, $m_1/N \to \Pr(X = x_1)$ and so forth (assuming ergodicity),

$$\bar{x} = \sum_{i=1}^{N} x_i \Pr(X = x_i).$$

[1] Some call this a probability *distribution* function, with acronym PDF (uppercase), as different from a probability *density* function, which has acronym pdf (lowercase).

- So, statistical means can converge to expectation in the limit. (Can show similar property for continuous RVs… but harder to do.)

VARIANCE: Second moment about the mean.

$$\begin{aligned}\operatorname{var}(X) &= \mathbb{E}[(X-\bar{x})^2] = \int_{-\infty}^{\infty}(x-\bar{x})^2 f_X(x)\,\mathrm{d}x\\ &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2,\end{aligned}$$

or is equal to the mean-square minus the square-mean.

STANDARD DEVIATION: Measure of dispersion about the mean of the samples of X: $\sigma_X = \sqrt{\operatorname{var}(X)}$.

- The expectation and variance capture key features of the actual pdf. Higher-order moments are available, but *we won't need them!*

KEY POINT FOR UNDERSTANDING VARIANCE: Chebychev's inequality

- Chebychev's inequality states (for positive ε)

$$\Pr(|X-\bar{x}| \ge \varepsilon) \le \frac{\sigma_X^2}{\varepsilon^2},$$

which implies that probability is concentrated around the mean.

- It may be proven as follows:

$$\Pr(|X-\bar{x}| \ge \varepsilon) = \int_{-\infty}^{\bar{x}-\varepsilon} f_X(x)\,\mathrm{d}x + \int_{\bar{x}+\varepsilon}^{\infty} f_X(x)\,\mathrm{d}x.$$

- For the two regions of integration $|x-\bar{x}|/\varepsilon \ge 1$ or $(x-\bar{x})^2/\varepsilon^2 \ge 1$. So,

$$\Pr(|X-\bar{x}| \ge \varepsilon) \le \int_{-\infty}^{\bar{x}-\varepsilon} \frac{(x-\bar{x})^2}{\varepsilon^2} f_X(x)\,\mathrm{d}x + \int_{\bar{x}+\varepsilon}^{\infty} \frac{(x-\bar{x})^2}{\varepsilon^2} f_X(x)\,\mathrm{d}x.$$

- Since $f_X(x)$ is positive, then we also have

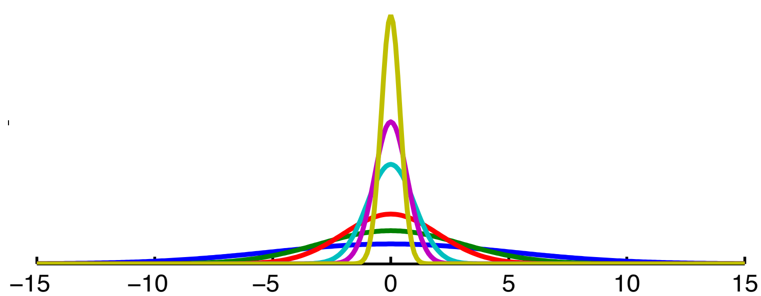
$$\Pr(|X-\bar{x}| \ge \varepsilon) \le \int_{-\infty}^{\infty} \frac{(x-\bar{x})^2}{\varepsilon^2} f_X(x)\,\mathrm{d}x = \frac{\sigma_X^2}{\varepsilon^2}.$$

- This inequality shows that probability is clustered around the mean, and that the variance is an indication of the dispersion of the pdf.
- That is, variance (later on, covariance too) informs us of how uncertain we are about the value of a random variable.
  - Low variance means that we are very certain of its value;
  - High variance means that we are very uncertain of its value.
- The mean and variance give us an estimate of the value of a random variable, and how certain we are of that estimate.

The most important distribution for this course

- The Gaussian (normal) distribution is of key importance to Kalman filters. (We will explain why this is true later—see "main point #7" on pg. 3–14.)
- Its pdf is defined as:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_X} \exp\left(-\frac{(x-\bar{x})^2}{2\sigma_X^2}\right).$$



- Symmetric about $\bar{x}$.
- Peak proportional to $\dfrac{1}{\sigma_X}$ at $\bar{x}$.
- Notation: $X \sim \mathcal{N}(\bar{x}, \sigma_X^2)$.
- Probability that X within $\pm\sigma_X$ of $\bar{x}$ is 68%; probability that X within $\pm 2\sigma_X$ of $\bar{x}$ is 96%; probability that X within $\pm 3\sigma_X$ of $\bar{x}$ is 99.7%.
  - A $\pm 3\sigma_X$ range almost certainly covers observed samples.
- "Narrow" distribution ➟ Sharp peak. High confidence in predicting X.
- "Wide" distribution ➟ Poor knowledge in what to expect for X.

3.2: Vector random variables

- With very little change in the preceding, we can also handle vectors of random variables.

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}; \qquad \text{let } x_0 = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

- X described by (scalar function) joint pdf $f_X(x)$ of vector X.
- $f_X(x_0)$ means $f_X(X_1 = x_1, X_2 = x_2 \cdots X_n = x_n)$.
- That is, $f_X(x_0)\,\mathrm{d}x_1\,\mathrm{d}x_2\,\cdots\,\mathrm{d}x_n$ is the probability that X is between x_0 and $x_0 + \mathrm{d}x$.
- Properties of joint pdf $f_X(x)$:

1. $f_X(x) \ge 0 \quad \forall \quad x$. Same as before.
2. $\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_X(x)\,\mathrm{d}x_1\,\mathrm{d}x_2\cdots\,\mathrm{d}x_n = 1$. Basically the same.
3. $\displaystyle\bar{x} = \mathbb{E}[X] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x_1\,\mathrm{d}x_2\cdots\,\mathrm{d}x_n$. Basically same.
4. Correlation matrix: Different.

$$\begin{aligned}\Sigma_X &= \mathbb{E}[XX^T] \qquad \text{(outer product)}\\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} xx^T f_X(x)\,\mathrm{d}x_1\,\mathrm{d}x_2\cdots\,\mathrm{d}x_n.\end{aligned}$$

5. *Co*variance matrix: Different. Define $\widetilde{X} = X - \bar{x}$. Then,

$$\begin{aligned}\Sigma_{\widetilde{X}} &= \mathbb{E}[(X-\bar{x})(X-\bar{x})^T]\\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}(x-\bar{x})(x-\bar{x})^T f_X(x)\,\mathrm{d}x_1\,\mathrm{d}x_2\cdots\,\mathrm{d}x_n.\end{aligned}$$

$\Sigma_{\tilde{X}}$ is symmetric and positive-semi-definite (psd). This means

$$y^T \Sigma_{\tilde{X}} y \ge 0 \quad \forall \quad y.$$

PROOF: For all $y \neq 0$,

$$\begin{aligned} 0 &\le \mathbb{E}[(y^T(X - \bar{x}))^2] \\ &= y^T \mathbb{E}[(X - \bar{x})(X - \bar{x})^T] y \\ &= y^T \Sigma_{\tilde{X}} y. \end{aligned}$$

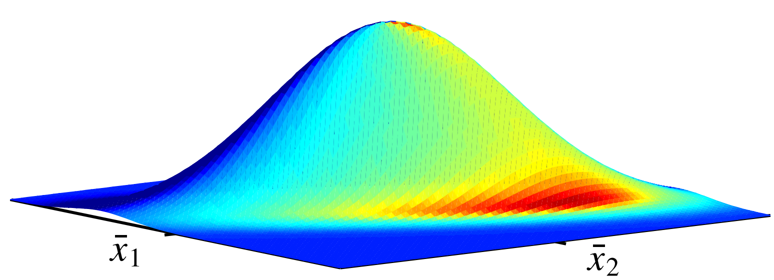
- Notice that correlation and covariance are the same for zero-mean random vectors.
- The covariance entries have specific meaning:

$$\begin{aligned} (\Sigma_{\tilde{X}})_{ii} &= \sigma_{X_i}^2 \\ (\Sigma_{\tilde{X}})_{ij} &= \rho_{ij}\sigma_{X_i}\sigma_{X_j} = (\Sigma_{\tilde{X}})_{ji}. \end{aligned}$$

  - The diagonal entries are the variances of each vector component;
  - The correlation coefficient ρ_{ij} is a measure of linear dependence between X_i and X_j. $|\rho_{ij}| \le 1$.

