1 Introduction to Physical Measurements: Nature's Inspirations

Frame 1

The concept of "measuring" can have several meanings. For our purposes in physics, we are interested in two main aspects:

1. To ascertain or determine the extent, quantity, or dimensions of something in terms of agreed-upon units.

2. To assess or evaluate.

In this section, we'll explore how distances are estimated, drawing inspiration from how nature solves these problems, from astronomical scales down to micro scales. We'll look at a couple of examples from the animal kingdom.

Go to Frame 2.

Frame 2

Let's consider a chameleon hunting its prey, a fly. We can infer that the chameleon must estimate the distance to the fly with considerable accuracy. The text suggests an approximate accuracy requirement of $\Delta x = 100\,\mu$m.

Why is such accuracy important for the chameleon? [a] To impress other chameleons. [b] To avoid missing the fly or sticking its tongue to the wall. [c] To calculate the fly's speed.

Choose an answer and go to Frame 3.

Frame 3

Your answer was [a — b — c].

The correct answer is [b]. If the chameleon misjudges the distance, it might miss the fly entirely or, if the fly is near a surface, stick its adhesive tongue to the wall instead of catching the prey.

What mechanism allows the chameleon to judge this distance so precisely? The text first considers stereoscopic vision.

Go to Frame 4.

Frame 4

Stereoscopic Vision (Binocular Vision)

This is the process of seeing depth by using two eyes. If we denote the distance between the chameleon's eyes as d, and the angle subtended by the fly at its eyes (the apparent angular shift) as φ, the distance r to the fly could be estimated.

The text provides a formula for r:

$$r = \frac{d}{\varphi}$$

(Note: This formula is a simplification, typically $r \approx d/\tan(\varphi)$ or $r \approx d/\varphi$ for small angles φ in radians).

However, scientists tested this. They covered one of the chameleon's eyes, and it could still accurately strike its prey. What does this observation imply about stereoscopic vision being the *sole* mechanism for its depth perception? ______________________________

Go to Frame 5.

Frame 5

The observation that a chameleon with one eye covered can still accurately judge distance implies that stereoscopic vision is **not the sole or primary mechanism** it uses for this task, or at least it has a very effective backup. If it relied only on stereoscopic vision, covering one eye would significantly impair its depth perception.

So, what other mechanism might the chameleon use? The text suggests lens accommodation.

Go to Frame 6.

Frame 6

Lens Accommodation

This refers to the eye's ability to change the focal length of its lens to focus on objects at different distances. The lens equation is recalled:

$$\frac{1}{a} = \frac{1}{f} - \frac{1}{b}$$

where a is the object distance (distance to prey), b is the image distance (distance from lens to retina), and f is the focal length of the lens.

The text states that for a chameleon, the image distance b (lens to retina) is fixed. The chameleon can change the focal length f of its lens and "feels" this change.

How does changing the focal length f (while b is fixed) allow the chameleon to estimate the object distance a? ________________________________

Go to Frame 7.

Frame 7

By changing its lens's focal length f to bring the image of the prey into sharp focus on its fixed retina (distance b), the chameleon effectively solves the lens equation for a. The muscular effort or sensory feedback from changing f provides the information needed to determine a. This is how it can estimate the distance to its prey.

Now, let's switch to another amazing animal: the bat. Go to Frame 8.

Frame 8

Apex Predator: Bat

Bats use an active sonar system. With it, they can determine the distance r, speed v, and even the texture/composition of their prey. Let's look at some mechanisms.

Stereoscopic Hearing (Binaural Hearing) How does a bat (or human) determine the direction of a sound source using two ears? (Hint: Think about timing differences). ____________

Go to Frame 9.

Frame 9

A bat determines the direction of a sound source through the **time delay** between the sound arriving at its left and right ears. This difference in arrival time is processed in the neural pathway behind the ears, often involving an amplification at an off-center point in the neural network, allowing the predator to pinpoint the sound's origin.

Now, how does a bat measure distance using sonar? The text describes a "delay-line" mechanism. Go to Frame 10.

Frame 10

Distance Measurement via "Delay Line"

Sound travels at a constant speed c (approx. 340 m/s in air). For a model system to work, it should also have consistent dynamics. When a bat emits a vocalization (a pulse):

1. The signal is sent to its vocal cords.
2. Simultaneously, a copy of this outgoing signal is sent along an internal neural "delay line" where it travels at a certain speed v_d.
3. The bat's ears listen for the echo from the prey, which arrives on a "prompt line."

When the echo is heard, the signal from the prompt line and the signal travelling along the delay line will coincide at a specific neuron, causing it to fire strongly (e.g., twice the normal response).

Knowing the speed v_d on the delay line and the time elapsed, how can the bat estimate the distance to the prey? ______________________________

Go to Frame 11.

Frame 11

The bat estimates the distance based on where along the delay line the coincidence occurs. Different neurons along the delay line correspond to different delay times. The specific neuron that fires indicates the total round-trip time of the sound pulse to the prey and back. Since the speed of sound c is known (or calibrated), the distance can be calculated ($r = c \times (\text{time}/2)$).

The text describes different types of sonar pulses. Go to Frame 12.

Frame 12

First Sonar Mode: Ranging (Pinging at a distance) The bat emits ultrasonic pulses, about 200 per second. Each pulse is very short, on the order of 10 ms. This is good for determining distance by timing the echo.

Second Sonar Mode: FM Sweeps (Frequency Modulated) The bat emits a "chirp" or "whistle" that sweeps through a range of frequencies, even an entire octave (e.g., from ν_0 to $2\nu_0$). Why does the bat do this, and what kind of information can it get from FM sweeps that it might not get from simple pings? (Hint: Absorption and reflection depend on frequency and material). _

Go to Frame 13.

Frame 13

The FM sweeps allow the bat to determine the **texture or composition** of the target. Because sound absorption and reflection are frequency-dependent and material-dependent, the returning echo from an FM sweep will be "distorted" or "colored" differently based on what it reflected off. The bat doesn't receive a clean pulse back but a "damaged" one. By analyzing how different frequencies in the sweep are attenuated or reflected, the bat can infer properties about the target's surface and material. These FM sweeps are actually composed of even shorter pulses (around 0.2 ms), with pauses in between for listening to echoes.

Third Sonar Mode: Combination The bat combines the first two modes, likely using pings for initial detection and ranging, and FM sweeps for detailed analysis of potential prey.

Go to Frame 14.

Frame 14

Fourth Sonar Mode: Doppler Shift

The bat can also use the Doppler effect to determine the target's speed. If a target is moving, the frequency of the reflected sound wave will be shifted. The formula given for the observed frequency ν from a source frequency ν_s when there's relative velocity v (and c is speed of sound) is:

$$\nu = \nu_s \left(1 + 2\frac{v}{c}\right)$$

(Note: This formula applies when the source and observer are the same, and the target is reflecting the wave. The factor of 2 accounts for the sound traveling to the target and then back.)

By measuring the frequency shift $(\nu - \nu_s)$, the bat can determine v. Go to Frame 15.

Frame 15

Connection Between Real and Model Systems

In the case of the bat, we saw that in the real system, the dynamics of sound travel are described by: Distance $S = ct$ (where c is speed of sound, t is time).

In the model system (the bat's internal delay line), the signal travel is described by: Distance $S_M = v_d t$ (where v_d is speed on delay line).

It's important that the dynamics are consistent. The text states "S_M is a readable quantity (meaning we can check its value at any time)." The connection between the real system and the model system is called a **sensor** (in this case, the bat's ear and neural processing).

What is a crucial characteristic of a good sensor in this context? ____________________

Go to Frame 16.

Frame 16

A crucial characteristic of a good sensor is that it should **not significantly affect the real system** it is measuring. The act of sensing (the bat's ear receiving the echo) should not, for example, alter the path of the sound wave or the prey's behavior in a way that makes the measurement invalid.

This concludes our brief introduction to how some animals perform remarkable physical measurements. End of Section.

2 Optimal Filtering and Combining Measurements

Frame 1

In many scientific and engineering contexts, we deal with a real system S and try to represent it with a model system M. We observe a variable X_S in the real system, and its counterpart in the model is X_M. The goal of optimal filtering is to find the best way to estimate or optimize the model M based on observations from S.

What are some general requirements for this process to be effective? The text lists three:

1. Weak coupling between S and M (the model should have minimal influence on the real system).

2. X_M must be a "readable" quantity (its value should be accessible, possibly dependent on time t).

3. We need a way to assess the degree of agreement between the model and the real system.

A fourth, more dynamic requirement is that the dynamics (e.g., linear differential equations) governing X_S and X_M should be as similar as possible.

The text mentions a specific criterion for assessing agreement: $\lim_{t\to\infty}\langle (X_M - X_S)^2\rangle = \ldots$. What kind of value would we ideally want for this limit? [a] As large as possible [b] As small as possible (close to zero) [c] A specific non-zero constant

Go to Frame 2.

Frame 2

Your answer was [a — b — c].

The correct answer is [b]. We want the long-term average of the squared difference between the model and the real system to be as small as possible, ideally approaching zero. This indicates good agreement.

Now, let's consider how to optimally combine measurements. Go to Frame 3.

Frame 3

Suppose we have two separate observations (measurements) of a true value x. Let's call our measurement z. This measurement z consists of the true value x plus some measurement noise r:

$$z = x + r$$

We are interested in the statistical distribution of the measurement noise r. The text states that r often follows a Gaussian (Normal) distribution.

What are the two main parameters that define a Gaussian distribution? 1. ________ 2. ________

Go to Frame 4.

Frame 4

A Gaussian (Normal) distribution is characterized by its: 1. Mean (average value) 2. Standard deviation (or variance, which is the square of the standard deviation)

The text states the noise r is distributed as $N(0, \sigma)$, which means it's a Normal distribution with:

- Mean = 0
- Standard deviation = σ (so variance = σ^2)

The probability density function (PDF) is given by:

$$\frac{dP}{dr} = N(0,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{r^2}{2\sigma^2}}$$

If the mean of the noise $\langle r \rangle$ is 0, what does this imply about the average measured value $\langle z \rangle$ if $z = x + r$ and x is a constant true value? ____

Go to Frame 5.

Frame 5

If $\langle r \rangle = 0$, then $\langle z \rangle = \langle x + r \rangle = \langle x \rangle + \langle r \rangle = x + 0 = x$. On average, the measurement z will give the true value x.

The text calculates the expectation value (average) of r, $\langle r \rangle$:

$$\langle r \rangle = \int_{-\infty}^{\infty} r \frac{dP}{dr} dr = 0$$

This is because r is an odd function and $e^{-r^2/(2\sigma^2)}$ is an even function, so their product integrated over a symmetric interval is zero.

What about the expectation value of r^2, denoted $\langle r^2 \rangle$? This is the variance of the noise. The text shows:

$$\langle r^2 \rangle = \int_{-\infty}^{\infty} r^2 \frac{dP}{dr} dr = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} r^2 e^{-r^2/(2\sigma^2)} dr$$

After a change of variables ($u^2 = r^2/(2\sigma^2)$, so $r = \sqrt{2}\sigma u$, $dr = \sqrt{2}\sigma du$), this integral evaluates to:

$$\langle r^2 \rangle = \sigma^2$$

This confirms that σ^2 is indeed the variance of the noise.

Go to Frame 6.

Frame 6

Now, suppose we have two such independent measurements, z_1 and z_2, with noise r_1 and r_2 respectively. $z_1 = x + r_1$, where $r_1 \sim N(0, \sigma_1)$ $z_2 = x + r_2$, where $r_2 \sim N(0, \sigma_2)$

We can write these as: Measurement 1: $(\bar{z}_1, \sigma_1)$ Measurement 2: $(\bar{z}_2, \sigma_2)$ (Here, $\bar{z}_i$ would be the measured value, which is an estimate of x, and σ_i is the standard deviation of the noise associated with that measurement).

The text reminds us of the Central Limit Theorem (CLT). What is the general idea of the CLT when summing many independent random variables? ____________________________________

Go to Frame 7.

Frame 7

The Central Limit Theorem (CLT) states that the sum (or average) of a large number of independent and identically distributed random variables will tend to be normally (Gaussian) distributed, regardless of the original distribution of the individual variables (as long as they have finite variance).

Example: "Brumov šum" (AC Hum Noise) The text gives an example of AC power supply hum as a type of noise, sometimes called "Brum" (likely referring to "hum" as "brummen" in German). This noise affects measurements. Let's say the voltage due to this hum can be described as:

$$U = U_0 \cos(\omega t)$$

The probability of making a measurement at a specific time t within one half-period $[0, T/2]$ (where $T = 2\pi/\omega$) is $dP/dt = 1/(T/2) = 2/T$, assuming any time is equally likely.

The text then shows a change of variables from t to U to find the probability distribution of U, dP/dU. $dU = -U_0\omega \sin(\omega t)dt$.

$$\frac{dP}{dU} = \frac{dP/dt}{dU/dt} = \frac{2/T}{-U_0\omega\sin(\omega t)} = \frac{2/T}{-U_0\omega\sqrt{1-\cos^2(\omega t)}} = \frac{2/T}{-U_0\omega\sqrt{1-(U/U_0)^2}} = \frac{1}{\pi\sqrt{U_0^2 - U^2}}$$

Is this distribution Gaussian? [Yes — No]

Go to Frame 8.

Frame 8

No, the distribution $dP/dU = \frac{1}{\pi\sqrt{U_0^2-U^2}}$ (for $-U_0 < U < U_0$) is an arcsine distribution, not Gaussian. It has peaks at $U = \pm U_0$.

However, the text notes: "Fortunately, the Central Limit Theorem saves us. So, if we have many contributions from hum, they will tend towards a Gaussian distribution. This is noticeable even for, say, $N = 10$ contributions." This means even if individual noise sources aren't Gaussian, their combined effect often can be approximated as Gaussian.

Go to Frame 9.

Frame 9

Averaging

Suppose we have N measurements z_i, each normally distributed around the true value x with the same variance σ^2:

$$\frac{dP}{dz_i} = N(x, \sigma)$$

We can form an average (mean) of these measurements:

$$\bar{z} = \frac{1}{N}\sum_{i=1}^{N} z_i$$

We know that the expected value of the noise for each measurement is $\langle z_i - x\rangle = \langle r_i\rangle = 0$. What is the expected value of the noise of the average, $\langle \bar{z} - x\rangle$? ____________________

Go to Frame 10.

Frame 10

$$\langle \bar{z} - x\rangle = \langle \frac{1}{N}\sum z_i - x\rangle = \langle \frac{1}{N}\sum(x + r_i) - x\rangle = \langle \frac{1}{N}(Nx + \sum r_i) - x\rangle$$
$$= \langle x + \frac{1}{N}\sum r_i - x\rangle = \langle \frac{1}{N}\sum r_i\rangle = \frac{1}{N}\sum\langle r_i\rangle = \frac{1}{N}\sum 0 = 0$$

The average $\bar{z}$ is also an unbiased estimator of x.

Now, let's consider the variance of this average. The variance of a single measurement z_i is $\langle(z_i - x)^2\rangle = \sigma^2$. What is the variance of $\bar{z}$? The text shows the calculation:

$$\langle(\bar{z} - x)^2\rangle = \langle\left(\frac{1}{N}\sum(z_i - x)\right)^2\rangle = \frac{1}{N^2}\langle\left(\sum r_i\right)^2\rangle$$

$$= \frac{1}{N^2}\langle\sum r_i^2 + \sum_{i\neq j} r_i r_j\rangle$$

If the noises r_i are independent, then $\langle r_i r_j\rangle = \langle r_i\rangle\langle r_j\rangle = 0 \cdot 0 = 0$ for $i \neq j$. So, the cross terms vanish.

$$\langle(\bar{z} - x)^2\rangle = \frac{1}{N^2}\sum\langle r_i^2\rangle = \frac{1}{N^2}\sum\sigma^2 = \frac{1}{N^2}(N\sigma^2) = \frac{\sigma^2}{N}$$

Thus, the distribution of the average $\bar{z}$ is $N(x, \sigma/\sqrt{N})$. The average has the same mean x but a *smaller* standard deviation.

This means averaging multiple measurements reduces the uncertainty. Go to Frame 11.

Frame 11

Now, suppose we have results from two sets of measurements:

- N measurements give an average $\bar{z}_1$ with standard deviation $\sigma_1 = \sigma/\sqrt{N}$.
- M measurements give an average $\bar{z}_2$ with standard deviation $\sigma_2 = \sigma/\sqrt{M}$.

We want to combine $\bar{z}_1$ and $\bar{z}_2$ optimally to get a new estimate $\bar{z}_3$. If we simply took all $N + M$ original measurements and averaged them, the new standard deviation would be $\sigma_3 = \sigma/\sqrt{N+M}$.

The text forms a weighted average:

$$\bar{z}_3 = \left(\frac{N}{N+M}\right)\bar{z}_1 + \left(\frac{M}{N+M}\right)\bar{z}_2$$

Let's express the weights in terms of variances $\sigma_1^2 = \sigma^2/N$ and $\sigma_2^2 = \sigma^2/M$. So $N = \sigma^2/\sigma_1^2$ and $M = \sigma^2/\sigma_2^2$. The total variance $\sigma_3^2 = \sigma^2/(N+M)$. This leads to the relationship:

$$\frac{1}{\sigma_3^2} = \frac{N+M}{\sigma^2} = \frac{N}{\sigma^2} + \frac{M}{\sigma^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}$$

The optimal combined estimate $\bar{z}_3$ using these variances is:

$$\bar{z}_3 = \frac{\sigma_3^2}{\sigma_1^2}\bar{z}_1 + \frac{\sigma_3^2}{\sigma_2^2}\bar{z}_2 = \left(\frac{1/\sigma_1^2}{1/\sigma_1^2 + 1/\sigma_2^2}\right)\bar{z}_1 + \left(\frac{1/\sigma_2^2}{1/\sigma_1^2 + 1/\sigma_2^2}\right)\bar{z}_2$$

$$\bar{z}_3 = \left(\frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}\right)\bar{z}_1 + \left(\frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}\right)\bar{z}_2$$

This shows that measurements with smaller variance (i.e., more precise) are given higher weight.

This can also be written recursively. If $\bar{z}_1$ is our current best estimate and $\bar{z}_2$ is a new measurement (or new average), the updated estimate is:

$$\bar{z}_{\text{new}} = \bar{z}_{\text{old}} + \frac{\sigma_{\text{old}}^2}{\sigma_{\text{old}}^2 + \sigma_{\text{new\_meas}}^2}(\bar{z}_{\text{new\_meas}} - \bar{z}_{\text{old}})$$

The term $(\bar{z}_{\text{new\_meas}} - \bar{z}_{\text{old}})$ is called the "innovation."