The most important multivariable distribution for this course

- The multivariable Gaussian is of key importance for Kalman filtering.



$$f_X(x) = \frac{1}{(2\pi)^{n/2}\left|\Sigma_{\widetilde{X}}\right|^{1/2}} \exp\left(-\frac{1}{2}(x-\bar{x})^T \Sigma_{\widetilde{X}}^{-1}(x-\bar{x})\right).$$

$$|\Sigma_{\widetilde{X}}| = \det(\Sigma_{\widetilde{X}}), \quad \Sigma_{\widetilde{X}}^{-1} \text{ requires positive-definite } \Sigma_{\widetilde{X}}.$$

- Notation: $X \sim \mathcal{N}(\bar{x}, \Sigma_{\widetilde{X}})$.
- Contours of constant $f_X(x)$ are hyper-ellipsoids, centered at $\bar{x}$, directions governed by $\Sigma_{\widetilde{X}}$. Principle axes decouple $\Sigma_{\widetilde{X}}$ (eigenvectors).
- Two-dimensional zero-mean case: (Let $\sigma_1 = \sigma_{X_1}$ and $\sigma_2 = \sigma_{X_2}$)

$$\Sigma_{\widetilde{X}} = \begin{bmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \qquad |\Sigma_{\widetilde{X}}| = \sigma_1^2\sigma_2^2(1-\rho_{12}^2).$$

$$f_{X_1,X_2}(x_1,x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho_{12}^2}} \exp\left(-\frac{\frac{x_1^2}{\sigma_1^2} - 2\rho_{12}\frac{x_1}{\sigma_1}\frac{x_2}{\sigma_2} + \frac{x_2^2}{\sigma_2^2}}{2(1-\rho_{12}^2)}\right).$$

3.3: *Uncorrelated* versus *independent*

INDEPENDENCE: Iff jointly-distributed RVs are independent, then

$$f_X(x_1, x_2, \ldots, x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_n}(x_n).$$

- Joint distribution can be split up into the product of individual distributions for each RV
- Equivalent condition: For all functions $f(\cdot)$ and $g(\cdot)$,

$$\mathbb{E}[f(X_1)g(X_2)] = \mathbb{E}[f(X_1)]\mathbb{E}[g(X_2)].$$

- "The particular value of the random variable X_1 has no impact on what value we would obtain for the random variable X_2."

UNCORRELATED: Two jointly-distributed RVs X_1 and X_2 are uncorrelated if their second moments are finite and

$$\text{cov}(X_1, X_2) = \mathbb{E}[(X_1 - \bar{x}_1)(X_2 - \bar{x}_2)] = 0$$

which implies $\rho_{12} = 0$.

- Uncorrelated means that there is no linear relationship between X_1 and X_2.

MAIN POINT #1: If jointly-distributed RV X_1 and X_2 are independent then they are uncorrelated. Independence implies uncorrelation.

- To see this, notice that independence means $\mathbb{E}[X_1X_2] = \mathbb{E}[X_1]\mathbb{E}[X_2]$. Therefore,

$$\text{cov}(X_1, X_2) = \mathbb{E}[X_1X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2] = 0,$$

therefore uncorrelated.

- Does uncorrelation imply independence? Consider the example

$$Y_1 = \sin(2\pi X) \qquad \text{and} \qquad Y_2 = \cos(2\pi X)$$

  with X uniformly distributed on $[0, 1]$.

- We can show that $\mathbb{E}[Y_1] = \mathbb{E}[Y_2] = \mathbb{E}[Y_1 Y_2] = 0$.
- So, $\text{cov}(Y_1, Y_2) = 0$ and Y_1 and Y_2 are uncorrelated.
- But, $Y_1^2 + Y_2^2 = 1$ and therefore Y_1 and Y_2 are clearly not independent.
- Note: $\text{cov}(X_1, X_2) = 0$ (uncorrelated) implies that

$$\mathbb{E}[X_1 X_2] = \mathbb{E}[X_1]\mathbb{E}[X_2]$$

  but independence requires

$$\mathbb{E}[f(X_1)g(X_2)] = \mathbb{E}[f(X_1)]\mathbb{E}[g(X_2)]$$

  for all $f(\cdot)$ and $g(\cdot)$.

- Therefore, independence is *much* stronger than uncorrelated.

COROLLARY: Consider a RV X with uncorrelated components. The covariance matrix $\Sigma_{\tilde{X}} = \mathbb{E}[(X - \bar{x})(X - \bar{x})^T]$ is diagonal.

PROOF: Notice that:

- The diagonal elements are $(\Sigma_{\tilde{X}})_{ii} = \mathbb{E}[(X_i - \bar{x}_i)^2] = \sigma_i^2$.
- The off-diagonal elements are

$$\begin{aligned}(\Sigma_{\tilde{X}})_{ij} &= \mathbb{E}[(X_i - \bar{x}_i)(X_j - \bar{x}_j)] \\ &= \mathbb{E}[X_i X_j - X_i \bar{x}_j - \bar{x}_i X_j + \bar{x}_i \bar{x}_j] \\ &= \mathbb{E}[X_i]\mathbb{E}[X_j] - 2\mathbb{E}[X_i]\mathbb{E}[X_j] + \mathbb{E}[X_i]\mathbb{E}[X_j] = 0.\end{aligned}$$

MAIN POINT #2: If jointly <u>normally</u> distributed RVs are uncorrelated, then they are independent. (This is a special case.)

3.4: Functions of random variables

MAIN POINT #3: Suppose we are given two random variables Y and X with $Y = g(X)$. Also assume that g^{-1} exists, g and g^{-1} are continuously differentiable, then

$$f_Y(y) = f_X(g^{-1}(y)) \left\|\left| \frac{\partial g^{-1}(y)}{\partial y} \right|\right\|,$$

where $\|\,|\cdot|\,\|$ means to take the absolute value of the determinant.

- So what? Say we know the pdf of X_1 and X_2 and

$$\left.\begin{aligned} Y_1 &= X_1 + X_2 \\ Y_2 &= X_2 \end{aligned}\right\} \quad Y = AX$$

we can now form the pdf of Y ➠ handy!

EXAMPLE: $Y = kX$ and $X \sim \mathcal{N}(0, \sigma_X^2)$.

- $X = \frac{1}{k}Y$ and so $g^{-1}(y) = \frac{1}{k}y$. Then, $\dfrac{\partial g^{-1}(y)}{\partial y} = \dfrac{1}{k}$.

$$\begin{aligned} f_Y(y) = f_X\left(\frac{y}{k}\right)\left|\frac{1}{k}\right| &= \frac{1}{|k|}\frac{1}{\sqrt{2\pi}\,\sigma_X}\exp\left(\frac{-(y/k)^2}{2\sigma_X^2}\right) \\ &= \frac{1}{\sqrt{2\pi\,(\sigma_X k)^2}}\exp\left(\frac{-y^2}{2(\sigma_X k)^2}\right). \end{aligned}$$

- Therefore, multiplication by gain k results in a normally-distributed RV with scale change in standard deviation. $Y \sim \mathcal{N}(0, k^2\sigma_X^2)$.

EXAMPLE: $Y = AX + B$ where A is a constant (non-singular) matrix, B is a constant vector, and $X \sim \mathcal{N}(\bar{x}, \Sigma_{\tilde{X}})$.

- $X = A^{-1}Y - A^{-1}B$ so $g^{-1}(y) = A^{-1}y - A^{-1}B$. Then, $\dfrac{\partial g^{-1}(y)}{\partial y} = A^{-1}$.