Go to Frame 12.

Frame 12

Quadratic Form (Least Squares Approach)

Another way to find the optimal combination of two measurements $z_1 \sim N(x, \sigma_1)$ and $z_2 \sim N(x, \sigma_2)$ is to minimize a cost function, often called $2J(x)$ (the sum of squared errors, weighted by their variances):

$$2J(x) = \frac{(z_1 - x)^2}{\sigma_1^2} + \frac{(z_2 - x)^2}{\sigma_2^2}$$

To find the value of x that minimizes this sum of squares (the "least squares estimate"), what mathematical operation do we perform on $2J(x)$ with respect to x? ______________________

Go to Frame 13.

Frame 13

To find the x that minimizes $2J(x)$, we take the derivative with respect to x and set it to zero:

$$\frac{d}{dx}[2J(x)] = 0$$

$$\frac{d}{dx}\left(\frac{(z_1 - x)^2}{\sigma_1^2} + \frac{(z_2 - x)^2}{\sigma_2^2}\right) = 0$$

$$\frac{-2(z_1 - x)}{\sigma_1^2} + \frac{-2(z_2 - x)}{\sigma_2^2} = 0$$

$$\frac{z_1 - x}{\sigma_1^2} + \frac{z_2 - x}{\sigma_2^2} = 0$$

Solving for x:

$$x\left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right) = \frac{z_1}{\sigma_1^2} + \frac{z_2}{\sigma_2^2}$$

$$x = \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right)^{-1}\left(\frac{z_1}{\sigma_1^2} + \frac{z_2}{\sigma_2^2}\right)$$

This is the same weighted average as found in Frame 11. Let's call this optimal estimate $\hat{x}$ (or z_3 from before).

Go to Frame 14.

Frame 14

Dispersion (Variance) of the Optimally Combined Estimate

The text asks to verify if this method of combining measurements is truly optimal by examining the dispersion (variance) of the combined estimate. Let $z_1 = x + r_1$ and $z_2 = x + r_2$, where r_1, r_2 are independent noises with $\langle r_1 \rangle = \langle r_2 \rangle = 0$, $\langle r_1^2 \rangle = \sigma_1^2$, $\langle r_2^2 \rangle = \sigma_2^2$. We form a linear combination $\hat{x} = \alpha z_1 + \beta z_2$. For $\hat{x}$ to be an unbiased estimate of x (i.e., $\langle \hat{x} \rangle = x$), what must be true about α and β? $\langle \alpha(x + r_1) + \beta(x + r_2) \rangle = \alpha x + \alpha \langle r_1 \rangle + \beta x + \beta \langle r_2 \rangle = (\alpha + \beta)x$. So, for $\langle \hat{x} \rangle = x$, we need $\alpha + \beta = 1$. Let $\beta = 1 - \alpha$. The combined noise is $r = \alpha r_1 + (1 - \alpha) r_2$. The variance of the combined estimate is $\langle r^2 \rangle$. Since r_1, r_2 are uncorrelated ($\langle r_1 r_2 \rangle = 0$):

$$\langle r^2 \rangle = \langle (\alpha r_1 + (1 - \alpha) r_2)^2 \rangle = \alpha^2 \langle r_1^2 \rangle + (1 - \alpha)^2 \langle r_2^2 \rangle + 2\alpha(1 - \alpha)\langle r_1 r_2 \rangle$$

$$\sigma_{\hat{x}}^2 = \alpha^2\sigma_1^2 + (1-\alpha)^2\sigma_2^2$$

To find the α that minimizes this variance, what do we do? ______________________________

Go to Frame 15.

Frame 15

To find the α that minimizes $\sigma_{\hat{x}}^2 = \alpha^2\sigma_1^2 + (1-\alpha)^2\sigma_2^2$, we differentiate with respect to α and set to zero:

$$\frac{d}{d\alpha}\sigma_{\hat{x}}^2 = 2\alpha\sigma_1^2 + 2(1-\alpha)(-1)\sigma_2^2 = 0$$

$$2\alpha\sigma_1^2 - 2\sigma_2^2 + 2\alpha\sigma_2^2 = 0$$

$$\alpha(\sigma_1^2 + \sigma_2^2) = \sigma_2^2$$

$$\alpha = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

And $\beta = 1-\alpha = \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}$. These are exactly the same weights we found from the least squares method and the weighted averaging based on number of measurements, confirming this is an optimal way to combine the estimates to minimize variance.

Go to Frame 16.

Frame 16

Correlation Between Measurements/Estimates

So far, we assumed the noises r_x and r_y (or r_1 and r_2) were uncorrelated. What if they are correlated? Suppose we have two sets of measurements, x and y, with means $\bar{r}_x = \bar{r}_y = 0$ and variances $\sigma_x^2 \neq 0$, $\sigma_y^2 \neq 0$. Now, assume there is a correlation between the noises, meaning $\langle r_x r_y \rangle \neq 0$. We define the **covariance** σ_{xy} as:

$$\sigma_{xy} = \langle (x-\bar{x})(y-\bar{y}) \rangle$$

(If $\bar{x}, \bar{y}$ are the true values, then $x - \bar{x} = r_x$ and $y - \bar{y} = r_y$, so $\sigma_{xy} = \langle r_x r_y \rangle$). The **correlation coefficient** ρ is defined as:

$$\rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

where $|\rho| \leq 1$. A negative ρ means anti-correlation.

The text shows that $\sigma_{xy} = \langle xy \rangle - \langle x \rangle \langle y \rangle$. (If x, y are measurements around a true value, this becomes $\langle (X_{true} + r_x)(X_{true} + r_y) \rangle - X_{true}^2 = \cdots \approx \langle r_x r_y \rangle$ if noise is centered at 0).

Go to Frame 17.

Frame 17

Combining Correlated Measurements/Estimates

Suppose we have two measurements $z_1 = x + w_1$ and $z_2 = x + w_2$, where w_1, w_2 are correlated noises. $\langle w_1^2 \rangle = \sigma_1^2$, $\langle w_2^2 \rangle = \sigma_2^2$, and $\langle w_1 w_2 \rangle = \rho\sigma_1\sigma_2 \neq 0$.

We can try to "decompose" one noise in terms of the other and an uncorrelated part. Let:

$$w_1 = \alpha w_2 + w$$

where w is a new noise component that is *uncorrelated* with w_2 (i.e., $\langle w w_2 \rangle = 0$). From $\langle w_1 w_2 \rangle = \langle (\alpha w_2 + w) w_2 \rangle = \alpha \langle w_2^2 \rangle + \langle w w_2 \rangle = \alpha\sigma_2^2$. Since $\langle w_1 w_2 \rangle = \rho\sigma_1\sigma_2$, we get $\alpha\sigma_2^2 = \rho\sigma_1\sigma_2 \implies \alpha = \rho\frac{\sigma_1}{\sigma_2}$.

Now find the variance of the uncorrelated part, $\sigma_w^2 = \langle w^2 \rangle$: $\sigma_1^2 = \langle w_1^2 \rangle = \langle (\alpha w_2 + w)^2 \rangle = \alpha^2 \langle w_2^2 \rangle + \langle w^2 \rangle + 2\alpha \langle w_2 w \rangle$ $\sigma_1^2 = \alpha^2 \sigma_2^2 + \sigma_w^2 + 0$ $\sigma_w^2 = \sigma_1^2 - \alpha^2 \sigma_2^2 = \sigma_1^2 - \left(\rho \frac{\sigma_1}{\sigma_2}\right)^2 \sigma_2^2 = \sigma_1^2 - \rho^2 \sigma_1^2 = \sigma_1^2 (1 - \rho^2)$. So, $w \sim N(0, \sigma_1 \sqrt{1 - \rho^2})$.

Go to Frame 18.

Frame 18

Now we use the quadratic form (least squares) again, but it's more complex due to correlation. The general form for $2J(x)$ for two correlated measurements involves the inverse of the covariance matrix. The text directly presents the result of minimizing a modified $2J(x)$ that accounts for correlation (using $w_1 = \alpha w_2 + w$ and w_2 is like $z_2 - x$, w is like $(z_1 - x) - \alpha(z_2 - x)$):

$$2J(x) = \left(\frac{w_2}{\sigma_2}\right)^2 + \left(\frac{w}{\sigma_w}\right)^2 = \frac{(z_2 - x)^2}{\sigma_2^2} + \frac{((z_1 - x) - \alpha(z_2 - x))^2}{\sigma_1^2 (1 - \rho^2)}$$

This is effectively transforming to uncorrelated variables. The text simplifies this to a quadratic form:

$$2J(x) = \frac{1}{1 - \rho^2} \left(\frac{(z_1 - x)^2}{\sigma_1^2} + \frac{(z_2 - x)^2}{\sigma_2^2} - \frac{2\rho(z_1 - x)(z_2 - x)}{\sigma_1 \sigma_2} \right)$$

To find the optimal $\hat{x}$ that minimizes this, we set $\frac{d}{dx}[2J(x)] = 0$. This leads to (after some algebra):

$$\hat{x} \left[\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} - \frac{2\rho}{\sigma_1 \sigma_2} \right] = \frac{z_1}{\sigma_1^2} + \frac{z_2}{\sigma_2^2} - \frac{\rho(z_1 + z_2)}{\sigma_1 \sigma_2}$$

What happens in the case of no correlation ($\rho = 0$)? Does this formula simplify to our previous result for uncorrelated measurements? [Yes — No]

Go to Frame 19.

Frame 19

Yes. If $\rho = 0$, the formula from Frame 18 becomes:

$$\hat{x} \left[\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} \right] = \frac{z_1}{\sigma_1^2} + \frac{z_2}{\sigma_2^2}$$

This is identical to the result in Frame 13 for uncorrelated measurements.

Now consider another extreme: perfect correlation, $\rho = 1$. What does the formula for $\hat{x}$ become? (The denominator term $1 - \rho^2$ in $2J(x)$ means we need to be careful if using that directly, but the derivative result is usually better behaved for limits). If $\rho = 1$:

$$\hat{x} \left[\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2} - \frac{2}{\sigma_1 \sigma_2} \right] = \frac{z_1}{\sigma_1^2} + \frac{z_2}{\sigma_2^2} - \frac{z_1 + z_2}{\sigma_1 \sigma_2}$$

The term in the square brackets on the left is $\left(\frac{1}{\sigma_1} - \frac{1}{\sigma_2}\right)^2$. The term on the right is $\frac{z_1\sigma_2 - z_1\sigma_1 + z_2\sigma_1 - z_2\sigma_2}{\sigma_1^2 \sigma_2} = \frac{(z_1 - z_2)(\sigma_2 - \sigma_1)}{\sigma_1^2 \sigma_2}$. So, if $\rho = 1$, it implies $z_1/\sigma_1 = z_2/\sigma_2$ (or $w_1/\sigma_1 = w_2/\sigma_2$, meaning the noises are perfectly proportional). In this case, the text shows $x = (\dots) = z_2$ (assuming $\sigma_1 \neq \sigma_2$; if $\sigma_1 = \sigma_2$ and $\rho = 1$, then $z_1 = z_2$, and any $x = z_1 = z_2$ works). The argument in the text suggests if $\rho = 1$, then $w_1 = (\sigma_1/\sigma_2) w_2$. The second measurement z_2 contains all the information of z_1 up to a scaling factor. Thus, $\hat{x} = z_2$ (or z_1, they are equivalent information-wise).

Go to Frame 20.

Frame 20

Special Case: Equal Dispersions (Variances)

What if the noises have equal variance, $\sigma_1 = \sigma_2 = \sigma$, but are still correlated with coefficient ρ? The general formula for $\hat{x}$ from Frame 18 (the derivative result):

$$\hat{x}\left[\frac{1}{\sigma^2} + \frac{1}{\sigma^2} - \frac{2\rho}{\sigma^2}\right] = \frac{z_1}{\sigma^2} + \frac{z_2}{\sigma^2} - \frac{\rho(z_1 + z_2)}{\sigma^2}$$

Multiply by σ^2:

$$\hat{x}[2 - 2\rho] = z_1 + z_2 - \rho(z_1 + z_2)$$

$$\hat{x}[2(1 - \rho)] = (z_1 + z_2)(1 - \rho)$$

If $\rho \neq 1$:

$$\hat{x} = \frac{(z_1 + z_2)(1 - \rho)}{2(1 - \rho)} = \frac{z_1 + z_2}{2}$$

What does this result mean? ______________________________

Go to Frame 21.

Frame 21

The result $\hat{x} = \frac{z_1+z_2}{2}$ when $\sigma_1 = \sigma_2 = \sigma$ (and $\rho \neq 1$) means that if the individual measurements have the same uncertainty, the optimal combined estimate is simply their **arithmetic mean (average)**, regardless of the correlation (as long as it's not perfect correlation, $\rho = 1$, where the problem becomes ill-defined or implies $z_1 = z_2$).

This is a familiar result: when measurements are of equal quality (same σ), we average them. The correlation affects the variance of this average, but not the way the average itself is formed.

This concludes this rather dense section on optimal filtering and combining measurements. The key takeaways are the methods for combining information based on their uncertainties (variances) and correlations. End of Section.

3 Tracking a Constant Scalar Quantity

Frame 1

Let $\hat{x}$ denote our estimate of a true, constant scalar quantity x. We are interested in how new information, arriving via measurements $z_i = x + r_i$, helps synchronize our model (our estimate $\hat{x}$) with the real system (the true value x).

The properties of our measurement noise r_i are crucial:

1. $r_i \sim N(0, \sigma)$: The noise is Gaussian with zero mean and standard deviation σ (variance σ^2).
2. $\langle r_i \rangle = 0$: The average noise is zero (unbiased measurements).
3. $\langle r_i^2 \rangle = \sigma^2$: The variance of the noise.
4. $\langle r_i r_j \rangle = \delta_{ij}\sigma^2$: The noise is uncorrelated from one measurement to the next. (δ_{ij} is the Kronecker delta, which is 1 if $i = j$ and 0 if $i \neq j$).

Why is the last property, $\langle r_i r_j \rangle = \delta_{ij}\sigma^2$ (uncorrelated noise), important for simplifying our analysis of multiple measurements? ____

Go to Frame 2.

Frame 2

The property that the measurement noise is uncorrelated ($\langle r_i r_j \rangle = 0$ for $i \neq j$) is important because it means the noise at any given moment is completely independent of the noise in previous (or future) moments. This greatly simplifies calculations of variance for sums or averages of measurements, as cross-terms involving products of different noise instances will average to zero (as seen in Frame 10 of the previous section).

Now, let's outline a scheme for iteratively tracking this scalar quantity. Go to Frame 3.

Frame 3

Tracking Scheme

We will develop an iterative scheme to update our estimate of the scalar quantity x.

Step n Suppose we have already processed $n-1$ measurements, or we are at the n-th step of our iteration. This could correspond to a time $(n \cdot T)$ having elapsed, where T is our measurement period. At this point, our best estimate of x is $\hat{x}_n$, and the variance of this estimate is $\hat{\sigma}_n^2$. The text states:

$$\hat{x}_n = \frac{1}{n}\sum_{i=1}^{n} z_i$$

$$\hat{\sigma}_n^2 = \frac{\sigma^2}{n}$$

This is the standard result for the mean of n independent measurements, each with noise variance σ^2.

Go to Frame 4.

Frame 4

Step $n+1$ Now, we acquire one more additional measurement, z_{n+1}. We want to update our estimate to $\hat{x}_{n+1}$ and its variance $\hat{\sigma}_{n+1}^2$.

The updated estimate $\hat{x}_{n+1}$ is the average of all $n+1$ measurements:

$$\hat{x}_{n+1} = \frac{1}{n+1}\sum_{i=1}^{n+1} z_i = \frac{1}{n+1}\left(\sum_{i=1}^{n} z_i + z_{n+1}\right)$$

Since $\sum_{i=1}^{n} z_i = n\hat{x}_n$, we can substitute this:

$$\hat{x}_{n+1} = \frac{1}{n+1}(n\hat{x}_n + z_{n+1})$$

This can be rewritten by adding and subtracting $\frac{1}{n+1}\hat{x}_n$:

$$\hat{x}_{n+1} = \frac{n\hat{x}_n + \hat{x}_n - \hat{x}_n + z_{n+1}}{n+1} = \frac{(n+1)\hat{x}_n + z_{n+1} - \hat{x}_n}{n+1}$$

$$\implies \hat{x}_{n+1} = \hat{x}_n + \frac{1}{n+1}(z_{n+1} - \hat{x}_n)$$

The term $(z_{n+1} - \hat{x}_n)$ is called the **innovation**. It's the difference between the new measurement and our previous best estimate. The factor $\frac{1}{n+1}$ is a **weight** or **gain**.

How does the variance of the estimate $\hat{\sigma}^2_{n+1}$ relate to $\hat{\sigma}^2_n$ and the measurement variance σ^2? The text shows (and we know $\hat{\sigma}^2_{n+1} = \sigma^2/(n+1)$): $\hat{\sigma}^{-2}_{n+1} = (n+1)/\sigma^2 = n/\sigma^2 + 1/\sigma^2 = \hat{\sigma}_n^{-2} + \sigma^{-2}$. This recursive update for the inverse variance is common in filtering.

The update for the estimate can also be written using variances (this form is more general for Kalman filters):

$$\implies \hat{x}_{n+1} = \hat{x}_n + \frac{\hat{\sigma}^2_{n+1}}{\sigma^2}(z_{n+1} - \hat{x}_n)$$

(Since $\hat{\sigma}^2_{n+1}/\sigma^2 = (\sigma^2/(n+1))/\sigma^2 = 1/(n+1)$).

Go to Frame 5.