- $f_X(x) = \dfrac{1}{(2\pi)^{n/2}|\Sigma_{\tilde{X}}|^{1/2}} \exp\left[-\dfrac{1}{2}(x-\bar{x})^T \Sigma_{\tilde{X}}^{-1}(x-\bar{x})\right].$
- Therefore,

$$\begin{aligned} f_Y(y) &= \frac{|A^{-1}|}{(2\pi)^{n/2}|\Sigma_{\tilde{X}}|^{1/2}} \exp\left[-\frac{1}{2}(A^{-1}(y-B)-\bar{x})^T \Sigma_{\tilde{X}}^{-1}(A^{-1}(y-B)-\bar{x})\right] \\ &= \frac{1}{(2\pi)^{n/2}(|A||\Sigma_{\tilde{X}}||A^T|)^{1/2}} \exp\left[-\frac{1}{2}(y-\bar{y})^T (A^{-1})^T \Sigma_{\tilde{X}}^{-1} A^{-1}(y-\bar{y})\right] \\ &= \frac{1}{(2\pi)^{n/2}|\Sigma_{\tilde{Y}}|^{1/2}} \exp\left[-\frac{1}{2}(y-\bar{y})\Sigma_{\tilde{Y}}^{-1}(y-\bar{y})\right], \end{aligned}$$

if $\Sigma_{\tilde{Y}} = A\Sigma_{\tilde{X}}A^T$ and $\bar{y} = A\bar{x} + B$. That is, $Y \sim \mathcal{N}(A\bar{x}+B, A\Sigma_{\tilde{X}}A^T)$.

CONCLUSION: Sum of Gaussians is Gaussian—very special case.

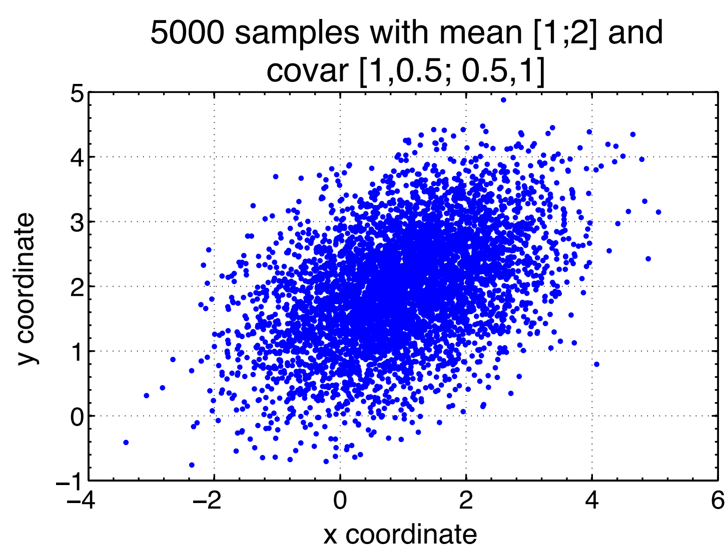
RANDOM NOTE: How to use `randn.m` to simulate non-zero mean Gaussian noise with covariance $\Sigma_{\tilde{Y}}$?

- $Y \sim \mathcal{N}(\bar{y}, \Sigma_{\tilde{Y}})$ but `randn.m` returns $X \sim \mathcal{N}(0, I)$.
- Try $y = \bar{y} + A^T x$ where A is square with the same dimension as $\Sigma_{\tilde{Y}}$; $A^T A = \Sigma_{\tilde{Y}}$. ($A$ is the Cholesky decomposition of positive-definite symmetric matrix $\Sigma_{\tilde{Y}}$).

```
ybar = [1; 2];
covar = [1, 0.5; 0.5, 1];
A = chol(covar);
x = randn([2, 1]);
y = ybar + A'*x;
```

- When $\Sigma_{\tilde{Y}}$ is non-positive definite (but also non-negative definite)

```
[L,D] = ldl(covar);
x = randn([2,5000]);
y = ybar(:,ones([1 5000]))
  +(L*sqrt(D))*x;
```



MAIN POINT #4: Normality is preserved in a linear transformation (extension of above example).

- Let X and W be independent vector random variables, a is a constant (vector): $X \sim \mathcal{N}(\bar{x}, \Sigma_{\tilde{X}})$, $W \sim \mathcal{N}(\bar{w}, \Sigma_{\tilde{W}})$.
- Let

$$Z = AX + BW + a$$

$$\mathbb{E}[Z] = \bar{z} = A\bar{x} + B\bar{w} + a$$

$$\Sigma_{\tilde{Z}} = A\Sigma_{\tilde{X}}A^T + B\Sigma_{\tilde{W}}B^T.$$

- Z is also a normal RV and $Z \sim \mathcal{N}(\bar{z}, \Sigma_{\tilde{Z}})$.

3.5: Conditioning

MAIN POINT #5: Conditional probabilities.

- Given jointly-distributed RVs, it is often of *extreme* interest to find the pdf of some of the RVs given known values for the rest.
- For example, given the joint pdf $f_{X,Y}(x, y)$ for RVs X and Y, we want the conditional pdf of $f_{X|Y}(x \mid y)$ which is the pdf of X for a known value $Y = y$.
- Can also think of it as $f_{X|Y}(X = x \mid Y = y)$.

DEFINE: Conditional pdf

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

is the probability that $X = x$ given that $Y = y$ has happened.

NOTE I: The marginal probability $f_Y(y)$ may be calculated as

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,\mathrm{d}x.$$

For each y, integrate out the effect of X.

NOTE II: If X, Y independent, $f_{X,Y}(x, y) = f_X(x) f_Y(y)$. Therefore

$$f_{X|Y}(x \mid y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \frac{f_X(x) f_Y(y)}{f_Y(y)} = f_X(x).$$

Knowing that $Y = y$ has occurred provides no information about X. (Is this what you would expect?)

DIRECT EXTENSION:

$$\begin{aligned} f_{X,Y}(x, y) &= f_{X|Y}(x \mid y) f_Y(y) \\ &= f_{Y|X}(y \mid x) f_X(x), \end{aligned}$$

Therefore,

$$f_{X|Y}(x \mid y) = \frac{f_{Y|X}(y \mid x) f_X(x)}{f_Y(y)}.$$

- This is known as Bayes' rule. It relates the *a posteriori* probability to the *a priori* probability.
- It forms a key step in the Kalman filter derivation.

MAIN POINT #6: Conditional expectation.

- Now that we have a way of expressing a conditional probability, we can also express a conditional mean.
- What do we expect the value of X to be given that $Y = y$ has happened?

$$\mathbb{E}[X = x \mid Y = y] = \mathbb{E}[X \mid Y] = \int_{-\infty}^{\infty} x f_{X|Y}(x \mid Y)\,\mathrm{d}x.$$

- Conditional expectation is not a constant (as expectation is) but a random variable. It is a function of the conditioning random variable (*i.e.,* Y).
- Note: Conditional expectation is *critical*. The Kalman filter is an algorithm to compute $\mathbb{E}[x_k \mid \mathbb{Z}_k]$, where we define $\mathbb{Z}_k$ later.

MAIN POINT #7: Central limit theorem.

- If $Y = \sum_i X_i$ and the X_i are independent and identically distributed IID), and the X_i have finite mean and variance, then Y will be approximately normally distributed.
- The approximation improves as the number of summed RVs gets large.

- Since the state of our dynamic system adds up the effects of lots of independent random inputs, it is reasonable to assume that the distribution of the state tends to the normal distribution.
- This leads to the key assumptions for the derivation of the Kalman filter, as we will see:
  - We will assume that state x_k is a normally-distributed RV;
  - We will assume that process noise w_k is a normally-distributed RV;
  - We will assume that sensor noise v_k is a normally-distributed RV;
  - We will assume that w_k and v_k are uncorrelated with each other.
- Even when these assumptions are broken in practice, the Kalman filter often works quite well.
- Exceptions to this rule tend to be with very highly nonlinear systems, for which particle filters must sometimes be employed to get good estimates.

3.6: Vector random (stochastic) processes

- A stochastic random process is a family of random vectors indexed by a parameter set ("time" in our case).
  - For example, the random process $X(t)$.
  - The value of the random process at any time t_0 is RV $X(t_0)$.
- Usually assume stationarity.
  - The pdf of the random variables are time-shift invariant.
  - Therefore, $\mathbb{E}[X(t)] = \bar{x}$ for all t and $\mathbb{E}[X(t_1)X^T(t_2)] = R_X(t_1 - t_2)$.

Properties

1. Autocorrelation: $R_X(t_1, t_2) = \mathbb{E}[X(t_1)X^T(t_2)]$. If stationary,

$$R_X(\tau) = \mathbb{E}[X(t)X^T(t+\tau)].$$

   - Provides a measure of correlation between elements of the process having time displacement τ.
   - $R_X(0) = \sigma_X^2$ for zero-mean X.
   - $R_X(0)$ is always the maximum value of $R_X(\tau)$.

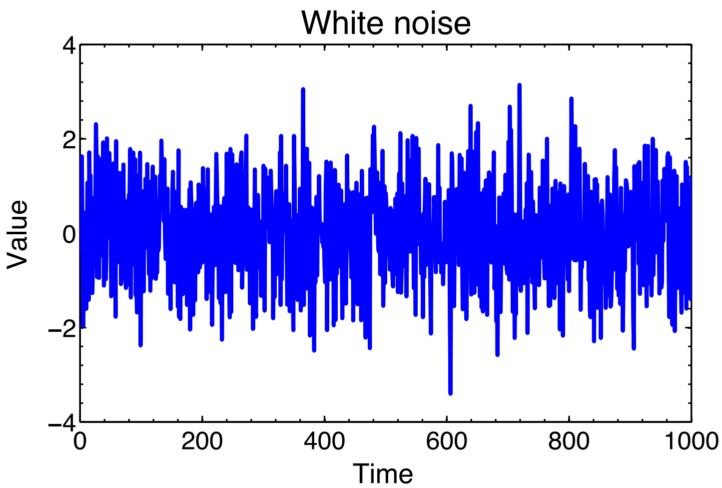
2. Autocovariance: $C_X(t_1, t_2) = \mathbb{E}[(X(t_1) - \mathbb{E}[X(t_1)])(X(t_2) - \mathbb{E}[X(t_2)])^T]$. If stationary,

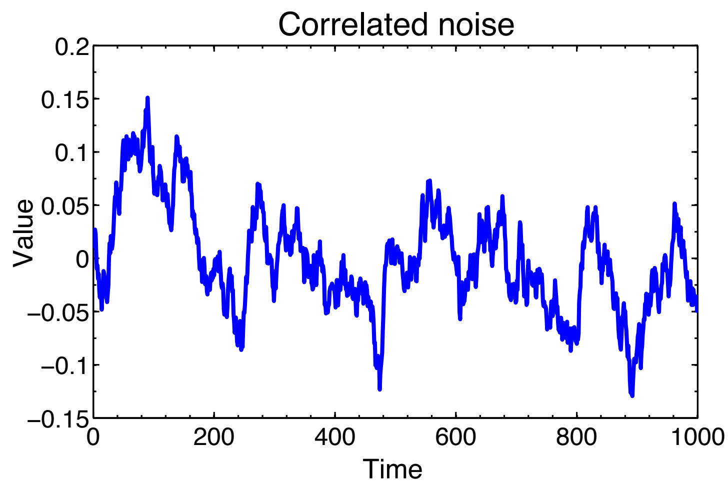
$$C_X(\tau) = \mathbb{E}[(X(t) - \bar{x})(X(t+\tau) - \bar{x})^T].$$

"White" noise

- Correlation from one time instant to the next is a key property of a random process. ➠ $R_X(\tau)$ and $C_X(\tau)$ very important.

- Some processes have a unique autocorrelation:
  1. Zero mean,
  2. $R_X(\tau) = \mathbb{E}[X(t)X(t+\tau)^T] = S_X\delta(\tau)$ where $\delta(\tau)$ is the Dirac delta.
- Therefore, the process is uncorrelated in time.
- Clearly an *abstraction*, but proves to be a *very* useful one.





Power spectral density (PSD)

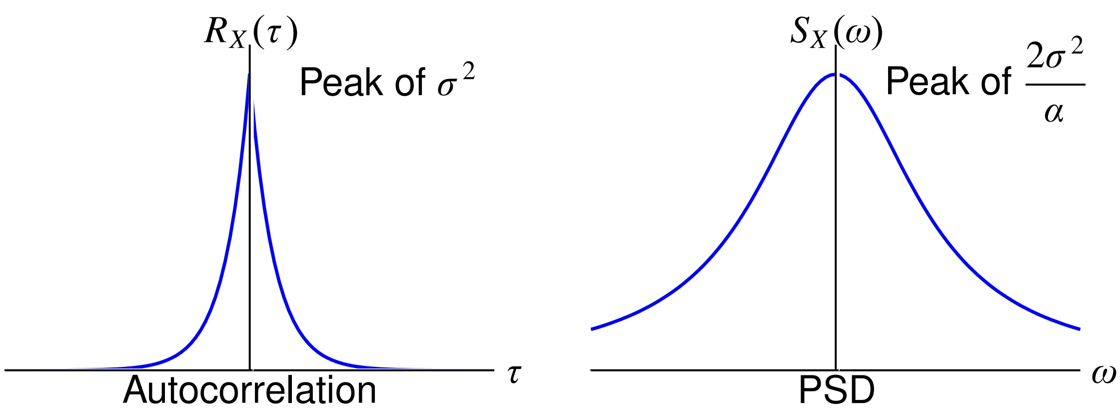
- Consider stationary random processes with autocorrelation defined as $R_X(\tau) = \mathbb{E}[X(t)X(t+\tau)^T]$.
- If a process varies slowly, time-adjacent samples are very correlated, and $R_X(\tau)$ will drop off slowly ➠ Mostly low-frequency content.
- If a process varies quickly, time-adjacent samples aren't very correlated, and $R_X(\tau)$ drops off quickly ➠ More high-freq. content.
- Therefore, the autocorrelation function tells us about the frequency content of the random signal.
- Consider scalar case. We define the power spectral density (PSD) as the Fourier transform of the autocorrelation function:

$$S_X(\omega) = \int_{-\infty}^{\infty} R_X(\tau) e^{-j\omega\tau}\,\mathrm{d}\tau$$
$$R_X(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} S_X(\omega) e^{j\omega\tau}\,\mathrm{d}\omega.$$

- $S_X(\omega)$ gives a frequency-domain interpretation of the "power" (energy) in the random process $X(t)$, and is real, symmetric and positive definite for real random processes.

EXAMPLE: $R_X(\tau) = \sigma^2 e^{-\alpha|\tau|}$ for $\alpha > 0$. Then,

$$S_X(\omega) = \int_{-\infty}^{\infty} \sigma^2 e^{-\alpha|\tau|} e^{-j\omega\tau}\,\mathrm{d}\tau = \frac{2\sigma^2\alpha}{\omega^2+\alpha^2}.$$



- α is a bandwidth measure.

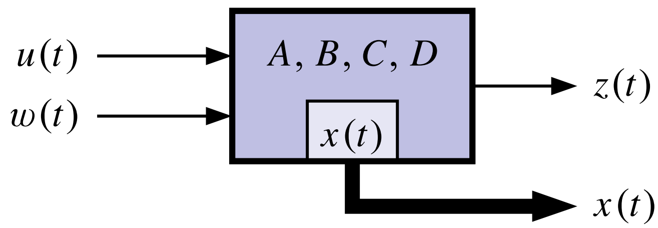
INTERPRETATION: 1. Area under PSD $S_X(\omega)$ for $\omega_1 \le \omega \le \omega_2$ provides a measure of energy in the signal in that frequency range.

2. White noise can be thought of as a limiting process as $\alpha \to \infty$. From our prior example, let $\alpha \to \infty$.

- As $\alpha \to \infty$, $R_X(\tau) \to \delta(\tau)$; therefore white.
- Bandwidth of spectral content is approximately α, so goes to infinity.
- Abstraction since implies that signal has infinite energy!