Frame 5

Convergence Assessment for $\hat{x} \to x$

Let's consider what happens as the sampling time $T \to 0$. In this limit, our discrete variables become continuous functions of time t: $\hat{x}_n \to \hat{x}(t)$ $\hat{\sigma}^2_n \to \hat{\sigma}^2_x(t)$ (variance of the estimate at time t) $z_n \to z(t)$ (the continuous measurement stream)

The update equation from Frame 4 was: $\hat{x}_{n+1} - \hat{x}_n = \frac{1}{n+1}(z_{n+1} - \hat{x}_n)$. Dividing by T (the time step between n and $n+1$):

$$\frac{\hat{x}_{n+1} - \hat{x}_n}{T} = \frac{1}{(n+1)T}(z_{n+1} - \hat{x}_n)$$

As $T \to 0$, the left side becomes the derivative $\dot{\hat{x}}(t)$. The term $nT \approx t$. So $(n+1)T \approx t$. The term $\frac{\hat{\sigma}^2_{n+1}}{\sigma^2 T}$ (using the alternative gain form) is what the text considers. Let $\sigma^2 T$ be $R(t)$, representing the effective measurement noise power over time. (If σ^2 is variance of noise r, and T is small, $R(t)$ is like a rate). The text uses $\frac{\hat{\sigma}^2_{n+1}}{\sigma^2 T}$ which becomes $\frac{\hat{\sigma}^2_x(t)}{R(t)}$ if we define $R(t) = \sigma^2 T$. The equation then becomes:

$$\dot{\hat{x}}(t) = \frac{\hat{\sigma}^2_x(t)}{R(t)}(z(t) - \hat{x}(t))$$

Assuming $\lim_{T\to 0} \frac{1}{(n+1)T}$ or more generally the gain factor $\frac{\hat{\sigma}_x^2(t)}{R(t)}$ has a meaningful limit. The text defines $R(t) = \lim(\sigma^2 T)$ such that $R(t) > 0$. So, the continuous-time update for the estimate is:

$$\dot{\hat{x}}(t) = \frac{\hat{\sigma}_x^2(t)}{R(t)}(z(t) - \hat{x}(t))$$

This differential equation describes how our estimate $\hat{x}(t)$ changes over time to track the measurement $z(t)$.

Go to Frame 6.

Frame 6

Convergence of the Estimate's Variance (Dispersion) in Continuous Time

From Frame 4, we had the update for the inverse variances: $\hat{\sigma}_{n+1}^{-2} = \hat{\sigma}_n^{-2} + \sigma^{-2}$ Subtract $\hat{\sigma}_n^{-2}$ from both sides: $\hat{\sigma}_{n+1}^{-2} - \hat{\sigma}_n^{-2} = \sigma^{-2}$ Divide by T:

$$\frac{\hat{\sigma}_{n+1}^{-2} - \hat{\sigma}_n^{-2}}{T} = \frac{1}{\sigma^2 T}$$

As $T \to 0$, the left side becomes $\frac{d}{dt}(\hat{\sigma}_x^2(t)^{-1})$. The right side becomes $1/R(t)$.

$$\frac{d}{dt}(\hat{\sigma}_x^2(t)^{-1}) = \frac{1}{R(t)}$$

The text actually works with variance $\hat{\sigma}_x^2$ directly: $\hat{\sigma}_{n+1}^2 = \sigma^2/(n+1)$ and $\hat{\sigma}_n^2 = \sigma^2/n$.

$$\frac{1}{T}(\hat{\sigma}_{n+1}^2 - \hat{\sigma}_n^2) = \frac{1}{T}\left(\frac{\sigma^2}{n+1} - \frac{\sigma^2}{n}\right) = \frac{\sigma^2}{T}\frac{n-(n+1)}{n(n+1)} = -\frac{\sigma^2}{Tn(n+1)}$$

As $T \to 0$, $n \to \infty$, and $nT \to t$. So, $Tn(n+1) \approx Tn^2 \approx (nT)^2/T = t^2/T$. The derivative becomes $\dot{\hat{\sigma}}_x^2(t)$. The expression given in the OCR seems to be:

$$\dot{\hat{\sigma}}_x^2 = -\frac{(\hat{\sigma}_x^2)^2}{R}$$

where $R = \sigma^2 T$. (This arises from $\frac{d}{dt}(\sigma^2/t) = -\sigma^2/t^2 = -(\sigma^2/t)^2/\sigma^2 = -(\hat{\sigma}_x^2)^2/(\sigma^2$ if $T = 1$ for R)). Let's use the more standard result that if $\frac{d}{dt}(1/V) = 1/R$, then $V = \frac{1}{\int(1/R)dt + C_0}$. If R is constant, $1/V = t/R + C_0 \implies V = R/(t + C_0 R)$. As $t \to \infty$, $V \to 0$.

So, $\hat{\sigma}_x^2(t) \to 0$ as $t \to \infty$. What does $\hat{\sigma}_x^2(t) \to 0$ mean for our estimate $\hat{x}(t)$? It means $\langle(\hat{x}(t) - x)^2\rangle \to 0$. This indicates that the estimate converges to the true value as we incorporate more information over time. The system achieves perfect synchronization between the model (estimate) and the real system (true value).

Go to Frame 7.

Frame 7

Measuring a Scalar Variable (When Dynamics for $x(t)$ in System S are Unknown)

Previously, we assumed x was constant. What if $x(t)$ can change, but we don't know its dynamics (how it changes on its own)? In this scenario, at any given moment, the measurement $z(t)$ is our only estimate for $x(t)$. It's as if we "forget" all past measurements. So, our estimate is $\hat{x}(t) = Z(t)$, and its variance is $\hat{\sigma}_x^2(t) = R(t)$ (the measurement noise variance).

What if we consider very small time intervals? We can approximate the dynamics as locally linear. Go to Frame 8.

Frame 8

Describing Dynamics in System S (Locally Linear)

If we assume the true value $x(t)$ changes slowly, or we look at small enough time intervals, we can approximate its dynamics with a first-order linear differential equation:

$$\dot{x}(t) = A(t)x(t) + C(t)$$

In discrete time, with time step T: $\dot{x} \approx (x_{n+1} - x_n)/T$. So, $x_{n+1} - x_n \approx T(A(t_n)x_n + C(t_n))$.

$$x_{n+1} \approx (1 + A(t_n)T)x_n + C(t_n)T$$

Let $\phi_n = 1 + A(t_n)T$ and $C_n^* = C(t_n)T$. (The OCR uses C_n for $C(t_n)T$). So, the model for how the true state evolves is:

$$x_{n+1} = \phi_n x_n + C_n^*$$

(This is our *model* of the true system's dynamics. It doesn't include noise yet).

Go to Frame 9.

Frame 9

Optimal Synchronization Procedure (Kalman Filter Foundation)

Suppose at time step n, we have an estimate $\hat{x}_n$ and its variance $\hat{\sigma}_n^2$. For step $n+1$, we first make a **prediction** (a priori estimate) of what x will be, based on our model of its dynamics, *before* we get the new measurement z_{n+1}. Let's call this prediction $\bar{x}_{n+1}$. Using the dynamics model from Frame 8:

$$\bar{x}_{n+1} = \phi_n \hat{x}_n + C_n^*$$

This is our best guess for x_{n+1} based on the old estimate $\hat{x}_n$ and the system dynamics.

What is the variance of this prediction, $\bar{\sigma}_{n+1}^2$? (The OCR uses $\bar{\sigma}_n^2$ for the predicted variance at $n+1$ based on state n). The definition given is $\bar{\sigma}_{n+1}^2 = \langle(\bar{x}_{n+1} - x_{n+1})^2\rangle$. Substituting: $\bar{x}_{n+1} - x_{n+1} = (\phi_n \hat{x}_n + C_n^*) - (\phi_n x_n + C_n^*) = \phi_n(\hat{x}_n - x_n)$. (This assumes no process noise for now, which is added in Frame 11).

$$\bar{\sigma}_{n+1}^2 = \langle(\phi_n(\hat{x}_n - x_n))^2\rangle = \phi_n^2\langle(\hat{x}_n - x_n)^2\rangle = \phi_n^2\hat{\sigma}_n^2$$

So, the predicted variance (a priori variance for step $n+1$) is $\bar{\sigma}_{n+1}^2 = \phi_n^2\hat{\sigma}_n^2$.

This is the **prediction step**. Go to Frame 10.

Frame 10

Optimal Synchronization Procedure (Kalman Filter - Update Step)

Now, at time $n+1$, we receive a new measurement z_{n+1} with measurement noise variance σ^2. We have:

- Our prediction (a priori estimate): $\bar{x}_{n+1}$ with variance $\bar{\sigma}_{n+1}^2$.
- Our new measurement: z_{n+1} with variance σ^2.

How do we optimally combine these two pieces of information to get our updated (a posteriori) estimate $\hat{x}_{n+1}$? This is exactly the problem of combining two estimates we solved in Frame 11 of the previous section (and Frame 13 of this document for uncorrelated noise). The updated estimate $\hat{x}_{n+1}$ is a weighted average of the prediction $\bar{x}_{n+1}$ and the measurement z_{n+1}:

$$\hat{x}_{n+1} = \bar{x}_{n+1} + K_{n+1}(z_{n+1} - \bar{x}_{n+1})$$

where K_{n+1} is the Kalman Gain. Using the formula for optimal weights (where $\bar{\sigma}^2_{n+1}$ acts like σ^2_{old} and σ^2 acts like $\sigma^2_{\text{new}_\text{meas}}$):

$$K_{n+1} = \frac{\bar{\sigma}^2_{n+1}}{\bar{\sigma}^2_{n+1} + \sigma^2}$$

The variance of this updated estimate $\hat{\sigma}^2_{n+1}$ is (from $1/\sigma_3^2 = 1/\sigma_1^2 + 1/\sigma_2^2$): $\hat{\sigma}^{-2}_{n+1} = \bar{\sigma}^{-2}_{n+1} + \sigma^{-2}$. Alternatively, it can be shown that $\hat{\sigma}^2_{n+1} = (1 - K_{n+1})\bar{\sigma}^2_{n+1} = \bar{\sigma}^2_{n+1} - \frac{(\bar{\sigma}^2_{n+1})^2}{\bar{\sigma}^2_{n+1}+\sigma^2}$. The text uses M_{n+1} for $\bar{\sigma}^2_{n+1}$ (predicted/a priori variance) and P_{n+1} for $\hat{\sigma}^2_{n+1}$ (updated/a posteriori variance). So, $K_{n+1} = M_{n+1}/(M_{n+1} + \sigma^2)$. And $P_{n+1} = M_{n+1} - K_{n+1}M_{n+1} = M_{n+1} - \frac{M^2_{n+1}}{M_{n+1}+\sigma^2}$.

This is the **update step**. The combination of prediction and update is the core of the Kalman filter. Go to Frame 11.

Frame 11

Dynamic Noise (Process Noise)

So far, our model for how the true state x_n evolves, $x_{n+1} = \phi_n x_n + C_n^*$, was assumed to be perfect. In reality, the system dynamics themselves might be noisy or not perfectly known. This is called **dynamic noise** or **process noise**, w_n. The state evolution equation becomes:

$$x_{n+1} = \phi_n x_n + C_n^* + \Gamma_n w_n$$

where w_n is typically zero-mean Gaussian noise, $\langle w_n \rangle = 0$, $\langle w_n w_m \rangle = \delta_{nm} Q_n$. Q_n is the variance of the process noise at step n. Γ_n is a scaling factor.

How does this process noise Q_n affect the variance of our prediction $\bar{\sigma}^2_{n+1}$ (which the text calls M_{n+1})? Previously (Frame 9), without process noise, $M_{n+1} = \phi_n^2 P_n$ (where $P_n = \hat{\sigma}_n^2$). With process noise, the uncertainty in our prediction increases. The new term added to the variance is due to w_n:

$$M_{n+1} = \phi_n^2 P_n + \Gamma_n^2 Q_n$$

The rest of the update step (for $\hat{x}_{n+1}$ and P_{n+1}) remains the same as in Frame 10, but uses this new, larger M_{n+1}. The presence of unknown dynamics (process noise) means our covariance (variance) can increase during the prediction step.

Go to Frame 12.

Frame 12

Transition to Continuous Picture (Riccati Equation)

Let's look at the continuous time limit for these recursive equations, especially for the variance. The text defines: $A(t) = (\phi_n - 1)/T$ $C(t) = C_n^*/T$ $R(t)$ is related to measurement noise σ^2 (e.g. $R = \sigma^2 T$) $Q(t)$ is related to process noise Q_n (e.g. $Q = Q_n/T$ if $\Gamma_n = 1$) $P(t)$ is the continuous version of $P_n = \hat{\sigma}_n^2$.

The update for the estimate $\hat{x}_{n+1} = \bar{x}_{n+1} + K_{n+1}(z_{n+1} - \bar{x}_{n+1})$ becomes a differential equation for $\hat{x}(t)$:

$$\lim_{T \to 0} \frac{\hat{x}_{n+1} - \hat{x}_n}{T} = \dot{\hat{x}}(t) = A(t)\hat{x}(t) + C(t) + \frac{P(t)}{R(t)}(Z(t) - \hat{x}(t))$$

(This assumes the measurement $Z(t)$ is directly of $x(t)$, i.e., $H = 1$ in standard Kalman terms. The term $Z(t) - \hat{x}(t)$ corresponds to the innovation).

The prediction variance was $M_{n+1} = \phi_n^2 P_n + \Gamma_n^2 Q_n$. The updated variance was $P_{n+1} = M_{n+1} - \frac{M_{n+1}^2}{M_{n+1}+\sigma^2}$. Taking the limit as $T \to 0$, the update for $P(t)$ (the variance of the estimate) becomes a differential equation known as the Riccati equation:

$$\dot{P}(t) = 2A(t)P(t) + \Gamma^2(t)Q(t) - \frac{P(t)^2}{R(t)}$$

The first term ($2AP$) shows how variance changes due to system dynamics (can grow or shrink). The second term ($\Gamma^2 Q$) shows how variance increases due to process noise. The third term ($-P^2/R$) shows how variance decreases due to new measurements.

This equation describes how the uncertainty (variance $P(t)$) of our estimate evolves over time. This is the heart of the continuous Kalman filter.

End of Section.

4 Tracking Vector Variables (Kalman Filter Introduction)

Frame 1

(Note: The original text mentions some initial parts might be less clear or missing due to absence from lectures. We will proceed with the provided material.)

Let's extend our tracking problem to **vector variables**. Imagine we are measuring the position $\boldsymbol{x}$ after a known period, and in our model system, we get a measurement $\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{r}$, where $\boldsymbol{r}$ is measurement noise. This measurement helps "sharpen" our estimate of a state vector, which might include multiple related quantities. For this example, the state vector we are interested in is:

$$\begin{bmatrix} \hat{x}_{\text{pos}} \\ \hat{v}_{\text{vel}} \end{bmatrix}$$

(an estimate of position and velocity).

The estimate is also sharpened because the components of the true state (e.g., true position x and true velocity v) might be correlated with the noise components in the measurement of x (denoted m_x) and the noise in the measurement of v (denoted m_v). The text introduces noisy "observations" of the true state components: $\bar{x} = x + m_x$ (observed position = true position + pos. noise) $\bar{v} = v + m_v$ (observed velocity = true velocity + vel. noise) And it's given that $\langle m_x m_v \rangle \neq 0$, meaning these observation noises are correlated.

Go to Frame 2.

Frame 2

Sharpening the Estimate (Scalar Components First)

Let's first consider updating the estimate for position, $\hat{x}_{\text{pos}}$, and then for velocity, $\hat{v}_{\text{vel}}$. The proposed update for the position estimate $\hat{x}$ (using $\bar{x}$ and $\bar{v}$ as noisy inputs, and z as a direct noisy measurement of x) is a linear combination:

$$\hat{x} = x + p_x = a_{xx}\bar{x} + a_{xv}\bar{v} + b_x z$$

And for the velocity estimate $\hat{v}$:

$$\hat{v} = v + p_v = a_{vx}\bar{x} + a_{vv}\bar{v} + b_v z$$

Here, p_x and p_v are the errors in our estimates. Our goal is to choose coefficients a_{ij} and b_i to minimize these errors (specifically, their variances).

Let's substitute $\bar{x} = x + m_x$, $\bar{v} = v + m_v$, and $z = x + r$ into the equation for $\hat{x}$:

$$x + p_x = a_{xx}(x + m_x) + a_{xv}(v + m_v) + b_x(x + r)$$

$$x + p_x = (a_{xx} + b_x)x + a_{xv}v + a_{xx}m_x + a_{xv}m_v + b_x r$$

For $\hat{x}$ to be an unbiased estimate of x (meaning $\langle p_x \rangle = 0$ and the coefficient of x on the RHS is 1, and coefficient of v is 0, assuming $\langle m_x \rangle = \langle m_v \rangle = \langle r \rangle = 0$):

1. Coefficient of v: $a_{xv} = 0$
2. Coefficient of x: $a_{xx} + b_x = 1 \implies b_x = 1 - a_{xx}$

So, the error p_x becomes:

$$p_x = a_{xx}m_x + (1 - a_{xx})r \quad (\text{since } a_{xv} = 0)$$

We expect $\langle p_x \rangle = a_{xx}\langle m_x \rangle + (1 - a_{xx})\langle r \rangle = 0$.

What quantity do we want to minimize to find the optimal a_{xx}? ____________________

Go to Frame 3.

Frame 3

To find the optimal a_{xx}, we want to minimize the variance of the estimation error p_x, which is $\langle p_x^2 \rangle$. Given $p_x = a_{xx}m_x + (1 - a_{xx})r$, and assuming m_x and r are uncorrelated (i.e., $\langle m_x r \rangle = 0$):

$$\langle p_x^2 \rangle = \langle (a_{xx}m_x + (1 - a_{xx})r)^2 \rangle$$

$$= a_{xx}^2 \langle m_x^2 \rangle + (1 - a_{xx})^2 \langle r^2 \rangle + 2a_{xx}(1 - a_{xx})\langle m_x r \rangle$$

Let $\langle m_x^2 \rangle = \sigma_x^2$ (variance of observation noise for x) and $\langle r^2 \rangle = \sigma_r^2$ (variance of direct measurement noise for x).

$$\langle p_x^2 \rangle = a_{xx}^2 \sigma_x^2 + (1 - a_{xx})^2 \sigma_r^2$$

To minimize this with respect to a_{xx}, we set $\frac{d}{da_{xx}} \langle p_x^2 \rangle = 0$:

$$2a_{xx}\sigma_x^2 + 2(1 - a_{xx})(-1)\sigma_r^2 = 0$$

$$a_{xx}\sigma_x^2 - \sigma_r^2 + a_{xx}\sigma_r^2 = 0$$

$$a_{xx}(\sigma_x^2 + \sigma_r^2) = \sigma_r^2 \implies a_{xx} = \frac{\sigma_r^2}{\sigma_x^2 + \sigma_r^2}$$

And $b_x = 1 - a_{xx} = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_r^2}$.