3.7: Discrete-time dynamic systems with random inputs

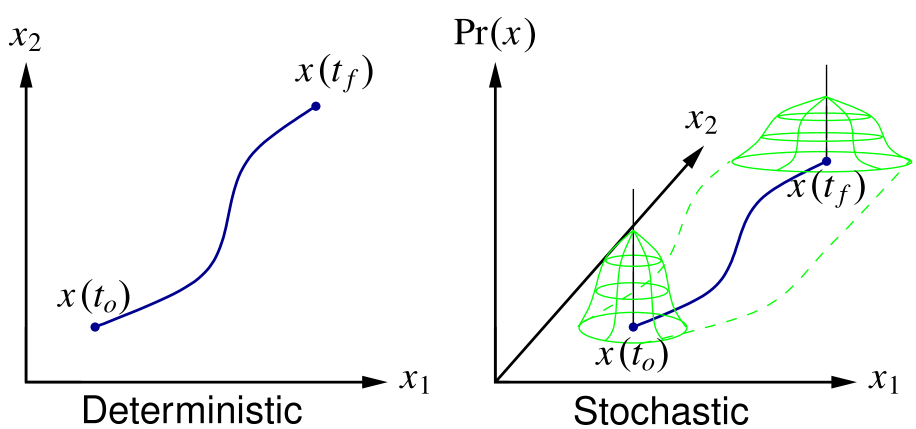
- What are the statistics of the state of a system driven by a random process?
- That is, how do the system dynamics influence the time-propagation of $x(t)$? ➠ Do not know $x(t)$ exactly, but can develop a pdf for $x(t)$.
- Primary example: A linear system driven by white noise $w(t)$, possibly in addition to deterministic input $u(t)$.



- Contrast to deterministic simulation, which can be easily simulated

$$\dot{x}(t) = Ax(t) + Bu(t), \qquad x(0) = x_0, \qquad x(0),\ u(t) \text{ known.}$$

- System response is completely specified. Therefore, no uncertainty.
- The stochastic propagation is different since there are inputs that are not known exactly. ➠ Best we can do is to say how uncertainty in the state changes with time.



- Deterministic: $x(t_0)$ and inputs known ➠ State trajectory deterministic.

- Stochastic: pdf for $x(t_0)$ and inputs known ➟ Can find only pdf for $x(t)$.
- We will work with Gaussian noises to a large extent, which are uniquely defined by the first- and second central moments of the statistics ➟ Gaussian assumption not essential.
- We will only ever track the first two moments.

NOTATION: Until now, we have always used capital letters for random variables. The state of a system driven by a random process is a random vector, so we could now call it $X(t)$ or X_k. However, it is more common to retain the standard notation $x(t)$ or x_k and understand from the context that we are now discussing an RV

Discrete-time systems

- We will start with a discrete system and then look at how to get an equivalent answer for continuous systems.
- Model:

$$x_k = A_{d,k-1}x_{k-1} + B_{d,k-1}u_{k-1} + w_{k-1},$$

where

x_k : State vector, a random process.

u_k : Deterministic control inputs.

w_k : Noise that drives the process (process noise), a random process.

- A_d and B_d assumed known. Can be time-varying. For now, assume fixed. Generalize later if necessary.
- Key assumptions about driving noise:

1. Zero mean: $\mathbb{E}[w_k] = 0 \quad \forall \quad k$.
2. White: $\mathbb{E}[w_{k_1} w_{k_2}^T] = \Sigma_{\widetilde{w}} \Delta(k_1 - k_2)$.

- $\Sigma_{\widetilde{w}}$ is called the spectral density of the noise signal w_k.
- Uncertainty at startup: Statistics of initial condition.

$$\mathbb{E}[x_0] = \bar{x}_0;$$

$$\mathbb{E}[(x_0 - \bar{x}_0) w_k^T] = 0 \quad \forall \quad k$$

$$\Sigma_{\tilde{x},0} = \mathbb{E}[(x_0 - \bar{x}_0)(x_0 - \bar{x}_0)^T].$$

- Key question: How do we propagate the statistics of this system?

Mean value:

- The mean value is

$$\begin{aligned}\mathbb{E}[x_k] = \bar{x}_k &= \mathbb{E}[A_d x_{k-1} + B_d u_{k-1} + w_{k-1}] \\ &= A_d \bar{x}_{k-1} + B_d u_{k-1}.\end{aligned}$$

Therefore, the mean propagation is

$$\begin{aligned}\bar{x}_0 &: \text{ Given} \\ \bar{x}_k &= A_d \bar{x}_{k-1} + B_d u_{k-1}.\end{aligned}$$

Deterministic simulation and mean values treated the same way.

Variations about mean

- To study the random variations about the mean, we need to form the second central moment of the statistics.

$$\Sigma_{\tilde{x},k} = \mathbb{E}[(x_k - \bar{x}_k)(x_k - \bar{x}_k)^T].$$

- Easiest to study if we note that

$$\begin{aligned} x_k - \bar{x}_k &= A_d x_{k-1} + B_d u_{k-1} + w_{k-1} \\ &\quad - A_d \bar{x}_{k-1} - B_d u_{k-1} \\ &= A_d(x_{k-1} - \bar{x}_{k-1}) + w_{k-1}. \end{aligned}$$

- Thus,

$$\Sigma_{\tilde{x},k} = \mathbb{E}[(A_d(x_{k-1} - \bar{x}_{k-1}) + w_{k-1})(A_d(x_{k-1} - \bar{x}_{k-1}) + w_{k-1})^T].$$

- Three terms:
  1. $\mathbb{E}[A_d(x_{k-1} - \bar{x}_{k-1})(x_{k-1} - \bar{x}_{k-1})^T A_d^T] = A_d \Sigma_{\tilde{x},k-1} A_d^T$.
  2. $\mathbb{E}[w_{k-1} w_{k-1}^T] = \Sigma_{\widetilde{w}}$.
  3. $\mathbb{E}[A_d(x_{k-1} - \bar{x}_{k-1}) w_{k-1}^T] = ?$

- The third term is a cross-correlation term.
  - But, x_{k-1} depends only on x_0 and inputs w_m for $m = 0 \ldots k-2$.
  - w_{k-1} is white noise uncorrelated with x_0.
  - Therefore, third term is zero.

Therefore, the covariance propagation is

$$\begin{aligned} \Sigma_{\tilde{x},0} &: \text{ Given} \\ \Sigma_{\tilde{x},k} &= A_d \Sigma_{\tilde{x},k-1} A_d^T + \Sigma_{\widetilde{w}}. \end{aligned}$$

In this equation, $A_d \Sigma_{\tilde{x},k-1} A_d^T$ is the homogeneous part; $\Sigma_{\widetilde{w}}$ is the driving term.

- Note: If A_d and $\Sigma_{\widetilde{w}}$ are constant and A_d stable, there is a steady-state solution.
  - As $k \to \infty$, $\Sigma_{\tilde{x},k} = \Sigma_{\tilde{x},k+1} = \Sigma_{\tilde{x}}$.

- Then, $\Sigma_{\tilde{x}} = A_d \Sigma_{\tilde{x}} A_d^T + \Sigma_{\widetilde{w}}$.

- This form is called a discrete Lyapunov equation. In MATLAB, `dlyap.m`

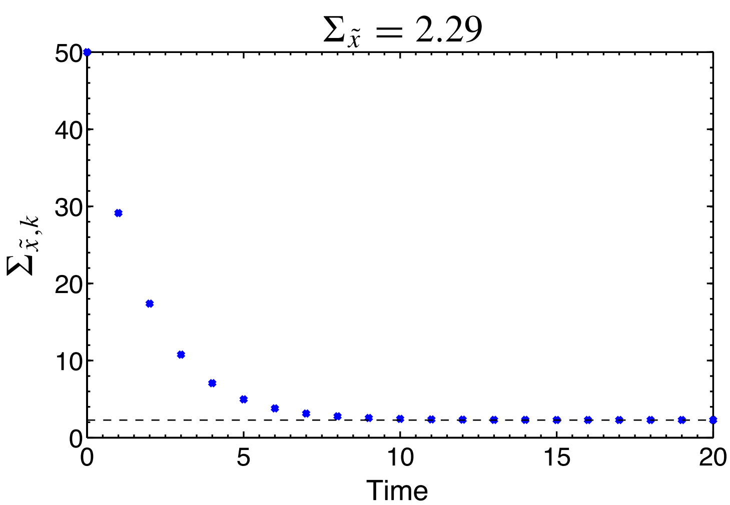
EXAMPLE: Can solve by hand in the scalar case.