These are familiar weights, similar to combining two independent estimates. The text performs a similar minimization for the velocity estimate $\hat{v}$, assuming $a_{vx} + b_v = 0$ and $a_{vv} = 1$ for unbiasedness (this seems to imply v is estimated purely from $\bar{v}$ and influenced by x and z only to cancel out their noise effects, or that $\bar{v}$ is the primary source of v information). The given result after minimization for a_{vx} (assuming $p_v = m_v + a_{vx}(m_x - r)$ based on $a_{vx} + b_v = 0, a_{vv} = 1$ implies certain structure) is:

$$a_{vx} = \frac{\langle m_x m_v \rangle}{\sigma_x^2 + \sigma_r^2}$$

(The derivation for a_{vx} is a bit condensed in the original, but it would involve minimizing $\langle p_v^2 \rangle$ where p_v includes terms with m_v, m_x, and r via a_{vx}, a_{vv}, b_v, and using the unbiasedness conditions.)

The sharpened estimates are then written in a "prediction + correction" form:

$$\hat{x} = \bar{x} + \frac{\sigma_x^2}{\sigma_x^2 + \sigma_r^2}(z - \bar{x})$$

$$\hat{v} = \bar{v} + \frac{\langle m_x m_v \rangle}{\sigma_x^2 + \sigma_r^2}(z - \bar{x})$$

Notice the correction term $(z - \bar{x})$ is the difference between the direct measurement z and the noisy observation $\bar{x}$.

Go to Frame 4.

Frame 4

Covariance Matrix

When dealing with multiple variables (vectors), their uncertainties and interdependencies are described by a covariance matrix. Let $\boldsymbol{P}$ be the covariance matrix of the estimate errors and $\boldsymbol{M}$ be the covariance matrix of the prediction errors (or observation noises like m_x, m_v). For a 2D state vector like $\begin{bmatrix} x \\ y \end{bmatrix}$: The elements are $M_{ij} = \langle (\bar{x}_i - x_i)(\bar{x}_j - x_j) \rangle$, where x_i, x_j are true values. E.g., for observations $\bar{x}, \bar{y}$ with noises m_x, m_y:

$$\boldsymbol{M} = \begin{bmatrix} \langle m_x^2 \rangle & \langle m_x m_y \rangle \\ \langle m_y m_x \rangle & \langle m_y^2 \rangle \end{bmatrix} = \begin{bmatrix} \sigma_x^2 & \rho \sigma_x \sigma_y \\ \rho \sigma_x \sigma_y & \sigma_y^2 \end{bmatrix}$$

where ρ is the correlation coefficient between m_x and m_y.

The determinant of this matrix $\boldsymbol{M}$ is $\det(\boldsymbol{M}) = \sigma_x^2 \sigma_y^2 - (\rho \sigma_x \sigma_y)^2 = \sigma_x^2 \sigma_y^2 (1 - \rho^2)$. The inverse $\boldsymbol{M}^{-1}$ is:

$$\boldsymbol{M}^{-1} = \frac{1}{\det(\boldsymbol{M})} \begin{bmatrix} \sigma_y^2 & -\rho \sigma_x \sigma_y \\ -\rho \sigma_x \sigma_y & \sigma_x^2 \end{bmatrix}$$

The multivariate normal (Gaussian) probability distribution for a vector $\boldsymbol{x}$ with mean $\bar{\boldsymbol{x}}$ and covariance matrix $\boldsymbol{M}$ is:

$$p(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^n \det(\boldsymbol{M})}} e^{-\frac{1}{2}(\boldsymbol{x} - \bar{\boldsymbol{x}})^T \boldsymbol{M}^{-1} (\boldsymbol{x} - \bar{\boldsymbol{x}})}$$

where n is the dimension of the vector. The term in the exponent is a quadratic form.

Go to Frame 5.

Frame 5

Sensor for Multiple Variables

Suppose we have s sensors and an n-dimensional state vector $\boldsymbol{x}$. The measurement $\boldsymbol{z}$ (an s-dimensional vector) can be related to the state vector by a sensor matrix $\boldsymbol{H}$ (size $s \times n$):

$$\boldsymbol{z} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{r}$$

where $\boldsymbol{r}$ is an s-dimensional measurement noise vector. The covariance matrix of this sensor noise is $\langle \boldsymbol{r}\boldsymbol{r}^T \rangle = \boldsymbol{R}$ (an $s \times s$ matrix).

Go to Frame 6.

Frame 6

Sharpening with a Vector Measurement (Kalman Filter Update)

Let's formulate the Kalman filter update for vector states. We have:

- A prediction (a priori estimate) of the state: $\bar{\boldsymbol{x}}$
- The covariance of this prediction error: $\boldsymbol{M} = \langle (\bar{\boldsymbol{x}} - \boldsymbol{x})(\bar{\boldsymbol{x}} - \boldsymbol{x})^T \rangle$
- A new vector measurement: $\boldsymbol{z}$
- The measurement model: $\boldsymbol{z} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{r}$
- The measurement noise covariance: $\boldsymbol{R} = \langle \boldsymbol{r}\boldsymbol{r}^T \rangle$

We want to find the updated (a posteriori) estimate $\hat{x}$ as a linear combination of the prediction $\bar{\boldsymbol{x}}$ and the measurement $\boldsymbol{z}$. The text represents this as: $\hat{x} = \boldsymbol{A}\bar{\boldsymbol{x}} + \boldsymbol{B}\boldsymbol{z} = \boldsymbol{x} + \boldsymbol{p}$ where $\boldsymbol{p}$ is the error in the final estimate. We want to minimize $\langle \boldsymbol{p}\boldsymbol{p}^T \rangle$, which is the trace of the posterior covariance matrix $\boldsymbol{P}$.

Substituting $\bar{\boldsymbol{x}} = \boldsymbol{x} + \boldsymbol{m}$ (where $\boldsymbol{m}$ is the prediction error with covariance $\boldsymbol{M}$) and $\boldsymbol{z} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{r}$:

$$\boldsymbol{x} + \boldsymbol{p} = \boldsymbol{A}(\boldsymbol{x} + \boldsymbol{m}) + \boldsymbol{B}(\boldsymbol{H}\boldsymbol{x} + \boldsymbol{r})$$

$$\boldsymbol{x} + \boldsymbol{p} = (\boldsymbol{A} + \boldsymbol{B}\boldsymbol{H})\boldsymbol{x} + \boldsymbol{A}\boldsymbol{m} + \boldsymbol{B}\boldsymbol{r}$$

For an unbiased estimate ($\langle \boldsymbol{p} \rangle = \boldsymbol{0}$ and coefficient of $\boldsymbol{x}$ is $\boldsymbol{I}$, the identity matrix):

$$\boldsymbol{A} + \boldsymbol{B}\boldsymbol{H} = \boldsymbol{I} \implies \boldsymbol{A} = \boldsymbol{I} - \boldsymbol{B}\boldsymbol{H}$$

The error in the estimate is:

$$\boldsymbol{p} = \boldsymbol{A}\boldsymbol{m} + \boldsymbol{B}\boldsymbol{r} = (\boldsymbol{I} - \boldsymbol{B}\boldsymbol{H})\boldsymbol{m} + \boldsymbol{B}\boldsymbol{r}$$

The posterior covariance matrix $\boldsymbol{P} = \langle \boldsymbol{p}\boldsymbol{p}^T \rangle$. Assuming $\boldsymbol{m}$ and $\boldsymbol{r}$ are uncorrelated:

$$\boldsymbol{P} = (\boldsymbol{I} - \boldsymbol{B}\boldsymbol{H})\boldsymbol{M}(\boldsymbol{I} - \boldsymbol{B}\boldsymbol{H})^T + \boldsymbol{B}\boldsymbol{R}\boldsymbol{B}^T$$

The method of Lagrangian multipliers is used to find $\boldsymbol{A}$ and $\boldsymbol{B}$ (or just $\boldsymbol{B}$, since $\boldsymbol{A}$ depends on it) that minimize $\mathrm{tr}(\boldsymbol{P})$ subject to the constraint $\boldsymbol{A} + \boldsymbol{B}\boldsymbol{H} = \boldsymbol{I}$.

The text jumps to a key result (often derived by setting $\frac{\partial \mathrm{tr}(\boldsymbol{P})}{\partial \boldsymbol{B}} = \boldsymbol{0}$): The "Kalman Gain" matrix $\boldsymbol{K}$ (which is our $\boldsymbol{B}$ here) is found to be: $\boldsymbol{K} = \boldsymbol{M}\boldsymbol{H}^T(\boldsymbol{H}\boldsymbol{M}\boldsymbol{H}^T + \boldsymbol{R})^{-1}$. And the updated estimate is:

$$\hat{x} = \bar{\boldsymbol{x}} + \boldsymbol{K}(\boldsymbol{z} - \boldsymbol{H}\bar{\boldsymbol{x}})$$

The term $(\boldsymbol{z} - \boldsymbol{H}\bar{\boldsymbol{x}})$ is the vector **innovation**. The updated covariance matrix $\boldsymbol{P}$ is:

$$\boldsymbol{P} = (\boldsymbol{I} - \boldsymbol{K}\boldsymbol{H})\boldsymbol{M}$$

The text provides an equivalent form, sometimes called the Joseph form, which is more robust: $\boldsymbol{P} = (\boldsymbol{I} - \boldsymbol{K}\boldsymbol{H})\boldsymbol{M}(\boldsymbol{I} - \boldsymbol{K}\boldsymbol{H})^T + \boldsymbol{K}\boldsymbol{R}\boldsymbol{K}^T$. The text also gives $\boldsymbol{P}^{-1} = \boldsymbol{M}^{-1} + \boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H}$ (Information Filter form). This shows how information (inverse covariance) from the prediction and measurement are added.

The equation $\hat{x} = \bar{\boldsymbol{x}} + \boldsymbol{P}\boldsymbol{H}^T\boldsymbol{R}^{-1}(\boldsymbol{z} - \boldsymbol{H}\bar{\boldsymbol{x}})$ is also given, which is another way to write the update using $\boldsymbol{K} = \boldsymbol{P}\boldsymbol{H}^T\boldsymbol{R}^{-1}$ (this is true if $\boldsymbol{P}$ is the *posterior* covariance used to calculate gain for the *next* iteration's prediction, or if it's specifically the form of gain $\boldsymbol{K} = \boldsymbol{M}\boldsymbol{H}^T S^{-1}$ where $S = \boldsymbol{H}\boldsymbol{M}\boldsymbol{H}^T + \boldsymbol{R}$).

Go to Frame 7.

Frame 7

Kalman Optimal Filter (Dynamics and Dynamic Noise)

Now we reintroduce dynamics and dynamic (process) noise. **Discrete Case** We are looking at the step from time $(n \cdot T)$ to $((n+1) \cdot T)$.

1. **State Prediction (Time Update):** The true state evolves: $\boldsymbol{x}_{n+1} = \boldsymbol{\phi}_n \boldsymbol{x}_n + \boldsymbol{C}_n + \boldsymbol{\Gamma}_n \boldsymbol{w}_n$ (where $\boldsymbol{w}_n$ is process noise with covariance $\boldsymbol{Q}_n$). Our predicted state (a priori estimate for $n+1$):

$$\bar{\boldsymbol{x}}_{n+1} = \boldsymbol{\phi}_n \hat{x}_n + \boldsymbol{C}_n$$

The covariance of this prediction error (a priori covariance for $n+1$), denoted $\boldsymbol{M}_{n+1}$:

$$\boldsymbol{M}_{n+1} = \langle (\bar{\boldsymbol{x}}_{n+1} - \boldsymbol{x}_{n+1})(\bar{\boldsymbol{x}}_{n+1} - \boldsymbol{x}_{n+1})^T \rangle$$

Substituting the equations: $\bar{\boldsymbol{x}}_{n+1} - \boldsymbol{x}_{n+1} = \boldsymbol{\phi}_n \hat{x}_n + \boldsymbol{C}_n - (\boldsymbol{\phi}_n \boldsymbol{x}_n + \boldsymbol{C}_n + \boldsymbol{\Gamma}_n \boldsymbol{w}_n) = \boldsymbol{\phi}_n(\hat{x}_n - \boldsymbol{x}_n) - \boldsymbol{\Gamma}_n \boldsymbol{w}_n$. Assuming error $(\hat{x}_n - \boldsymbol{x}_n)$ is uncorrelated with process noise $\boldsymbol{w}_n$:

$$\boldsymbol{M}_{n+1} = \boldsymbol{\phi}_n \boldsymbol{P}_n \boldsymbol{\phi}_n^T + \boldsymbol{\Gamma}_n \boldsymbol{Q}_n \boldsymbol{\Gamma}_n^T$$

(where $\boldsymbol{P}_n$ is the posterior covariance from step n).

2. **Measurement Update (Correction):** Kalman Gain: $\boldsymbol{K}_{n+1} = \boldsymbol{M}_{n+1} \boldsymbol{H}^T (\boldsymbol{H} \boldsymbol{M}_{n+1} \boldsymbol{H}^T + \boldsymbol{R})^{-1}$ Updated state estimate (a posteriori):

$$\hat{x}_{n+1} = \bar{\boldsymbol{x}}_{n+1} + \boldsymbol{K}_{n+1}(\boldsymbol{z}_{n+1} - \boldsymbol{H} \bar{\boldsymbol{x}}_{n+1})$$

Updated error covariance (a posteriori):

$$\boldsymbol{P}_{n+1} = (\boldsymbol{I} - \boldsymbol{K}_{n+1} \boldsymbol{H}) \boldsymbol{M}_{n+1}$$

These are the discrete Kalman filter equations.

Go to Frame 8.

Frame 8

Transition to Continuous Time

For the continuous case, we have dynamic noise $w(t)$ and measurement noise $r(t)$, assumed uncorrelated. The state estimate update becomes (as seen in the previous chapter for scalars):

$$\dot{\hat{x}}(t) = \boldsymbol{A}(t)\hat{x}(t) + \boldsymbol{C}(t) + \boldsymbol{P}(t)\boldsymbol{H}^T \boldsymbol{R}^{-1}(t)(\boldsymbol{Z}(t) - \boldsymbol{H}\hat{x}(t))$$

(Here $\boldsymbol{A}(t)$ is the system dynamics matrix, $\boldsymbol{C}(t)$ is a control input vector, $\boldsymbol{P}(t)$ is the estimate error covariance, $\boldsymbol{H}$ is the measurement matrix, $\boldsymbol{R}(t)$ is the measurement noise covariance, $\boldsymbol{Z}(t)$ is the measurement).

The covariance matrix $\boldsymbol{P}(t)$ evolves according to the **Matrix Riccati Differential Equation**:

$$\dot{\boldsymbol{P}}(t) = \boldsymbol{A}(t)\boldsymbol{P}(t) + \boldsymbol{P}(t)\boldsymbol{A}^T(t) + \boldsymbol{\Gamma}(t)\boldsymbol{Q}(t)\boldsymbol{\Gamma}^T(t) - \boldsymbol{P}(t)\boldsymbol{H}^T \boldsymbol{R}^{-1}(t)\boldsymbol{H}\boldsymbol{P}(t)$$

The terms represent:

- $\boldsymbol{AP} + \boldsymbol{PA}^T$: How covariance evolves due to system dynamics.
- $\boldsymbol{\Gamma Q \Gamma}^T$: Increase in covariance due to process noise.
- $-\boldsymbol{PH}^T \boldsymbol{R}^{-1} \boldsymbol{HP}$: Decrease in covariance due to measurements.

Go to Frame 9.

Frame 9

Example: Brownian Motion of a Colloidal Particle

Consider 1D motion. The system dynamics are governed by Stokes' law for drag and random forces from molecular collisions (Newton's second law):

$$m\ddot{x} = -6\pi\eta r_p \dot{x} + F_x(t)$$

(m=mass, x=position, η=viscosity, r_p=particle radius, $F_x(t)$=random force). Let $\tau = m/(6\pi\eta r_p)$ be a characteristic time. State vector $\boldsymbol{x}_{\text{state}} = \begin{bmatrix} x \\ v \end{bmatrix}$ (where $v = \dot{x}$). The system equations are: $\dot{x} = v$ $\dot{v} = -\frac{1}{\tau}v + \frac{F_x(t)}{m}$ The random force term $F_x(t)/m$ is the process noise $w(t)$. Its "spectral density" is Q. $\langle \frac{F_x(t)}{m}\frac{F_x(t')}{m} \rangle = Q\delta(t-t')$. The system matrix $\boldsymbol{A} = \begin{bmatrix} 0 & 1 \\ 0 & -1/\tau \end{bmatrix}$, $\boldsymbol{\Gamma} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

The Riccati equation for $\boldsymbol{P} = \begin{bmatrix} p_{xx} & p_{xv} \\ p_{vx} & p_{vv} \end{bmatrix}$ becomes a set of coupled differential equations. The text focuses on the steady-state solution ($\dot{\boldsymbol{P}} = \boldsymbol{0}$). For velocity variance $p_{vv} = \langle v^2 \rangle$: $\frac{d}{dt}\langle v^2 \rangle = -\frac{2}{\tau}\langle v^2 \rangle + Q = 0 \implies \langle v^2 \rangle_\infty = \frac{Q\tau}{2}$. From thermodynamics (equipartition theorem), $\frac{1}{2}m\langle v^2 \rangle_\infty = \frac{1}{2}k_B T_{\text{abs}}$ (where k_B is Boltzmann's constant, T_{abs} is absolute temperature). So $\langle v^2 \rangle_\infty = k_B T_{\text{abs}}/m$. This implies $Q = \frac{2k_B T_{\text{abs}}}{m\tau}$. This relates the process noise strength Q to physical parameters.

Go to Frame 10.

Frame 10

Continuing with the Brownian motion example: For the covariance term $p_{xv} = \langle xv \rangle$: $\frac{d}{dt}\langle xv \rangle = \langle v^2 \rangle - \frac{1}{\tau}\langle xv \rangle$. In steady state, $\langle xv \rangle_\infty = \tau \langle v^2 \rangle_\infty = \tau\frac{k_B T_{\text{abs}}}{m}$.

For the position variance $p_{xx} = \langle x^2 \rangle$: $\frac{d}{dt}\langle x^2 \rangle = 2\langle xv \rangle$. In steady state, this would imply $\langle xv \rangle_\infty = 0$, which contradicts the above unless $\tau = 0$. This indicates that $\langle x^2 \rangle$ does *not* reach a steady state but grows with time (diffusion). Integrating $\frac{d}{dt}\langle x^2 \rangle = 2\langle xv \rangle_\infty$ gives: $\langle x^2(t) \rangle - \langle x^2(0) \rangle = 2\langle xv \rangle_\infty t = 2\frac{\tau k_B T_{\text{abs}}}{m}t$. Let $D = \frac{\tau k_B T_{\text{abs}}}{m}$ be the diffusion constant. Then $\langle x^2(t) \rangle - \langle x_0^2 \rangle = 2Dt$. This is Einstein's relation for Brownian motion.

Now, if we add position measurements ($z = Hx + r$, with $H = [1\ 0]$), the Riccati equation terms involving R (measurement noise variance) will prevent P_{xx} from growing indefinitely, and a steady state for all components of P can be found.

Go to Frame 11.

Frame 11

Example: RC Circuit Voltage Measurement

Consider an RC circuit. The voltage across the capacitor u_C (denoted u here) follows: $\dot{u} = -\frac{1}{\tau}u + W(t)$, where $\tau = RC$, and $W(t)$ is dynamic noise (e.g., Johnson noise in the resistor) with $\langle W(t)W(t') \rangle = Q\delta(t-t')$. Here, $A = -1/\tau$, $\Gamma = 1$. We measure the capacitor voltage $z = u + r$, so $H = 1$, and measurement noise has variance R. The steady-state Riccati equation for the variance $P = \langle (\hat{u} - u)^2 \rangle$ is: $\dot{P} = 2AP + \Gamma^2 Q - P^2 H^T R^{-1} HP = 0$

$$-\frac{2}{\tau}P + Q - \frac{P^2}{R} = 0$$

This is a quadratic equation for P_∞: $P^2 + \frac{2R}{\tau}P - QR = 0$. Solving for P_∞ (taking the positive root for variance):

$$P_\infty = \frac{-\frac{2R}{\tau} \pm \sqrt{(\frac{2R}{\tau})^2 - 4(1)(-QR)}}{2} = -\frac{R}{\tau} + \sqrt{\left(\frac{R}{\tau}\right)^2 + QR}$$

The text denotes $\alpha = \frac{R}{\tau}\sqrt{1 + \frac{QR\tau^2}{R^2}} = \frac{R}{\tau}\sqrt{1 + \frac{Q\tau^2}{R}}$. So $P_\infty = -\frac{R}{\tau} + \alpha$.

The full time-dependent solution $P(t)$ is also given, showing how the variance of the estimate converges from an initial P_0 to P_∞.

What does P_∞ represent in the context of tracking the voltage u? ________________________
Go to Frame 12.

Frame 12

P_∞ represents the **steady-state (minimum achievable) variance of the error in our estimate of the voltage** u. After the filter has run for a long time, this is the best precision we can expect for our voltage estimate, given the system dynamics, process noise Q, and measurement noise R.

The text defines an effective time constant for the filter's estimation error convergence: $\tau_{\text{eff}}^{-1} = \tau^{-1} + \frac{Q\tau}{2R}$ (This is an approximation for small Q, derived from a specific interpretation of α). Special cases:

- If $Q = 0$ (no process noise, u is truly constant after transients): $P_\infty \approx -\frac{R}{\tau} + \frac{R}{\tau}(1 + \frac{Q\tau^2}{2R}) = \frac{Q\tau}{2}$ (if expanded for small Q). If $Q = 0$ exactly, $P_\infty = 0$. $\tau_{\text{eff}} = \tau$.
- If $Q \to \infty$ (process noise dominates, true u very erratic): P_∞ becomes large. $\tau_{\text{eff}} \to 0$, meaning the filter tries to react very quickly, but its error variance will be high.

This concludes the overview of vector tracking and the Kalman filter concepts. End of Section.

5 Sensors: Simplified Kalman Schemes and Feedback Loops

Frame 1

We have previously discussed the Kalman filter. Let's look at a schematic representation of its core update equation:

$$\dot{\hat{x}} = \boldsymbol{A}\hat{x} + \boldsymbol{C}(t) + \boldsymbol{K}(t)[z - \boldsymbol{H}\hat{x}]$$

(A diagram like the one in the OCR page 1, top, would be very helpful here, showing the feedback loop where the innovation $z - \boldsymbol{H}\hat{x}$ is multiplied by the gain $\boldsymbol{K}(t)$ and added to the dynamics part $\boldsymbol{A}\hat{x} + \boldsymbol{C}(t)$ to drive the integrator producing $\hat{x}$.)

The last term, $\boldsymbol{K}(t)[z - \boldsymbol{H}\hat{x}]$, measures the degree of synchronization between the system S (represented by measurement z) and the model M (represented by estimate $\hat{x}$). When S and M are well-aligned, the innovation $z - \boldsymbol{H}\hat{x}$ is small. The gain $\boldsymbol{K}(t)$ can be anything, and the text implies that if it converges to a steady-state value $\boldsymbol{K}_\infty$, it will be "good enough."

What is the role of the term $\boldsymbol{H}\hat{x}$ in the innovation? [a] It's the direct measurement from the sensor. [b] It's the predicted measurement based on our current estimate of the state. [c] It's the process noise.

Go to Frame 2.

Frame 2

Your answer was [a — b — c].

The correct answer is [b]. $\boldsymbol{H}\hat{x}$ represents what we *expect* the measurement to be, given our current estimate of the state $\hat{x}$ and our knowledge of how the state maps to measurements (the $\boldsymbol{H}$ matrix). The innovation is the difference between the actual measurement z and this predicted measurement.

Now let's consider the sensor itself as a universal measurement system. Go to Frame 3.

Frame 3

Sensor as a Universal Measurement System

What do we desire from an ideal sensor? The text lists five qualities:

1. The output of the sensor should be a well-defined quantity (e.g., a voltage $\hat{x}(t) = U(t)$).
2. It should depend only on one specific quantity being measured (x).
3. The sensor should itself eliminate or average out much of the measurement noise.
4. The sensor should have minimal backward influence on the observed system.
5. The relation $\hat{x}(t) = U(t)$ should hold, meaning the output should be a readable quantity.

A sensor connects the true quantity $z(t)$ (which is $x(t)$ plus noise) to its output estimate $\hat{x}(t)$ via a differential equation. Schematically (see OCR page 1, middle diagram): Input $Z(t) \rightarrow$ [RED K(t)] $\rightarrow$ Output $\hat{X}(t)$ (where "RED K(t)" represents the sensor's dynamics/filtering process).

Go to Frame 4.

Frame 4

Order of a Sensor

The "order" of a sensor is defined by the order of the differential equation that relates its input $z(t)$ to its output $\hat{x}(t)$. Equivalently, a sensor of order u (where $u > 0$ is an integer) can be considered an optimal tracking system for variables $\hat{x}(t)$ in system S whose u-th derivative with respect to time is at most constant (or zero plus white noise $W(t)$):

$$\frac{d^{(u)}}{dt^{(u)}}x(t) = 0 + W(t)$$

Comment on a Thermometer: Think of a thermometer warming up under your arm. It always has some lag. If the body temperature were changing non-linearly (e.g., quadratically or cubically), the thermometer's temperature might "run away" or never accurately track the true temperature due to this lag and its own response characteristics. This illustrates why the dynamic response (order) of a sensor is important.

Go to Frame 5.

Frame 5

First-Order Sensor

In the real system S, we might have: $\dot{x} = W(t)$ (the true quantity x changes randomly, $\langle W(t)^2 \rangle = Q$) $z = x + r(t)$ (we measure x with noise r, $\langle r(t)^2 \rangle = R$)

The Kalman filter equations for optimal filtering state that, for this system (where $A = 0, C = 0, \Gamma = 1, H = 1$ in the general Kalman equations): The estimate evolves as: $\dot{\hat{x}} = K(z - \hat{x})$ The error variance P evolves as: $\dot{P} = -P^2/R + Q$ The Kalman gain is $K = P/R$.

If the gain $K(t)$ becomes constant K_∞ (when P reaches steady state P_∞), then $\dot{P} = 0$. So, $-P_\infty^2/R + Q = 0 \implies P_\infty^2 = QR \implies P_\infty = \sqrt{QR}$. And $K_\infty = P_\infty/R = \sqrt{QR}/R = \sqrt{Q/R}$.

Let's define a time constant $\tau = 1/K_\infty = \sqrt{R/Q}$. The differential equation for the estimate $\hat{x}$ becomes:

$$\frac{1}{K_\infty}\dot{\hat{x}} + \hat{x} = z(t) \implies \tau\dot{\hat{x}} + \hat{x} = z(t)$$

This is the standard differential equation for a **first-order sensor**. It's an optimal indicator for tracking a constant (or slowly varying quantity with process noise Q).

Go to Frame 6.

Frame 6

Example: Thermometer

Consider a thermometer in a cup of coffee at temperature T_z. The thermometer itself has temperature $\hat{T}(t)$. Heat flow P_j through the thermometer wall (surface area S_{area}, thickness d, thermal conductivity λ) is:

$$P_j = \frac{\lambda S_{\text{area}}}{d}(T_z - \hat{T})$$

This heat flow changes the thermometer's internal energy: $P = dQ/dt = mc_p d\hat{T}/dt$. Equating them: $mc_p \frac{d\hat{T}}{dt} = \frac{\lambda S_{\text{area}}}{d}(T_z - \hat{T})$. Rearranging gives:

$$\left(\frac{mc_p d}{\lambda S_{\text{area}}}\right)\dot{\hat{T}} + \hat{T} = T_z(t)$$

If we define $K_c = \left(\frac{mc_p d}{\lambda S_{\text{area}}}\right)$, then $K_c\dot{\hat{T}} + \hat{T} = T_z(t)$. This is exactly the form of a first-order sensor equation, where K_c is the time constant τ.

What does this equation tell us about how the thermometer's reading $\hat{T}$ responds to the coffee's temperature T_z? ____

Go to Frame 7.

Frame 7

The equation $K_c\dot{\hat{T}} + \hat{T} = T_z(t)$ shows that the thermometer's reading $\hat{T}$ will exponentially approach the coffee's temperature T_z with a characteristic time constant K_c. It won't respond instantaneously.

The text now considers typical inputs $z(t)$ to a first-order sensor and their responses. This is about understanding the sensor's behavior when $z(t) \neq$ const. It mentions Green's functions, which describe the sensor's response to an impulse.

Consider the first-order sensor equation: $\tau\dot{\hat{x}} + \hat{x} = z(t)$. **Typical Inputs $z(t)$ for a First-Order Sensor:**

1. $z(t) = \delta(t)$ (Dirac delta function - an impulse at $t = 0$)
2. $z(t) = H_0(t)$ (Heaviside step function - 0 for $t < 0$, H_0 for $t \geq 0$)
3. $z(t) = \alpha t$ (Ramp input)
4. $z(t) = \cos(\omega t)$ (Sinusoidal input - expect same frequency, phase shift, and amplitude change in output $\hat{x}(t)$)

(Diagrams for these inputs and typical responses are shown on OCR page 3).

Let's solve for the homogeneous part of $\tau\dot{\hat{x}} + \hat{x} = z(t)$ first: $\tau\dot{\hat{x}}_H + \hat{x}_H = 0 \implies \tau\frac{d\hat{x}_H}{dt} = -\hat{x}_H$. The solution is $\hat{x}_H(t) = Ce^{-\lambda t}$. Substitute into the homogeneous DE: $\tau C(-\lambda)e^{-\lambda t} + Ce^{-\lambda t} = 0 \implies (-\lambda\tau + 1)Ce^{-\lambda t} = 0$. This requires $\lambda = 1/\tau$. So, $\hat{x}_H(t) = Ce^{-t/\tau}$.

Now, consider the impulse input $z(t) = \delta(t)$. For $t > 0$, $z(t) = 0$, so $\hat{x}(t) = Ce^{-t/\tau}$ for $t > 0$. To find C, we integrate the DE from $-\epsilon$ to $+\epsilon$ and let $\epsilon \to 0$: $\lim_{\epsilon\to 0}\left[\int_{-\epsilon}^{\epsilon}\tau\dot{\hat{x}}dt + \int_{-\epsilon}^{\epsilon}\hat{x}dt\right] = \lim_{\epsilon\to 0}\int_{-\epsilon}^{\epsilon}\delta(t)dt$. The first term gives $\tau[\hat{x}(\epsilon) - \hat{x}(-\epsilon)]$. The second term goes to 0 if $\hat{x}$ is finite. The RHS is 1. Assuming $\hat{x}(t < 0) = 0$, then $\hat{x}(-\epsilon) = 0$. So $\tau\hat{x}(0^+) = 1 \implies \hat{x}(0^+) = 1/\tau$. Thus, for an impulse input, the response (Green's function) is $\hat{x}(t) = \frac{1}{\tau}e^{-t/\tau}$ for $t > 0$.

Go to Frame 8.

Frame 8

Second-Order Sensor

A system (like a mass on a spring with damping, or an RLC circuit) whose dynamics are described by a second-order differential equation acts as a second-order sensor. The text refers to the previous chapter's example of Brownian motion for a particle, which led to: For the state (x, v): $\frac{d^2x}{dt^2} = 0 + W$ (in the absence of drag, if x is position). More generally, the sensor dynamics are given by $\ddot{\hat{x}} + 2\xi\omega_0\dot{\hat{x}} + \omega_0^2\hat{x} =$ terms involving input $z(t)$. The example from Brownian motion (velocity equation): $\dot{v} = -(1/\tau)v + F_x(t)/m$. The optimal Kalman filter for a system like $\ddot{x} = 0 + W$ (constant velocity model with acceleration noise) leads to equations for the error covariance matrix $\boldsymbol{P}$. The steady-state solution leads to constant Kalman gains. The resulting differential equation for the estimate $\hat{x}$ (position) can be written in the standard form of a second-order system:

$$\ddot{\hat{x}} + 2\xi\omega_0\dot{\hat{x}} + \omega_0^2\hat{x} = \omega_0^2 z + 2\xi\omega_0\dot{z}$$

where $\omega_0^2 = \sqrt{Q/R}$ and $2\xi\sqrt{Q/R} = \sqrt{2\sqrt{QR}}/R \cdot \sqrt{Q/R}$ (from specific P_{ij} values in the OCR). A key result is that for the optimal second-order Kalman filter tracking a system like $\ddot{x} = W$, the damping ratio is $\xi = 1/\sqrt{2} \approx 0.707$. This provides good responsiveness without excessive overshoot.

The Green's function for this second-order sensor (response to $z(t) = \delta(t)$) is also discussed. If initially at rest: $\hat{x}(0) = 0, \dot{\hat{x}}(0) = \omega_0^2$. (This results from integrating the DE across the delta function). The homogeneous solution $x_H(t) = C_1 e^{\lambda_1 t} + C_2 e^{\lambda_2 t}$ where $\lambda_{1,2} = -\omega_0[\xi \mp \sqrt{\xi^2 - 1}]$. For $\xi = 1/\sqrt{2}$ (optimal), $\xi^2 - 1 = -1/2$, so $\lambda_{1,2} = -\omega_0[\frac{1}{\sqrt{2}} \mp i\frac{1}{\sqrt{2}}]$. The Green's function is $G(t) = \sqrt{2}\omega_0 \sin(\frac{\omega_0}{\sqrt{2}}t)e^{-\frac{\omega_0}{\sqrt{2}}t}$. (A sketch of this damped sinusoidal response is on OCR page 6).

Go to Frame 9.

Frame 9

Transfer Function

The **transfer function** $H(s)$ of a sensor relates the output $\hat{X}(s)$ to the input $Z(s)$ in the Laplace domain. (This is a brief foray into control theory mathematics). The Laplace Transform is defined as:

$$\mathcal{L}(f(t)) = F(s) = \int_0^\infty e^{-st} f(t) dt$$

Some important properties:

1. $\mathcal{L}(1) = 1/s$
2. $\mathcal{L}(e^{at}) = 1/(s-a)$
3. $\mathcal{L}(f(t)e^{at}) = F(s-a)$
4. $\mathcal{L}(\frac{d}{dt}f(t)) = sF(s) - f(0)$. If $f(0) = 0$, then $\mathcal{L}(\dot{f}(t)) = sF(s)$. Similarly, $\mathcal{L}(\ddot{f}(t)) = s^2 F(s)$ if initial conditions are zero.

A general n-th order linear sensor can be described by: $a_n \frac{d^n \hat{x}}{dt^n} + \cdots + a_1 \dot{\hat{x}} + a_0 \hat{x} = b_m \frac{d^m z}{dt^m} + \cdots + b_1 \dot{z} + b_0 z$. Taking the Laplace Transform (assuming zero initial conditions): $(a_n s^n + \cdots + a_1 s + a_0)X(s) = (b_m s^m + \cdots + b_1 s + b_0)Z(s)$. The transfer function $H(s) = X(s)/Z(s)$ is:

$$H(s) = \frac{b_m s^m + \cdots + b_0}{a_n s^n + \cdots + a_0}$$

Go to Frame 10.

Frame 10

Transfer Function of First-Order Sensor The DE is $\tau\dot{\hat{x}} + \hat{x} = z(t)$. Laplace transform: $(\tau s + 1)X(s) = Z(s)$. Transfer function:

$$H(s) = \frac{X(s)}{Z(s)} = \frac{1}{1 + \tau s}$$

Transfer Function of Second-Order Sensor The DE (simplified form without $\dot{z}$ term for now): $\ddot{\hat{x}} + 2\xi\omega_0\dot{\hat{x}} + \omega_0^2\hat{x} = \omega_0^2 z(t)$. Laplace transform: $(s^2 + 2\xi\omega_0 s + \omega_0^2)X(s) = \omega_0^2 Z(s)$. Transfer function:

$$H(s) = \frac{\omega_0^2}{s^2 + 2\xi\omega_0 s + \omega_0^2}$$

(The OCR text has a slightly different form if $2\xi\omega_0\dot{z}$ term is kept, leading to $H(s) = \frac{2\xi\omega_0 s + \omega_0^2}{s^2 + 2\xi\omega_0 s + \omega_0^2}$).

The comment "Here we then checked in lectures if this really works as expected..." refers to verifying that if an input signal $z(s)$ is multiplied by $H(s)$, the output $x(s)$ makes sense, especially for sinusoidal inputs $z(t) = z_0 e^{i\omega t}$. For sinusoidal inputs, we replace s with $i\omega$ in $H(s)$ to get the frequency response $H(i\omega)$. The output will be $x(t) = x_0 e^{i\omega t} e^{i\delta} = |H(i\omega)| z_0 e^{i(\omega t+\delta)}$, where $|H(i\omega)|$ is the amplitude ratio and $\delta = \arg(H(i\omega))$ is the phase shift.

Go to Frame 11.