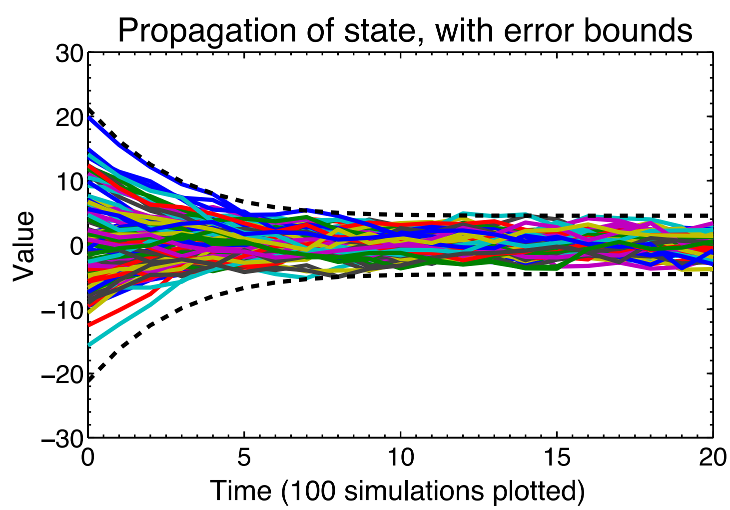
- Consider: $x_k = \alpha x_{k-1} + w_{k-1}$. $\mathbb{E}[w_{k-1}] = 0$; $\Sigma_{\widetilde{w}} = 1$; $\bar{x}_0 = 0$.
- Then,

$$\begin{aligned} \Sigma_{\tilde{x}} &= \alpha \Sigma_{\tilde{x}} \alpha^T + 1 \\ \Sigma_{\tilde{x}}(1-\alpha^2) &= 1 \\ \Sigma_{\tilde{x}} &= \frac{1}{1-\alpha^2}. \end{aligned}$$

  Valid for $|\alpha| < 1$; otherwise unstable.

- In the example below, we plot 100 random trajectories for $\alpha = 0.75$ and $\Sigma_{\tilde{x},0} = 50$.
- Compare propagation of $\Sigma_{\tilde{x},k}$ with steady-state `dlyap.m` solution.





- The error bounds plotted are 3σ bounds (*i.e.*, $\pm 3\sqrt{\Sigma_{\tilde{x},k}}$).

3.8: Continuous-time dynamic systems with random inputs

- For continuous-time systems we have

$$\dot{x}(t) = A(t)x(t) + B_u(t)u(t) + B_w(t)w(t),$$

where

$x(t)$: State vector, a random process.

$u(t)$: Deterministic control inputs.

$w(t)$: Noise driving the process (process noise), a random process.

Further:

$$\mathbb{E}[w(t)] = 0; \qquad \mathbb{E}[w(t)w(\tau)^T] = S_w\delta(t-\tau)$$

$$\mathbb{E}[x(0)] = \bar{x}(0); \qquad \mathbb{E}[(x(0)-\bar{x}(0))(x(0)-\bar{x}(0))^T] = \Sigma_{\tilde{x}}(0).$$

- A, B_u and B_w assumed known. Can be time-varying. For now, assume fixed. Generalize later if necessary.
- S_w is the spectral density of $w(t)$.
- Easiest analysis is to discretize model; use results obtained earlier; let $\Delta t \to 0$.
  - Drop deterministic inputs $u(t)$ for simplicity. (They affect only the mean, and in known ways.)

Continuous-time propagation of statistics

Mean value

- Starting with the discrete case, we have

$$\bar{x}_k = A_d\bar{x}_{k-1} + B_d u_{k-1}$$

$$A_d = e^{A\Delta t} \approx I + A\Delta t + \mathcal{O}(\Delta t^2).$$

- Let $\Delta t \to 0$ and consider case when $u_k = 0$ for all k.

$$\bar{x}_k = (A\Delta t + I)\bar{x}_{k-1}$$
$$\frac{\bar{x}_k - \bar{x}_{k-1}}{\Delta t} = A\bar{x}_{k-1}.$$

- As $\Delta t \to 0$

$$\dot{\bar{x}}(t) = A\bar{x}(t)$$

as expected.

Therefore, the mean propagation is

$$\begin{aligned} \bar{x}(0) &: \text{ Given} \\ \dot{\bar{x}}(t) &= A\bar{x}(t) + Bu(t). \end{aligned}$$

Deterministic simulation and mean values treated the same way.

Variations about mean

- The result for discrete-time systems was: $\Sigma_{\tilde{x},k} = A_d \Sigma_{\tilde{x},k-1} A_d^T + \Sigma_{\widetilde{w}}$.
- But, we need a way to relate discrete $\Sigma_{\widetilde{w}}$ to a continuous spectral density S_w before we can proceed.
- Recall the discrete system response in terms of continuous system matrices:

$$\begin{aligned} x_k &= e^{A\Delta t} x_{k-1} + \int_{(k-1)\Delta t}^{k\Delta t} e^{A(k\Delta t - \tau)} B_w w(\tau)\,\mathrm{d}\tau \\ &= e^{A\Delta t} x_{k-1} + w_{k-1}. \end{aligned}$$

- Integral explicitly accounts for variations in the noise during Δt.

RECALL: Discrete white noise of the form

$$\mathbb{E}[w_k w_l^T] = \begin{cases} \Sigma_{\widetilde{w}}, & k = l; \\ 0, & k \neq l. \end{cases}$$

- But noise input term to our converted state dynamics is defined as an integral

$$w_{k-1} = \int_{(k-1)\Delta t}^{k\Delta t} e^{A(k\Delta t-\tau)} B_w w(\tau)\,\mathrm{d}\tau .$$

- Form outer product to get equivalent $\Sigma_{\widetilde{w}}$

$$\begin{aligned}\Sigma_{\widetilde{w}} &= \mathbb{E}\left[\left(\int_{(k-1)\Delta t}^{k\Delta t} e^{A(k\Delta t-\tau)} B_w w(\tau)\,\mathrm{d}\tau\right)\left(\int_{(k-1)\Delta t}^{k\Delta t} e^{A(k\Delta t-\gamma)} B_w w(\gamma)\,\mathrm{d}\gamma\right)^T\right] \\ &= \mathbb{E}\left[\int_{(k-1)\Delta t}^{k\Delta t}\int_{(k-1)\Delta t}^{k\Delta t} e^{A(k\Delta t-\tau)} B_w w(\tau) w(\gamma)^T B_w^T e^{A^T(k\Delta t-\gamma)}\,\mathrm{d}\tau\,\mathrm{d}\gamma\right].\end{aligned}$$

- Big ugly mess but has two saving graces
  1. Expectation can go inside integrals.
  2. $\mathbb{E}[w(\tau)w(\gamma)^T] = S_w\delta(\tau-\gamma)$ ➠ One of the integrals drops out!

- So, we have

$$\Sigma_{\widetilde{w}} = \int_{(k-1)\Delta t}^{k\Delta t} e^{A(k\Delta t-\tau)} B_w S_w B_w^T e^{A^T(k\Delta t-\tau)}\,\mathrm{d}\tau .$$

KEY POINT: While S_w may have simple form, $\Sigma_{\widetilde{w}}$ will be full matrix in general.

- To solve integral, we can approximate: As $\Delta t \to 0$, then $k\Delta t - \tau \to 0$ and $e^{A(k\Delta t-\tau)} \approx I + A(k\Delta t - \tau) + \cdots$
- That is, $e^{A(k\Delta t-\tau)} \approx I$. Then,

$$\Sigma_{\widetilde{w}} \approx \left(B_w S_w B_w^T\right)\Delta t .$$

- We will see a better method to evaluate $\Sigma_{\widetilde{w}}$ when $\Delta t \neq 0$, but for now we continue with this result to determine the continuous-time system covariance propagation.

- Start with $\Sigma_{\tilde{x},k} = A_d \Sigma_{\tilde{x},k-1} A_d^T + \Sigma_{\widetilde{w}}$, and substitute $\Sigma_{\widetilde{w}} \approx B_w S_w B_w^T \Delta t$. Also, use $A_d \approx (I + A\Delta t)$

$$\begin{aligned}\Sigma_{\tilde{x},k} &\approx (I + A\Delta t)\Sigma_{\tilde{x},k-1}(I + A\Delta t)^T + B_w S_w B_w^T \Delta t \\ &= \Sigma_{\tilde{x},k-1} + \Delta t(A\Sigma_{\tilde{x},k-1} + \Sigma_{\tilde{x},k-1}A^T + B_w S_w B_w^T) + \mathcal{O}(\Delta t^2) \\ \frac{\Sigma_{\tilde{x},k} - \Sigma_{\tilde{x},k-1}}{\Delta t} &= A\Sigma_{\tilde{x},k-1} + \Sigma_{\tilde{x},k-1}A^T + B_w S_w B_w^T + \mathcal{O}(\Delta t).\end{aligned}$$

- As $\Delta t \to 0$

$$\dot{\Sigma}_{\tilde{x}}(t) = A\Sigma_{\tilde{x}}(t) + \Sigma_{\tilde{x}}(t)A^T + B_w S_w B_w^T,$$

initialized with $\Sigma_{\tilde{x}}(0)$.

- This is a matrix differential equation.
- Symmetric, so don't need to solve for every element.
- Two effects
  - $A\Sigma_{\tilde{x}}(t) + \Sigma_{\tilde{x}}(t)A^T$: Homogeneous part. Contractive for stable A. Reduces covariance.
  - $B_w S_w B_w^T$: Impact of process noise. Tends to increase covariance.
- Steady-state solution: Effects balance for systems with constant dynamics (A, B_w, S_w) and stable A.

$$A\Sigma_{\tilde{x},ss} + \Sigma_{\tilde{x},ss}A^T + B_w S_w B_w^T = 0.$$

- This is a continuous-time Lyapunov equation. In MATLAB, `lyap.m`

> The covariance propagation is
>
> $$\begin{aligned}\Sigma_{\tilde{x}}(0) &: \text{ Given} \\ \dot{\Sigma}_{\tilde{x}}(t) &= A\Sigma_{\tilde{x}}(t) + \Sigma_{\tilde{x}}(t)A^T + B_w S_w B_w^T.\end{aligned}$$
>
> In this equation, $A\Sigma_{\tilde{x}}(t) + \Sigma_{\tilde{x}}(t)A^T$ is the homogeneous part; $B_w S_w B_w^T$ is the driving term.

EXAMPLE: $\dot{x}(t) = Ax(t) + B_w w(t)$, where A and B_w are scalars.

- Then, $\dot{\Sigma}_{\tilde{x}} = 2A\Sigma_{\tilde{x}} + B_w^2 S_w$. This can be solved in closed form to find

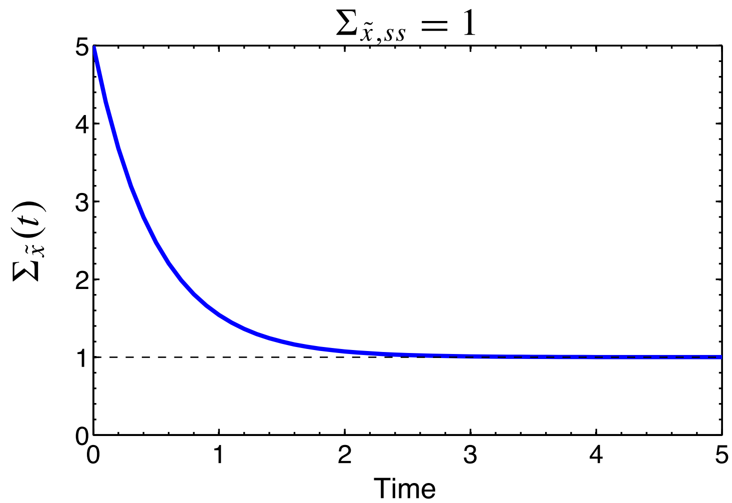
$$\Sigma_{\tilde{x}}(t) = \frac{B_w^2 S_w}{2A}(e^{2At} - 1) + \Sigma_{\tilde{x}}(0)e^{2At}.$$

- If $A < 0$ (stable) then the initial condition contribution goes to zero.

$$\Sigma_{\tilde{x},ss} = \frac{-B_w^2 S_w}{2A}$$

independent of $\Sigma_{\tilde{x}}(0)$.

- Example: Let $A = -1$, $\Sigma_{\tilde{x}}(0) = 5$. Find $\Sigma_{\tilde{x}}(t)$. Solution:
- Increased $\Sigma_{\tilde{x},ss}$ as more noise added: $B_w^2 S_w$ term.
- Decreased $\Sigma_{\tilde{x},ss}$ as A becomes "more stable".



3.9: Relating $\Sigma_{\widetilde{w}}$ to S_w precisely: A little trick

- Want to find $\Sigma_{\widetilde{w}}$ precisely, regardless of Δt. Assume we know S_w and plant dynamics.

$$\Sigma_{\widetilde{w}} = \int_0^{\Delta t} e^{A(\Delta t-\tau)} B_w S_w B_w^T e^{A^T(\Delta t-\tau)}\,\mathrm{d}\tau.$$

- Define

$$Z = \left[\begin{array}{c:c} -A & B_w S_w B_w^T \\ \hdashline 0 & A^T \end{array}\right], \qquad C = e^{Z\Delta t} = \left[\begin{array}{c:c} c_{11} & c_{12} \\ \hdashline 0 & c_{22} \end{array}\right]$$

- We can compute $C = \mathcal{L}^{-1}\left[(sI - Z)^{-1}\right]\Big|_{t=\Delta t}$ and find that

$$c_{11} = \mathcal{L}^{-1}\left[(sI - z_{11})^{-1}\right]\Big|_{t=\Delta t} = \mathcal{L}^{-1}\left[(sI + A)^{-1}\right]\Big|_{t=\Delta t} = e^{-A\Delta t}$$

$$c_{12} = \mathcal{L}^{-1}\left[(sI - z_{11})^{-1} z_{12} (sI - z_{22})^{-1}\right]\Big|_{t=\Delta t}$$

$$= \mathcal{L}^{-1}\left[(sI + A)^{-1} B_w S_w B_w^T (sI - A^T)^{-1}\right]\Big|_{t=\Delta t}$$

$$c_{22} = \mathcal{L}^{-1}\left[(sI - z_{22})^{-1}\right]\Big|_{t=\Delta t} = \mathcal{L}^{-1}\left[(sI - A^T)^{-1}\right]\Big|_{t=\Delta t} = e^{A^T\Delta t} = A_d^T$$

- So, $c_{22}^T = A_d$. Also $c_{22}^T c_{12} = \Sigma_{\widetilde{w}}$. With one `expm.m` command, can quickly switch from continuous-time to discrete-time.

- To show the second identity, recognize that c_{12} is a convolution of two impulse responses: one due to the system $(sI + A)^{-1} B_w$ and the other due to the system $S_w B_w^T (sI - A^T)^{-1}$.

- The first of these has impulse response $I e^{-At} B_w$ and the second has impulse response $S_w B_w^T e^{A^T t}$. Convolving them we get

$$\int_0^{\Delta t} I A^{-\tau} B_w S_w B_w^T e^{A^T(\Delta t-\tau)}\,\mathrm{d}\tau.$$

- Finally, consider

$$c_{22}^T c_{12} = \int_0^{\Delta t} e^{A(\Delta t-\tau)} B_w S_w B_w^T e^{A^T(\Delta t-\tau)}\,\mathrm{d}\tau = \Sigma_{\widetilde{w}}.$$

EXAMPLE: Let

$$\dot{x}(t) = \begin{bmatrix} 0 & 1 \\ -1 & -1 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} w(t)$$

with $S_w = 0.06$ and $\Delta t = 2\pi/16$.

$$Z = \left[\begin{array}{cc:cc} 0 & -1 & 0 & 0 \\ 1 & 1 & 0 & 0.06 \\ \hdashline 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{array}\right]; \qquad C = \left[\begin{array}{cc:cc} 0.91 & -0.47 & -0.006 & -0.0046 \\ 0.47 & 1.38 & 0.0046 & 0.023 \\ \hdashline 0 & 0 & 0.93 & -0.32 \\ 0 & 0 & 0.32 & 0.62 \end{array}\right].$$

- So,

$$A_d = \begin{bmatrix} 0.93 & 0.32 \\ -0.32 & 0.62 \end{bmatrix}; \quad \Sigma_{\widetilde{w}} = \begin{bmatrix} 0.0009 & 0.0030 \\ 0.0030 & 0.0156 \end{bmatrix}; \quad B_w S_w B_w^T = \begin{bmatrix} 0 & 0 \\ 0 & 0.06 \end{bmatrix}.$$