Frame 11

Sensor Response to Periodic Signals (Bode Diagrams)

When the input is a periodic signal, like $z(t) = z_0 e^{i\omega t}$, the output will be $\hat{x}(t) = x_0 e^{i(\omega t+\delta)}$. The relationship is given by the frequency response $H(i\omega)$ (obtained by setting $s = i\omega$ in $H(s)$): $H(i\omega) = |H(i\omega)| e^{i\delta(\omega)}$. The output amplitude is $x_0 = |H(i\omega)| z_0$. The phase of the output is shifted by $\delta(\omega)$ relative to the input. $\tan\delta(\omega) = \frac{\mathrm{Im}(H(i\omega))}{\mathrm{Re}(H(i\omega))}$.

Bode diagrams are plots of:

1. Amplitude response: $20\log_{10}|H(i\omega)|$ (in decibels, dB) vs. $\log_{10}\omega$.
2. Phase response: $\delta(\omega)$ (in degrees or radians) vs. $\log_{10}\omega$.

(Note: Decibel is defined for power ratios $10\log(P_2/P_1)$. For amplitude ratios, it's $20\log(A_2/A_1)$ because power is often proportional to amplitude squared).

Go to Frame 12.

Frame 12

Bode Plot for First-Order System (Low-Pass Filter) $H(s) = \frac{1}{1+\tau s} \implies H(i\omega) = \frac{1}{1+i\omega\tau}$. Magnitude: $|H(i\omega)| = \frac{1}{\sqrt{1+(\omega\tau)^2}}$. Phase: $\delta(\omega) = -\arctan(\omega\tau)$.

Limiting cases for magnitude $20\log|H(i\omega)|$:

- $\omega\tau \ll 1$ (low frequencies, $\omega \to 0$): $|H| \approx 1 \implies 20\log|H| \approx 0$ dB.
- $\omega\tau \gg 1$ (high frequencies, $\omega \to \infty$): $|H| \approx 1/(\omega\tau) \implies 20\log|H| \approx -20\log(\omega\tau)$. This is a line with slope -20 dB per decade of frequency.
- $\omega\tau = 1$ (cutoff/corner frequency): $|H| = 1/\sqrt{2} \implies 20\log|H| \approx 20(-0.15) \approx -3$ dB.

(A sketch like OCR page 11, top diagram, would be here). This is a **Low-Pass Filter (LPF)**. It acts as an integrator for high frequencies.

Limiting cases for phase $\delta(\omega)$:

- $\omega\tau \to 0 : \delta \to -\arctan(0) = 0$.
- $\omega\tau \to \infty : \delta \to -\arctan(\infty) = -\pi/2$ radians $(-90°)$.
- $\omega\tau = 1 : \delta = -\arctan(1) = -\pi/4$ radians $(-45°)$.

(A sketch like OCR page 11, bottom diagram, would be here).

Go to Frame 13.

Frame 13

Bode Plot for $H(i\omega) = (i\omega)^n$ $|H(i\omega)| = |\omega^n e^{in\pi/2}| = \omega^n$. $20\log|H| = 20n\log\omega$. This is a line with slope $20n$ dB/decade. Phase $\delta = n\pi/2$. (A sketch like OCR page 12, middle, shows lines for n=1, n=2, n=-1).

Example: High-Pass Filter (HPF) An RC circuit with output taken across the resistor forms a high-pass filter. Impedance $Z = R + 1/(i\omega C) = (1 + i\omega RC)/(i\omega C)$. Output voltage $U_{\text{out}} = IR = (U_{\text{in}}/Z)R$. $H(i\omega) = \frac{U_{\text{out}}}{U_{\text{in}}} = \frac{R}{Z} = \frac{i\omega RC}{1+i\omega RC} = \frac{i\omega\tau}{1+i\omega\tau}$ (where $\tau = RC$).

Magnitude $20\log|H(i\omega)|$:

- $\omega\tau \ll 1 : |H| \approx \omega\tau \implies 20\log|H| \approx 20\log(\omega\tau)$ (slope +20 dB/decade).
- $\omega\tau \gg 1 : |H| \approx 1 \implies 20\log|H| \approx 0$ dB.
- $\omega\tau = 1 : |H| = 1/\sqrt{2} \implies -3$ dB.

(Sketch is on OCR page 14, top. It's a mirror image of the LPF magnitude plot). This filter passes high frequencies and attenuates low frequencies. For low frequencies (where it attenuates), it acts as a differentiator ($H \approx i\omega\tau$).

Go to Frame 14.

Frame 14

Bode Plot for Second-Order System $H(i\omega) = \frac{\omega_0^2}{(i\omega)^2+2\xi\omega_0(i\omega)+\omega_0^2} = \frac{1}{1-(\omega/\omega_0)^2+i(2\xi\omega/\omega_0)}$. Let $x = \omega/\omega_0$. $H(ix\omega_0) = \frac{1}{(1-x^2)+i(2\xi x)}$. Magnitude $|H(ix\omega_0)| = \frac{1}{\sqrt{(1-x^2)^2+(2\xi x)^2}}$. Phase $\delta = -\arctan\left(\frac{2\xi x}{1-x^2}\right)$.

Magnitude $20\log|H|$:

- $x \ll 1(\omega \ll \omega_0) : |H| \approx 1 \implies 0$ dB.
- $x \gg 1(\omega \gg \omega_0) : |H| \approx 1/x^2 = (\omega_0/\omega)^2 \implies -40\log x = -40\log(\omega/\omega_0)$. Slope -40 dB/decade.
- $x = 1(\omega = \omega_0) : |H| = 1/(2\xi)$. If ξ is small (e.g., $\xi \ll 1$), $|H|$ can be very large (resonance peak). If $\xi = 1/\sqrt{2} \approx 0.707$, then $|H| \approx 1/\sqrt{2}$ (actually $1/(2\cdot 0.707) \approx 1/1.414 \approx 0.707$, so -3dB point is not exactly at $x = 1$ unless ξ is large). For $\xi = 1/\sqrt{2}$, at $x = 1$, $|H| = 1/(\sqrt{2}) \approx 0.707$, so $20\log|H| \approx -3$dB.

(Sketch is on OCR page 15, top, showing resonance peaks for small ξ). This is a resonant low-pass filter.

Phase δ:

- $x \to 0 : \delta \to 0$.
- $x \to \infty : \delta \to -\arctan(0^-) = -\pi$ radians ($-180°$) (since $1 - x^2$ becomes large negative).
- $x = 1(\omega = \omega_0) : \delta = -\arctan(\pm\infty) = -\pi/2$ radians ($-90°$) (denominator $1 - x^2 = 0$).

(Sketch is on OCR page 16, top).

Go to Frame 15.

Frame 15

Impact of the Sensor on the Observed System

We want to quantify how much a sensor perturbs the system S it's measuring. (Diagram on OCR page 16, middle, shows system S with an "internal" voltage U_{AB} and a sensor M connected to it). Let the sensor draw power P_M from the system: $P_M = U_M^2/Z_M$ (if Z_M is resistive part of sensor impedance, U_M is voltage across sensor). This power is needed for the sensor to operate.

Thevenin's Theorem Any linear electrical network with two terminals A and B can be replaced by an equivalent circuit consisting of an ideal voltage source U_{AB} (Thevenin voltage) in series with an internal impedance Z_{AB} (Thevenin impedance). (Diagram shows U_{AB} in series with Z_{AB}, and the measurement system Z_M connected across them). The current $I = U_{AB}/(Z_{AB} + Z_M)$. The voltage across the measurement system (sensor) is $U_M = IZ_M = U_{AB}\frac{Z_M}{Z_{AB}+Z_M} = U_{AB}\frac{1}{1+Z_{AB}/Z_M}$.

For U_M to be very close to U_{AB} (i.e., the sensor accurately measures the system's open-circuit voltage and doesn't "load" it down), what should be the relationship between Z_M and Z_{AB}? [a] $Z_M \ll Z_{AB}$ [b] $Z_M \approx Z_{AB}$ [c] $Z_M \gg Z_{AB}$

Go to Frame 16.

Frame 16

Your answer was [a — b — c].

The correct answer is [c]. For $U_M \approx U_{AB}$, we need $Z_{AB}/Z_M \to 0$, which means the sensor's input impedance Z_M should be much larger than the system's Thevenin (output) impedance Z_{AB}. This is the concept of high **input impedance** for a measuring device like a voltmeter.

The power delivered to the sensor is $P_M = U_M^2/Z_M = U_{AB}^2\frac{Z_M}{(Z_{AB}+Z_M)^2}$. To maximize this power transfer (e.g., for an antenna feeding a receiver), we set $dP_M/dZ_M = 0$. This occurs when $Z_M = Z_{AB}$ (impedance matching). However, for minimal disturbance, we want minimal power drawn by the sensor. From $P_M/P_{\max} \approx 4Z_{AB}/Z_M$ (for $Z_M \gg Z_{AB}$, where $P_{\max} = U_{AB}^2/(4Z_{AB})$ is max power when $Z_M = Z_{AB}$), we see that if Z_M is large, P_M is small.

So, the "takeaway message":

1. For a sensor to accurately measure a quantity (like voltage) without disturbing the system, its input impedance (Z_{in} of sensor) should be very high: $Z_{\text{in}} \to \infty$.

2. For a source (like a signal generator, or the output of a sensor that drives another device) to deliver its signal effectively without being loaded down, its output impedance (Z_{out} of source) should be very low: $Z_{\text{out}} \to 0$.

End of Section.

6 Operational Amplifiers (Op-amps) and Instrumentation Amplifiers

Frame 1

An **operational amplifier (op-amp)** is an electronic circuit designed to amplify weak electrical signals. It typically has two inputs:

- An "inverting" input (often denoted U_1 or V_-)
- A "non-inverting" input (often denoted U_2 or V_+)

and one output (U_{izh} or V_{out}). Its primary purpose is to amplify the *difference* in voltage between these two inputs. Op-amps also require power supply inputs, which are usually omitted in simplified schematics.

A schematic representation is shown (see OCR page 1, top diagram, depicting a triangle symbol for the op-amp with U_1, U_2 inputs, U_{izh} output, and power supply connections often implied or drawn to "GND" or power rails).

The transfer function of an op-amp can be roughly described as:

$$H = A_{DC} H_{LPF}$$

What do A_{DC} and H_{LPF} represent in this context? 1. A_{DC}: ____________ 2. H_{LPF}: ____________

Go to Frame 2.

Frame 2

In the transfer function $H = A_{DC} H_{LPF}$: 1. A_{DC} is the **DC open-loop gain** of the op-amp. 2. H_{LPF} is the transfer function of a **low-pass filter**, representing the op-amp's frequency-dependent response (gain typically decreases at high frequencies).

The output voltage U_{izh} is given by the formula:

$$U_{\text{izh}} = A_{DC}(U_1 - U_2)$$

(Note: The original text seems to use U_1 as non-inverting and U_2 as inverting if the formula is $A_{DC}(U_1 - U_2)$ and later connects output to U_2 for negative feedback. Standard notation often has V_+ for non-inverting and V_- for inverting, with $V_{out} = A(V_+ - V_-)$. We'll follow the text's variable usage.)

The term "open-loop gain" (A_{DC}) means this is the gain in the absence of any feedback connections. A_{DC} is typically very large, for example, $A_{DC} \sim 10^6$.

What is a potential problem if A_{DC} is very large and we apply even a small difference voltage ($U_1 - U_2$)? ____

Go to Frame 3.

Frame 3

If A_{DC} is very large (e.g., 10^6), even a tiny input difference voltage ($U_1 - U_2$) would theoretically result in an enormous output voltage ($10^6 \times$ difference). In practice, the output voltage is limited by the power supply voltages. This leads to **saturation** (the output "hits the rails") and signal **distortion** (clipping).

To control this high gain and make op-amps useful, what circuit technique is commonly employed? [a] Positive feedback [b] Negative feedback [c] Using smaller power supplies

Go to Frame 4.

Frame 4

Your answer was [a — b — c].

The correct answer is [b] **Negative feedback**. Negative feedback is used to create stable amplifiers with well-defined gains. (Positive feedback is sometimes used in regenerative circuits like oscillators or Schmitt triggers, but not typically for linear amplification).

Negative Feedback Configuration (Non-inverting Amplifier Example from OCR) To create a negative feedback loop, the output of the op-amp is connected back to the inverting input (here, U_2) often through a voltage divider formed by resistors. The OCR shows a specific configuration that it calls a "non-inverting amplifier" where the input U_{in} is applied to the non-inverting terminal U_1, and the output U_{out} is fed back to the inverting terminal U_2 via a resistor F (acting as part of a voltage divider, with another resistor often implied to ground from U_2).

(See diagram on OCR page 1, middle. Input U_{in} to V_+ (U_1), output U_{out} (U_{izh}) connected via F to V_- (U_2). There's often another resistor from V_- to ground, let's call it R_G. Then $U_2 = U_{\text{out}} \frac{R_G}{F+R_G}$. The diagram in the OCR seems to simplify this, perhaps implying F is a factor of a voltage divider or F is the feedback resistor and the other is implicitly 1 unit). The OCR text seems to use F as a *feedback factor* rather than a single resistor value directly. Let's assume $U_2 = F \cdot U_{\text{out}}$. The op-amp equation is $U_{\text{out}} = A(U_1 - U_2)$. Substituting $U_1 = U_{\text{in}}$ and $U_2 = FU_{\text{out}}$:

$$U_{\text{out}} = A(U_{\text{in}} - FU_{\text{out}})$$

$$U_{\text{out}} = AU_{\text{in}} - AFU_{\text{out}}$$

$$U_{\text{out}}(1 + AF) = AU_{\text{in}}$$

The closed-loop gain $H = U_{\text{out}}/U_{\text{in}}$ is:

$$H = \frac{A}{1 + AF}$$

If A (the open-loop gain) is very large, what does H approximate to? (Hint: If $AF \gg 1$, then $1 + AF \approx AF$). ____

Go to Frame 5.

Frame 5

If A is very large such that $AF \gg 1$, then $1 + AF \approx AF$. So, the closed-loop gain H becomes:

$$H \approx \frac{A}{AF} = \frac{1}{F}$$

This is a very important result: the closed-loop gain of an op-amp circuit with negative feedback primarily depends on the external feedback components (represented by F), and *not* on the op-amp's large and often imprecise open-loop gain A.

The text states: "We see: $F = 1 \implies H = U_{\text{out}}/U_{\text{in}} = 1$." This configuration where $F = 1$ (meaning all the output is fed back to the inverting input, $U_2 = U_{\text{out}}$) is called a **voltage follower** or **buffer**. Its gain is 1.

What is a key advantage of a voltage follower, even though its voltage gain is only 1? (Hint: think about input and output impedances from the previous chapter). ____________________

Go to Frame 6.

Frame 6

A key advantage of a voltage follower (buffer) is that it has a very **high input impedance** and a very **low output impedance**. This means it can connect a source with a high output impedance to a load with a low input impedance without significant signal loss or loading effects. It "buffers" the source from the load. The text notes: "This allows us to connect to the input without drawing power from the signal. As we found in the previous chapter, it has $Z_{\text{in}} \to \infty$."

Go to Frame 7.

Frame 7

Instrumentation Amplifier

Now let's look at an instrumentation amplifier. This is a more complex circuit typically built from three op-amps. Its purpose is to amplify the difference between two input signals while rejecting any signal common to both inputs (common-mode noise). It provides high input impedance and precise, stable gain. (A diagram like the one on OCR page 2, top, is essential here).

The diagram shows:

- Two input op-amps (buffers or non-inverting amplifiers) connected to inputs U_1 and U_2.
- A resistor R_0 (often called R_G for gain) connecting the inverting inputs of these first two op-amps.
- Resistors R_1 connecting the outputs of the first stage op-amps to their respective inverting inputs.
- A third op-amp configured as a differential amplifier, taking inputs from the outputs of the first two op-amps ($\tilde{U}_1, \tilde{U}_2$). This stage uses resistors R_2', R_3'.

Let's analyze the first stage (inputs U_1, U_2, outputs $\tilde{U}_1, \tilde{U}_2$). Applying Kirchhoff's current law at the node between the two R_1 resistors and R_0 (or by noting that due to high open-loop gain, the inverting and non-inverting inputs of each op-amp are virtually at the same potential for the first stage op-amps when feedback is active): Current through R_1 from U_1 side: $(\tilde{U}_1 - U_1)/R_1$ Current through R_1 from U_2 side: $(U_2 - \tilde{U}_2)/R_1$ (current direction reversed) Current through R_0: $(U_1 - U_2)/R_0$ (since inverting inputs follow non-inverting inputs). The text directly states (likely from op-amp analysis where current into op-amp inputs is zero, and $V_+ = V_-$): The current through the series R_1, R_0, R_1 (conceptually) is the same. $(\tilde{U}_1 - U_1)/R_1 + (\tilde{U}_2 - U_2)/R_1 = (U_1 - U_2)/R_0 -$ this is incorrect.

The correct analysis for the first stage yields: The voltage across R_0 is $U_1 - U_2$ (since V_+ tracks V_- for the input op-amps). The current through R_0 is $I_{R0} = (U_1 - U_2)/R_0$. This same current flows through both R_1 resistors. So, $\tilde{U}_1 = U_1 + I_{R0} R_1 = U_1 + \frac{R_1}{R_0}(U_1 - U_2)$ And $\tilde{U}_2 = U_2 - I_{R0} R_1 = U_2 - \frac{R_1}{R_0}(U_1 - U_2)$ The differential output of the first stage is $\Delta\tilde{U} = \tilde{U}_1 - \tilde{U}_2$:

$$\Delta\tilde{U} = (U_1 - U_2) + \frac{2R_1}{R_0}(U_1 - U_2) = (U_1 - U_2)\left(1 + \frac{2R_1}{R_0}\right)$$

The text uses a slightly different derivation but arrives at a similar form for $\Delta\tilde{U} = (\tilde{U}_1 - \tilde{U}_2) = (2\frac{R_1}{R_0} + 1)(U_1 - U_2) = (2\frac{R_1}{R_0} + 1)\Delta U$. This is the gain of the first differential stage.

Go to Frame 8.