- Exact $\Sigma_{\widetilde{w}}$ and approximate $\Sigma_{\widetilde{w}}$ very different due to large Δt.
- Compare predictions of steady-state performance.
  - Continuous:
    $$A\Sigma_{\tilde{x},ss} + \Sigma_{\tilde{x},ss}A^T + B_w S_w B_w^T = 0; \qquad \Sigma_{\tilde{x},ss} = \begin{bmatrix} 0.03 & 0 \\ 0 & 0.03 \end{bmatrix}.$$
  - Discrete: $\Sigma_{\tilde{x},ss} = A_d \Sigma_{\tilde{x},ss} A_d^T + \Sigma_{\widetilde{w}}$; $\qquad \Sigma_{\tilde{x},ss} = \begin{bmatrix} 0.03 & 0 \\ 0 & 0.03 \end{bmatrix}$.
- Because we used discrete white noise scaled to be equivalent to the continuous noise, the steady-state predictions are the same.
- Conversion now complete:

DISCRETE: A_d, $\Sigma_{\widetilde{w}}$, B_d

$$\text{Given } \Sigma_{\tilde{x},0} \;:\; \Sigma_{\tilde{x},k} = A_d \Sigma_{\tilde{x},k-1} A_d^T + \Sigma_{\widetilde{w}}$$

$$\text{Given } \bar{x}_0 \;:\; \bar{x}_k = A_d \bar{x}_{k-1} + B_d u_{k-1}.$$

CONTINUOUS: A, S_w, B_u, B_w

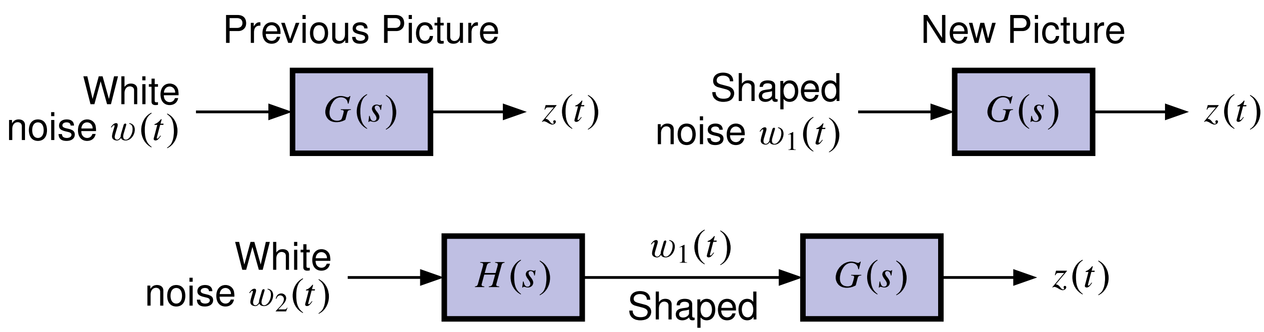
$$\text{Given } \Sigma_{\tilde{x}}(0) \;:\; \dot{\Sigma}_{\tilde{x}}(t) = A \Sigma_{\tilde{x}}(t) + \Sigma_{\tilde{x}}(t) A^T + B_w S_w B_w^T$$

$$\text{Given } \bar{x}(0) \;:\; \dot{\bar{x}}(t) = A\bar{x}(t) + B_u u(t).$$

- All matrices may be time varying.
- Equivalent $\Sigma_{\widetilde{w}}$ for given S_w available.

3.10: Shaping filters

- We have assumed that the inputs to the linear dynamic systems are white (Gaussian) noises.
- Therefore, the PSD is flat ➠ The input has content at all frequencies.
- Pretty limiting assumption, but one that can be easily fixed ➠ Can use second linear system to "shape" the noise and modify the PSD as desired.



- Therefore, we can drive our linear system with noise that has a desired PSD by introducing a shaping filter $H(s)$ that itself is driven by white noise.
- The combined system $GH(s)$ looks exactly the same as before, *but* the system $G(s)$ is not driven by pure white noise any more.
- Analysis quite simple: *Augment* original model with filter states.
- Original system has

$$\dot{x}(t) = Ax(t) + B_w w_1(t)$$

$$z(t) = Cx(t)$$

- New shaping filter with white input and desired PSD output has

$$\dot{x}_s(t) = A_s x_s(t) + B_s w_2(t)$$

$$w_1(t) = C_s x_s(t).$$

- Combine into one system

$$\begin{bmatrix} \dot{x}(t) \\ \hdashline \dot{x}_s(t) \end{bmatrix} = \left[\begin{array}{c:c} A & B_w C_s \\ \hdashline 0 & A_s \end{array}\right] \begin{bmatrix} x(t) \\ \hdashline x_s(t) \end{bmatrix} + \begin{bmatrix} 0 \\ \hdashline B_s \end{bmatrix} w_2(t)$$

$$z(t) = \left[\begin{array}{c:c} C & 0 \end{array}\right] \begin{bmatrix} x(t) \\ \hdashline x_s(t) \end{bmatrix}$$

- Augmented system just a larger-order linear system driven by white noise.

KEY QUESTION: Given a PSD for w_1, how do we find $H(s)$ (or, A_s, B_s, C_s).

- Fact: If w_2 is unit-variance white noise, then PSD for w_1 is

$$S_{w_1}(\omega) = H(-j\omega)H^T(j\omega).$$

- Find $H(j\omega)$ by <u>spectral factorization</u> of $S_{w_1}(\omega)$. Take the minimum-phase part (keep the poles and zeros that have negative real part).

EXAMPLE: Let

$$S_{w_1}(\omega) = \frac{2\sigma^2\alpha^2}{\omega^2 + \alpha^2}$$
$$= \frac{\sqrt{2}\sigma\alpha}{\alpha + j\omega} \cdot \frac{\sqrt{2}\sigma\alpha}{\alpha - j\omega}.$$

Then

$$H(s) = \frac{\sqrt{2}\sigma\alpha}{s + \alpha},$$

$$\dot{x}_s(t) = -\alpha x_s(t) + \sqrt{2}\alpha\sigma\, w_2(t)$$

$$w_1(t) = x_s(t).$$

EXAMPLE: Wind model.

- Wind gusts have some high-frequency content, but typically are dominated by low-frequency disturbances ➠ White noise a good model?
- We would typically think of a PSD for the wind that puts more emphasis on the low-frequency component of the signal.
- Think of this as the output of a shaping filter $H(s) = \dfrac{1}{\tau_w s + 1}$.

NOTE: If the bandwidth of the wind filter $(1/\tau_w) >$bandwidth of plant dynamics, then white noise is a good approximation to system input.

- Otherwise, must use shaping filters in plant model and drive with white noise.
- The size of the A matrix becomes a limitation. Use the simplest noise model possible.

Appendix: Plett notation versus textbook notation

- I use $\mathbb{E}[\cdot]$ to denote statistical expectation; the textbooks use $E[\cdot]$ or $E(\cdot)$.
- I use $\Sigma_{(\cdot)}$ to denote the correlation among the variables $(\cdot)$.
  - For example, $\Sigma_{XY} = \mathbb{E}[XY^T]$, and $\Sigma_X = \mathbb{E}[XX^T]$.
  - The same notation naturally applies to computing covariance,

$$\Sigma_{\widetilde{X}\widetilde{Y}} = \mathbb{E}[\widetilde{X}\widetilde{Y}^T] = \mathbb{E}[(X - \bar{x})(Y - \bar{y})^T]$$

    and

$$\Sigma_{\widetilde{X}} = \mathbb{E}[(X - \bar{x})(X - \bar{x})^T].$$

  - If the variables are zero mean, then $\Sigma_{\widetilde{X}\widetilde{Y}} = \Sigma_{XY}$, and $\Sigma_{\widetilde{X}} = \Sigma_X$.
- The textbooks tend to use different symbols for different covariances. For example,
  - cov(state error) $= \mathbb{E}[\tilde{x}\tilde{x}^T] = P$,
  - cov(process noise) $= Q$, and
  - cov(sensor noise) $= R$.