Frame 8

Now for the second stage (the differential amplifier with inputs $\tilde{U}_1, \tilde{U}_2$ and output U_{out}). The standard formula for a differential amplifier with four resistors R_a, R_b, R_c, R_d (where $\tilde{U}_1$ goes to R_a then to inverting input, $\tilde{U}_2$ goes to R_c then to non-inverting input, R_b is feedback, R_d from non-inverting to ground) is $U_{\text{out}} = \frac{R_d}{R_c+R_d}\left(1+\frac{R_b}{R_a}\right)\tilde{U}_2 - \frac{R_b}{R_a}\tilde{U}_1$. If $R_a = R_c = R_2'$ and $R_b = R_d = R_3'$, then this simplifies to:

$$U_{\text{out}} = \frac{R_3'}{R_2'}(\tilde{U}_2 - \tilde{U}_1) = -\frac{R_3'}{R_2'}(\tilde{U}_1 - \tilde{U}_2) = -\frac{R_3'}{R_2'}\Delta\tilde{U}$$

The text has a slightly different setup or intermediate variables U. It derives for the second stage: $U = \tilde{U}_2\frac{R_3'}{R_2'+R_3'}$ (voltage at non-inverting input of 3rd op-amp). And $(\tilde{U}_1 - U)/R_2' = (U - U_{\text{out}})/R_3'$ (currents into inverting node). Leading to $U_{\text{out}} = -\frac{R_3}{R_2}[\tilde{U}_1 - (1+\frac{R_2}{R_3})\frac{R_3'}{R_2'+R_3'}\tilde{U}_2]$. (Using R_2, R_3 from a diagram not fully shown for this stage). If perfectly balanced, $R_2 = R_2'$ and $R_3 = R_3'$, the gain for the differential signal $\Delta\tilde{U}$ is $-R_3'/R_2'$.

Combining with the first stage gain $\Delta\tilde{U} = (1+\frac{2R_1}{R_0})\Delta U$: Overall gain $A_{\text{diff}} = \frac{U_{\text{out}}}{\Delta U} = -\frac{R_3'}{R_2'}\left(1+\frac{2R_1}{R_0}\right)$. The crucial part is how it rejects common-mode signals.

Go to Frame 9.

Frame 9

Common Mode Rejection Ratio (CMRR)

The goal is to make the pre-factor for $\tilde{U}_2$ in the expression for U_{out} (from the full derivation in the OCR page 3, top) equal to 1, when it's actually part of $-(R_3/R_2)\tilde{U}_1 + (\text{factor})\tilde{U}_2$. The text aims for the output to be $U_{\text{out}} = -G(\tilde{U}_1 - \tilde{U}_2)$. The condition derived for ideal differential amplification (perfect common-mode rejection) is:

$$(1+\frac{R_2}{R_3})\frac{R_3'}{R_2'+R_3'} = 1$$

This simplifies to the well-known condition for a differential amplifier to reject common-mode signals:

$$\frac{R_2}{R_3} = \frac{R_2'}{R_3'}$$

(i.e., the resistor ratios must be matched).

In practice, this ratio is never perfect. Let's say $\frac{R_3'}{R_2'+R_3'}(1+\frac{R_2}{R_3}) = \frac{1+\epsilon}{1-\epsilon}$ where $\epsilon \to 0$ represents a small mismatch. Then the output voltage is (from OCR page 3):

$$U_{\text{out}} \approx -\frac{R_3}{R_2(1-\epsilon)}[\Delta\tilde{U} - \epsilon(\tilde{U}_1 + \tilde{U}_2)]$$

The term $\tilde{U}_1 + \tilde{U}_2 = 2U_{CM}$ (where U_{CM} is the common-mode input to the second stage). The differential gain $A_d \approx -R_3/(R_2(1-\epsilon))$. The common-mode gain $A_{cm} \approx -R_3/(R_2(1-\epsilon)) \cdot (-\epsilon)$. The **Common Mode Rejection Ratio (CMRR)** is defined as the ratio of differential gain to common-mode gain:

$$\text{CMRR} = \left|\frac{A_d}{A_{cm}}\right| = \left|\frac{-\frac{R_3}{R_2(1-\epsilon)}}{\frac{R_3\epsilon}{R_2(1-\epsilon)}}\right| = \frac{1}{|\epsilon|}$$

A good instrumentation amplifier has a very high CMRR (e.g., 10^6, which is 120 dB since $\text{CMRR(dB)} = 20\log_{10}(\text{CMRR})$). This means it strongly amplifies differences but rejects signals common to both inputs.

This detailed analysis shows the principles behind designing precise amplifiers that are resilient to common-mode noise. End of Section.

7 Thermal Noise in a Resistor and its Propagation

Frame 1

Consider charge moving through a conductor against resistance (i.e., current flow). This random thermal motion of charge carriers gives rise to a fluctuating voltage known as **thermal noise** or Johnson-Nyquist noise. (A diagram like OCR page 1, top, showing a resistor R with a noise voltage source $e(t)$ or $U_g(t)$ in series would be helpful).

Let's write down some properties for the noise voltage $U_g(t)$ across a resistor R due to current $I(t)$: $U_g(t) = IR$ (Ohm's Law for the instantaneous noise current and voltage). The key statistical properties of this thermal noise voltage are:

1. $\langle U_g(t) \rangle = 0$: The average noise voltage is zero.
2. $\langle U_g^2(t) \rangle \neq 0$: The mean square voltage (related to noise power) is non-zero.

Go to Frame 2.

Frame 2

Now, let's consider a simple RC circuit: a voltage source $U_g(t)$ (representing the thermal noise of a resistor R) in series with the resistor R and a capacitor C. (Diagram like OCR page 1, middle, showing U_g source, R, C in series, with U_C across the capacitor). The current is I. The charge on the capacitor is e, so $e = CU_C$ and $de/dt = I = C\dot{U}_C$. Applying Kirchhoff's voltage law:

$$U_g - IR - \frac{e}{C} = 0$$

Substituting $I = C\dot{U}_C$ and $e/C = U_C$:

$$U_g - RC\dot{U}_C - U_C = 0$$

Rearranging to get the dynamics for U_C:

$$RC\dot{U}_C + U_C = U_g$$

Let $\tau = RC$ (the time constant of the circuit).

$$\dot{U}_C = -\frac{1}{\tau}U_C + \frac{1}{\tau}U_g$$

This equation describes how the voltage across the capacitor U_C responds to the noise voltage U_g. This is in the form of a Kalman dynamics equation for U_C (our estimate/filtered value), where U_g/τ acts as the "dynamic noise" or input driving the system, and $-U_C/\tau$ is the system dynamics part. The text implies that $\langle U_C \rangle = 0$ and the variance (covariance matrix, though scalar here) is $\langle U_C^2 \rangle = P$.

Go to Frame 3.

Frame 3

From the previous chapter (Frame 11, Section on Vector Variables, or general Kalman filter for scalar), the steady-state variance $P = \langle U_C^2 \rangle$ (when $\dot{P} = 0$) for a system $\dot{x} = Ax + \Gamma w$ (where $x = U_C, A = -1/\tau, \Gamma = 1/\tau, w = U_g$) with process noise Q_{eff} (representing variance of Γw) is given by the Riccati equation: $\dot{P} = 2AP + \Gamma^2 Q_{\text{eff}} = 0$ (assuming no measurement noise term for now, as U_C is the state). Here, the "process noise" driving U_C is U_g/τ. Let $\langle U_g^2 \rangle = Q_U$ (this Q_U is related to physical Q later). Then the effective driving noise variance for the $\dot{U}_C$ equation is $Q_{\text{eff}} = \langle (U_g/\tau)^2 \rangle = Q_U/\tau^2$. So, in steady state:

$$-\frac{2}{\tau}P + \frac{1}{\tau^2}Q_U = 0$$

$$P = \frac{Q_U}{2\tau}$$

Thus, $\langle U_C^2 \rangle = P_\infty = Q_U/(2\tau)$.

Now, let's connect this to thermodynamics. The capacitor stores energy. The average energy stored in a capacitor in thermal equilibrium at temperature T_{abs} is given by the equipartition theorem:

$$\langle W_C \rangle = \frac{1}{2}C\langle U_C^2 \rangle = \frac{1}{2}k_B T_{\text{abs}}$$

(where k_B is Boltzmann's constant). So, $\langle U_C^2 \rangle = k_B T_{\text{abs}}/C$. Equating the two expressions for $\langle U_C^2 \rangle$: $P_\infty = k_B T_{\text{abs}}/C = Q_U/(2\tau)$. Since $\tau = RC$, we have $Q_U = 2\tau k_B T_{\text{abs}}/C = 2RCk_B T_{\text{abs}}/C = 2k_B T_{\text{abs}} R$. This $Q_U = \langle U_g^2 \rangle$ is the mean square value of the open-circuit thermal noise voltage from resistor R. This is a form of the Nyquist formula for total noise power.

Go to Frame 4.

Frame 4

Dynamic Noise (Frequency Characteristics)

We assume the thermal noise $U_g(t)$ is **white noise**. What does this mean for its correlation in time? [a] Correlated over long times [b] Correlated only over very short times (effectively uncorrelated for different t, t') [c] Perfectly correlated for all times

Go to Frame 5.

Frame 5

Your answer was [a — b — c].

The correct answer is [b]. White noise is, by definition, completely uncorrelated in time. Its autocorrelation function is a Dirac delta function:

$$\langle U_g(t)U_g(t+T') \rangle = Q_0 \delta(T')$$

(Here Q_0 is a constant representing the noise power spectral density, related to $Q_U = 2k_B T_{\text{abs}} R$ found earlier, but integrated over all frequencies, or Q_0 is $S_0/2$ where S_0 is the double-sided PSD). The OCR uses $Q\delta(T)$ which might imply Q is $2k_B T_{\text{abs}} R$ if integrated.

Let's estimate the magnitude of this noise. If $R = 1\ \text{M}\Omega = 10^6\Omega$, $\tau = 1\mu\text{s} = 10^{-6}$ s, and $T_{\text{abs}} = 300$ K (room temp). $k_B \approx 1.38 \times 10^{-23}$ J/K. $\langle U_C^2 \rangle = P_\infty = k_B T_{\text{abs}}/C$. Since $\tau = RC$, $C = \tau/R = 10^{-6}/10^6 = 10^{-12}$ F $= 1$ pF. $P_\infty = (1.38 \times 10^{-23} \times 300)/10^{-12} \approx 4.14 \times 10^{-9}$ V^2. The RMS voltage $\sqrt{P_\infty} = \sqrt{\langle U_C^2 \rangle} \approx \sqrt{4.14 \times 10^{-9}} \approx 6.4 \times 10^{-5}$ V $= 64\mu$V. The OCR calculation directly uses $P_\infty = Q/(2\tau)$ and $Q = 2k_B TR/C \cdot \tau = 2k_B TR$ (this Q seems to be Q_U from earlier). $P_\infty = (2k_B T_{\text{abs}} R)/(2\tau) = k_B T_{\text{abs}} R/\tau$. $\sqrt{P_\infty} = \sqrt{1.38 \times 10^{-23} \times 300 \times 10^6/10^{-6}}$ V $\approx \sqrt{4.14 \times 10^{-9}}$ V $\approx 64\mu$V. The OCR has a slightly different formula/substitution yielding $10\mu V$. The discrepancy might be in the definition of Q used in $P = Q/(2\tau)$. If Q in $P = Q/(2\tau)$ is the Q_0 from the autocorrelation (power spectral density value), then the relation is different. The standard result for an RC filter driven by white voltage noise U_g with double-sided PSD $N_0/2 = 2k_BTR$ is $\langle U_C^2 \rangle = k_BT/C$.

Go to Frame 6.

Frame 6

Spectral Density of Dynamic Noise

The relation $\langle U_g(t)U_g(t+T') \rangle = \text{const} \cdot \delta(T')$ implies that the noise power is uniformly distributed across all frequencies (this is the definition of white noise). This is the idealization. If this were truly valid for all frequencies, the total noise power ($\int S(\omega)d\omega$, where $S(\omega)$ is power spectral density) would be infinite. (A diagram like OCR page 2, middle, shows a flat spectral density $d\langle U^2 \rangle/d\omega$ vs ω). The integral $P = \int_0^\infty \frac{d\langle U^2 \rangle}{d\omega} d\omega$ would diverge if $d\langle U^2 \rangle/d\omega$ is a non-zero constant.

To resolve this, we consider that physical noise sources are only "white" up to a certain cutoff frequency. Alternatively, we can analyze the system by considering its response to individual frequency components independently.

If a voltage U is composed of two frequency components: $U = U_0(\omega_1)\cos(\omega_1 t) + U_0(\omega_2)\cos(\omega_2 t + \delta_p)$ Then the mean square voltage $\langle U^2 \rangle$ is:

$$\langle U^2 \rangle = \frac{1}{2}U_0^2(\omega_1) + \frac{1}{2}U_0^2(\omega_2) + \langle U_0(\omega_1)U_0(\omega_2)\cos(\omega_1 t)\cos(\omega_2 t + \delta_p) \rangle$$

If $\omega_1 \neq \omega_2$, the last term (cross-term) averages to zero over a long time because the frequencies are different. This means we can treat frequencies independently when calculating average power.

Go to Frame 7.

Frame 7

Wiener-Khinchin Theorem

This theorem provides a fundamental link between the time-domain and frequency-domain descriptions of a stationary random process (like noise). It states (roughly): The **power spectral density** of a noise signal is the **Fourier Transform** of its **autocorrelation function**.

1. Autocorrelation function $c(T')$: Measures how related the signal $U(t)$ is to a time-shifted version of itself $U(t+T')$, averaged over t.

$$c(T') = \int_{-\infty}^{\infty} U(t)U(t+T')dt$$

(This is for deterministic signals; for random signals, it's an expectation $\langle U(t)U(t+T') \rangle$).

2. Fourier Transform pair: $U(t) = \int_{-\infty}^{\infty} U_\nu e^{-2\pi i\nu t} d\nu$ $U_\nu = \int_{-\infty}^{\infty} U(t)e^{+2\pi i\nu t} dt$ (Using ν for frequency here, $\omega = 2\pi\nu$).

The theorem then states that the power spectral density $S_U(\nu) = |U_\nu|^2$ is related to $c(T')$:

$$c(T') = \int_{-\infty}^{\infty} |U_\nu|^2 e^{-2\pi i\nu T'} d\nu$$

And conversely:

$$|U_\nu|^2 = \int_{-\infty}^{\infty} c(T')e^{+2\pi i\nu T'} dT'$$

(The OCR uses $1/\pi$ and ω which corresponds to a different Fourier transform convention, often $S_U(\omega) = \int c(T')e^{-i\omega T'}dT'$). The text uses: $|U_\omega|^2 = \frac{1}{\pi}\int_{-\infty}^{\infty} e^{-i\omega T'}c(T')dT'$. This $|U_\omega|^2$ is $d\langle U^2 \rangle/d\omega$. For our case, $c(T') = \langle U_g(t)U_g(t+T') \rangle = Q_0\delta(T')$. (Using Q_0 for the constant in the delta function). Then $d\langle U_g^2\rangle/d\omega = \frac{1}{\pi}\int Q_0\delta(T')e^{-i\omega T'}dT' = Q_0/\pi$. The original text uses $\langle U(t)U(t+T')\rangle = 2k_BTR\delta(T')$. This means the power spectral density $d\langle U_g^2\rangle/d\omega = \frac{2k_BT_{\text{abs}}R}{\pi} = \text{constant}$. This is the spectral density of the thermal noise voltage source itself.

Go to Frame 8.

Frame 8

Nyquist's Derivation (Experimental Setup)

Nyquist considered an experimental setup: a coaxial cable of length L and characteristic impedance $Z = R$, terminated at both ends by resistors R. (Diagram on OCR page 4, middle, shows this setup). He was interested in the number of electromagnetic modes n in the cable per unit frequency $dn/d\omega$. The condition for standing waves in the cable (shorted or open ends, approximately): $L = n\frac{\lambda}{2}$. Since $\lambda = c/\nu = 2\pi c/\omega$ (where c is speed of light in cable): $L = n\frac{\pi c}{\omega} \implies n = \frac{\omega L}{\pi c}$. So, the density of modes is:

$$\frac{dn}{d\omega} = \frac{L}{\pi c}$$

Each of these modes is considered a degree of freedom. According to Bose-Einstein statistics, the average energy $\epsilon(\omega)$ per mode at temperature T_{abs} is:

$$\epsilon(\omega) = \hbar\omega\frac{1}{e^{\hbar\omega/k_BT_{\text{abs}}} - 1}$$

The average power dP flowing in these modes in a frequency interval $d\omega$ is (from the text's derivation, accounting for energy flow): The text relates this to the average power dissipated in a resistor: $dP = Rd\langle I^2\rangle$. Using $\langle I^2\rangle = \langle U^2\rangle/(4R^2)$ for power waves in a matched system. The spectral density of mean square noise voltage across one resistor is $d\langle U^2\rangle/d\omega$. The text arrives at $d\langle U^2\rangle/d\omega = \frac{2}{\pi}R\hbar\omega\frac{1}{e^{\hbar\omega/k_BT_{\text{abs}}-1}}$ by combining the density of modes with the average energy per mode, and relating power to voltage. (The intermediate steps on OCR page 4 are condensed. The idea is power per mode $\times$ modes per unit frequency = power per unit frequency).

This is Nyquist's formula for the spectral density of thermal noise voltage.

Go to Frame 9.

Frame 9

Let's look at the Nyquist formula for $d\langle U^2\rangle/d\omega$:

$$\frac{d\langle U^2\rangle}{d\omega} = \frac{2R}{\pi}\frac{\hbar\omega}{e^{\hbar\omega/k_BT_{\text{abs}}} - 1}$$

Consider two limits: 1. High frequencies: $\hbar\omega \gg k_BT_{\text{abs}}$. What happens to the exponential term and then to the spectral density? __2. Low frequencies: $\hbar\omega \ll k_BT_{\text{abs}}$. Use $e^x \approx 1 + x$ for small x. What does the spectral density become? ____

Go to Frame 10.

Frame 10

1. High frequencies ($\hbar\omega \gg k_BT_{\text{abs}}$): $e^{\hbar\omega/k_BT_{\text{abs}}} \gg 1$. So $e^{\hbar\omega/k_BT_{\text{abs}}} - 1 \approx e^{\hbar\omega/k_BT_{\text{abs}}}$. $\frac{d\langle U^2\rangle}{d\omega} \approx \frac{2R}{\pi}\hbar\omega e^{-\hbar\omega/k_BT_{\text{abs}}}$. The density drops off exponentially (quantum regime).

2. Low frequencies ($\hbar\omega \ll k_BT_{\text{abs}}$): $e^{\hbar\omega/k_BT_{\text{abs}}} - 1 \approx (1 + \hbar\omega/k_BT_{\text{abs}}) - 1 = \hbar\omega/k_BT_{\text{abs}}$.

$$\frac{d\langle U^2\rangle}{d\omega} \approx \frac{2R}{\pi}\frac{\hbar\omega}{\hbar\omega/k_BT_{\text{abs}}} = \frac{2Rk_BT_{\text{abs}}}{\pi}$$

This is constant! This is the "white noise" approximation valid at lower frequencies (classical regime). The cutoff frequency where $\hbar\omega \approx k_B T_{\text{abs}}$ is around $10^{13} - 10^{14}$ Hz at room temperature, so for most electronics, the constant spectral density is a very good approximation. This constant value $S_0 = 2Rk_B T_{\text{abs}}/\pi$ (for single-sided $\omega \in [0, \infty)$) or $4Rk_B T_{\text{abs}}$ (for double-sided PSD in Hz, $S_0(f) = 4kTR$) is often used.

Go to Frame 11.

Frame 11

Propagation of Thermal Noise Through a Linear Circuit

If an input voltage $U_{\text{in}}(i\omega)$ (a frequency component of noise) is applied to a linear circuit with transfer function $H(i\omega)$, the output voltage component is $U_{\text{out}}(i\omega) = H(i\omega)U_{\text{in}}(i\omega)$. For power (or mean square voltage, which is proportional to power): $|U_{\text{out}}(i\omega)|^2 = |H(i\omega)|^2 |U_{\text{in}}(i\omega)|^2$. In terms of spectral densities:

$$\frac{d\langle U_{\text{out}}^2\rangle}{d\omega} = |H(i\omega)|^2 \frac{d\langle U_{\text{in}}^2\rangle}{d\omega}$$

If the input is thermal noise from a resistor R_{source} at temperature T_{abs}, then $\frac{d\langle U_{\text{in}}^2\rangle}{d\omega} = \frac{2R_{\text{source}}k_B T_{\text{abs}}}{\pi}$. So the output noise spectral density is:

$$\frac{d\langle U_{\text{out}}^2\rangle}{d\omega} = |H(i\omega)|^2 \frac{2R_{\text{source}}k_B T_{\text{abs}}}{\pi}$$

This is known as **Nyquist's Theorem** (for noise propagation). The text also shows an equivalent form: $\frac{d\langle U^2\rangle}{d\omega} = \frac{2k_B T_{\text{abs}}}{\pi}\text{Re}(Z_{\text{out}})$, where Z_{out} is the output impedance of the noisy network looking back from the output terminals. This is true if the noise source itself is $R = \text{Re}(Z_{out})$. For an RC filter, $\text{Re}(Z_{out}) = R|H(i\omega)|^2$.

Go to Frame 12.

Frame 12

Input and Output Impedance (Recap)

The text briefly revisits input and output impedance.

- **Open Circuit (OC):** No load connected, current is zero, voltage is maximum.
- **Short Circuit (SC):** Output terminals connected directly, voltage is zero, current is maximum.

A voltage divider consists of two impedances Z_1 and Z_2 in series, with input U_{in} across both, and output U_{out} across Z_2. Input impedance: $Z_{\text{in}} = Z_1 + Z_2$. Output voltage: $U_{\text{out}} = U_{\text{in}}\frac{Z_2}{Z_1+Z_2}$. Output impedance Z_{out} (looking back into the output terminals with U_{in} source shorted) is Z_1 in parallel with Z_2: $\frac{1}{Z_{\text{out}}} = \frac{1}{Z_1} + \frac{1}{Z_2} \implies Z_{\text{out}} = \frac{Z_1 Z_2}{Z_1+Z_2}$.

For an RC low-pass filter (input to R, then C to ground, output across C): $Z_1 = R, Z_2 = 1/(i\omega C)$. $Z_{\text{in}} = R + 1/(i\omega C)$. $Z_{\text{out}} = \frac{R \cdot 1/(i\omega C)}{R+1/(i\omega C)} = \frac{R}{1+i\omega RC}$. So $\text{Re}(Z_{\text{out}}) = \text{Re}\left(\frac{R(1-i\omega RC)}{1+(\omega RC)^2}\right) = \frac{R}{1+(\omega RC)^2} = R|H(i\omega)|^2$. This confirms the equivalent form of Nyquist's theorem stated in the previous frame.

This concludes the section on thermal noise and its characteristics. End of Section.

8 Measurements of Constant Quantities / Statistics

Frame 1

We've encountered some statistical ideas before; this will be a broader refresher, with a focus on dealing with noise in measurements. We'll assume that measurement noise is often distributed according to a Gaussian (Normal) distribution. The model for a measurement z of a true quantity x (where H is a measurement matrix, here scalar $H = 1$):

$$r = (z - Hx) \sim N(a, \sigma)$$

Here, r is the measurement error (noise). Typically, we assume the mean of the noise $a = 0$. We are interested in estimating the parameters a (true mean of noise, ideally 0) and σ (standard deviation of noise). A question arises: is the distribution still truly Gaussian if we "fit" a Gaussian to it, especially as the number of samples $n \to \infty$? (This hints at the robustness of Gaussian assumptions and the Central Limit Theorem).

Go to Frame 2.

Frame 2

Suppose we have a sample of n measurements, $z_1, z_2, \ldots, z_n$. Two fundamental sample statistics are: 1. **Sample Mean ($\bar{z}$):**

$$\bar{z} = \frac{1}{n} \sum_{i=1}^{n} z_i$$

2. **Sample Variance (s^2):**

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (z_i - \bar{z})^2$$

Why do we divide by $n - 1$ for the sample variance instead of n? ______________________________

Go to Frame 3.

Frame 3

We divide by $n - 1$ for the sample variance to make s^2 an **unbiased estimator** of the true population variance σ^2. If we divided by n, the sample variance would, on average, slightly underestimate the true variance. The term $n - 1$ is referred to as the "degrees of freedom."

Now, let's introduce the t-statistic. Go to Frame 4.

Frame 4

T-statistic

We define a new statistic, T, as follows:

$$T = \frac{\bar{z} - a_0}{s/\sqrt{n}}$$

where:

- $\bar{z}$ is the sample mean.
- s is the sample standard deviation ($\sqrt{s^2}$).

- n is the sample size.
- a_0 is a hypothesized true mean of the population from which the sample z_i was drawn. (The OCR uses 'a', here a_0 to avoid confusion with noise mean).

This is called the **T-statistic**. It is used when the true population standard deviation σ is unknown and is estimated by the sample standard deviation s. Notice that T depends on the choice of the hypothesized parameter a_0.

The probability density function (PDF) of the T-statistic follows **Student's t-distribution** with $n-1$ degrees of freedom. The text gives its form as:

$$\frac{dP}{dT} = \frac{1}{\sqrt{n-1}B(\frac{n-1}{2}, \frac{1}{2})} \left(1 + \frac{T^2}{n-1}\right)^{-n/2} = S(n-1)$$

where $B(x, y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt$ is the Beta function. This distribution is bell-shaped like a Gaussian but has "heavier tails," especially for small n. As $n \to \infty$, it approaches a Gaussian distribution. (A sketch of the t-distribution, similar to OCR page 1 bottom, would be here).

Go to Frame 5.

Frame 5

Procedure for Using T-statistic (Hypothesis Testing)

The T-statistic is used for hypothesis testing, typically to test if a sample mean $\bar{z}$ is significantly different from a hypothesized population mean a_0.

1. Acquire a sample of n measurements, z_i.
2. Choose a hypothesized value a_0 for the true mean (based on $N(a_0, \sigma)$ being the assumed underlying distribution).
3. Calculate the sample mean $\bar{z}$ and sample variance s^2 (or standard deviation s).
4. Calculate the T-statistic: $T_{\text{calc}} = (\bar{z} - a_0)/(s/\sqrt{n})$.
5. Choose a **significance level** (or risk level) α. This is the probability of rejecting the hypothesis when it is true (a Type I error). Common values are $\alpha = 0.05$ (5%) or $\alpha = 0.01$ (1%).
6. From tables of the t-distribution (or software), find the critical t-value, t_{crit}, such that $P(|T| > t_{\text{crit}}) = \alpha$ for $n-1$ degrees of freedom.
7. **Decision Rule:** If $|T_{\text{calc}}| > t_{\text{crit}}$, we "reject the hypothesis that the true mean is a_0" at the α significance level. Otherwise, we "fail to reject the hypothesis" (or "cannot reject the hypothesis"). We don't say we "accept" it, just that there's not enough evidence to reject it.

Go to Frame 6.

Frame 6

Confidence Interval

Instead of just rejecting or not rejecting a specific a_0, we can construct a **confidence interval** for the true mean. A $(1-\alpha)$ confidence interval is an interval $[a_<, a_>]$ such that we are $(1-\alpha) \times 100\%$

confident that the true population mean lies within it. It is defined by the range of a_0 values for which we would *not* reject the null hypothesis.

$$P(t_< \leq T \leq t_>) = 1 - \alpha$$

If the t-distribution is symmetric, $t_< = -t_{\text{crit}}$ and $t_> = +t_{\text{crit}}$. So, $-t_{\text{crit}} \leq \frac{\bar{z}-a_0}{s/\sqrt{n}} \leq t_{\text{crit}}$. Rearranging for a_0:

$$\bar{z} - t_{\text{crit}}\frac{s}{\sqrt{n}} \leq a_0 \leq \bar{z} + t_{\text{crit}}\frac{s}{\sqrt{n}}$$

So, $a_< = \bar{z} - t_{\text{crit}}\frac{s}{\sqrt{n}}$ and $a_> = \bar{z} + t_{\text{crit}}\frac{s}{\sqrt{n}}$. We expect the true mean to be outside this interval with probability α.

Go to Frame 7.

Frame 7

Chi-squared (χ^2) Distribution

Suppose, as before, we have a sample z_i from a Gaussian distribution $N(a, \sigma)$. We can construct another statistic called χ^2:

$$\chi^2 = (n-1)\frac{s^2}{\sigma_0^2} = \sum_{i=1}^{n} \frac{(z_i - \bar{z})^2}{\sigma_0^2}$$

(The second form is if $\bar{z}$ is the true mean a; if $\bar{z}$ is the sample mean, the $(n-1)s^2/\sigma_0^2$ form is more standard, where σ_0^2 is a hypothesized population variance). This statistic follows a χ^2 distribution with $n-1$ degrees of freedom. It's used to test hypotheses about the population variance σ^2.

The PDF of the $\chi^2(k)$ distribution (with $k = n-1$ degrees of freedom) is:

$$\frac{dP}{d\chi^2} = \frac{1}{2^{k/2}\Gamma(k/2)}(\chi^2)^{k/2-1}e^{-\chi^2/2}$$

where $\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}dx$ is the Gamma function. (A sketch of the χ^2 distribution, which is skewed to the right, is on OCR page 2, middle). This distribution is also tabulated. We can set rejection regions based on a significance level α. For example, with probability $1-\alpha$, we expect $\chi^2 \in [\chi^2_<, \chi^2_>]$. If our calculated χ^2 falls outside, we might reject the hypothesized σ_0^2 (or conclude that the data is not from $N(a, \sigma_0)$).

Go to Frame 8.

Frame 8

Theorem for Sum of Squares (Cochran's Theorem Idea)

A very important theorem in statistics states: The sum of squares of n independent, standard normally distributed random variables ($\sim N(0,1)$) follows a χ^2 distribution with n degrees of freedom. Symbolically, if $X_i \sim N(0,1)$ are independent, then $\sum_{i=1}^n X_i^2 \sim \chi^2(n)$.

The text provides a "proof" idea: If we have $x_i \sim N(0,1)$, then $\sum x_i^2 = \chi^2$. (This is the definition). If we have $z_i \sim N(\bar{z}, \sigma)$ (this is slightly confusing notation, usually $z_i \sim N(\mu, \sigma)$ and $\bar{z}$ is the sample mean), we first normalize the sample by forming $(z_i - \bar{z})/\sigma$. These are approximately $N(0,1)$ if $\bar{z}$ is close to the true mean μ. Then $\sum_i \frac{(z_i-\bar{z})^2}{\sigma^2}$ is approximately $\chi^2(n-1)$ (losing one degree of freedom because we used the sample mean $\bar{z}$ to estimate the true mean). This leads back to $\chi^2 = (n-1)s^2/\sigma^2 \sim \chi^2(n-1)$.

Go to Frame 9.

Frame 9

Theorem for Ratio of Distributions (Relating to F-distribution or t-distribution)

The text paraphrases a theorem: Let $X \sim N(0,1)$ (a standard normal variable) and $Y \sim \chi^2(n)$ (a chi-squared variable with n degrees of freedom), and assume X and Y are independent. Then the statistic $T = \frac{X}{\sqrt{Y/n}}$ follows a Student's t-distribution with n degrees of freedom, $S(n)$.

The proof sketch involves substituting the definitions: $X = \frac{\bar{z}-a}{\sigma/\sqrt{n}}$ (if σ is known, this is $N(0,1)$) $Y = (n-1)s^2/\sigma^2 \sim \chi^2(n-1)$. Then $T = \frac{(\bar{z}-a)/(\sigma/\sqrt{n})}{\sqrt{((n-1)s^2/\sigma^2)/(n-1)}} = \frac{(\bar{z}-a)/(\sigma/\sqrt{n})}{s/\sigma} = \frac{\bar{z}-a}{s/\sqrt{n}}$. This recovers the definition of the T-statistic with $n-1$ degrees of freedom (not n as the theorem statement suggests, this is a common point of care).

Go to Frame 10.

Frame 10

Goodness-of-Fit Tests

These tests are used to determine how well a proposed probability distribution $dP/d\xi$ matches observed experimental data.

Pearson's χ^2 Test

1. We have a sample z_i. We hypothesize it comes from a distribution $dP/d\xi$.
2. Divide the range of possible values of z into ρ bins (classes or categories). The bins do not need to be of equal width. (See diagram OCR page 4, top).
3. For each bin k:
   - N_k: Observed number of sample points falling into bin k.
   - $P_k = \int_{\xi_k}^{\xi_{k+1}} (dP/d\xi) d\xi$: Theoretical probability that a measurement falls into bin k, based on the hypothesized distribution.
   - NP_k: Expected number of sample points in bin k if the hypothesis is true (where $N = \sum N_k$ is total sample size).
4. The χ^2 statistic is calculated as:

$$\chi^2_{\text{calc}} = \sum_{k=1}^{\rho} \frac{(N_k - NP_k)^2}{NP_k}$$

5. This χ^2_{calc} approximately follows a χ^2 distribution with $df = \rho - 1 - m$ degrees of freedom, where m is the number of parameters of the hypothesized distribution that were estimated from the data itself. (If all parameters are pre-specified, $m = 0$).
6. We compare χ^2_{calc} to a critical value χ^2_{crit} from tables for a chosen significance level α. If $\chi^2_{\text{calc}} > \chi^2_{\text{crit}}$, we reject the hypothesis that the data comes from $dP/d\xi$.

A condition for this test to be valid is that the expected count in each bin, NP_k, should not be too small (e.g., $NP_k \geq 5$). If some bins have small expected counts, they should be merged. The OCR mentions that if the P_k are well chosen, $N_k - NP_k$ will just be statistical noise, and then a Poisson distribution $dP/dN = (\bar{N}^v e^{-\bar{N}})/v!$ applies to N_k, which approaches $N(\bar{N}, \sqrt{\bar{N}})$ for large $\bar{N}$.

Go to Frame 11.

Frame 11

Example: Radioactive Decay

We observe radioactive decay, timing the intervals t_k between successive decays. Suppose we have $N = 100$ such time intervals, and no interval is longer than $1.6s$. We want to test the hypothesis that these times follow an exponential distribution $dP = \frac{1}{\tau_0}e^{-t/\tau_0}dt$ with a hypothesized $\tau_0 = 1s$, at a significance level $\alpha = 5\%$.

How would you set up bins for this χ^2 test? What are N_k and NP_k? (The OCR creates two bins: $0 \le t \le 1.6s$ and $t > 1.6s$). Let Bin 0 be $0 \le t \le 1.6s$. Let Bin 1 be $t > 1.6s$.

- Observed counts: $N_0 = 100$ (all 100 measurements fell in $0 \le t \le 1.6s$). $N_1 = 0$.
- Expected probabilities for $\tau_0 = 1s$: $P_0 = \int_0^{1.6} \frac{1}{1}e^{-t/1}dt = [-e^{-t}]_0^{1.6} = 1 - e^{-1.6} \approx 1 - 0.2019 = 0.7981 \approx 0.80$. $P_1 = \int_{1.6}^{\infty} e^{-t}dt = [-e^{-t}]_{1.6}^{\infty} = e^{-1.6} \approx 0.20$.
- Expected counts for $N = 100$: $NP_0 = 100 \times 0.80 = 80$. $NP_1 = 100 \times 0.20 = 20$.

Now calculate χ^2_{calc}:

$$\chi^2_{\text{calc}} = \frac{(100-80)^2}{80} + \frac{(0-20)^2}{20} = \frac{400}{80} + \frac{400}{20} = 5 + 20 = 25$$

Degrees of freedom: $\rho = 2$ bins. We used a pre-specified $\tau_0 = 1s$, so $m = 0$. $df = 2 - 1 - 0 = 1$. From tables, $\chi^2_{\text{crit}}(df = 1, \alpha = 0.05) = 3.84$. Since $\chi^2_{\text{calc}} = 25 > \chi^2_{\text{crit}} = 3.84$, what do we conclude?

Go to Frame 12.

Frame 12

Since our calculated $\chi^2_{\text{calc}} = 25$ is much larger than the critical value $\chi^2_{\text{crit}}(df = 1, \alpha = 0.05) = 3.84$ (the OCR uses $\chi^2_{crit} = 5.99$ which is for $df = 2$ or a different α, but the principle is the same if $25 > 5.99$), we **reject the null hypothesis** that the decay times follow an exponential distribution with $\tau_0 = 1s$, at the 5% significance level. Our data is significantly different from what this hypothesis would predict.

The OCR notes: "So: we can reject the hypothesis $\tau = 1s$ at a 5% significance level." (The diagram on OCR page 5 shows the χ^2 distribution, the critical value, and the calculated value falling in the rejection region).

Go to Frame 13.

Frame 13

Fisher's Test and Likelihood Function (Parameter Estimation)

Sometimes we want to test a distributional form $dP/dz(q_1, \ldots, q_m)$ where the parameters q_i are not known and need to be estimated from the data for an optimal fit. Suppose we have m measurements, divided into ρ bins $[z_{k-1}, z_k)$. The probability for the k-th bin is $P_k = \int_{z_{k-1}}^{z_k}(dP/dz)dz = P_k(q_i)$. The **Likelihood Function** L^* is the probability of observing the exact histogram (counts N_k in each bin) given the parameters q_i:

$$L^* = \prod_{k=1}^{\rho}[P_k(q_i)]^{N_k}$$

(This is from multinomial distribution, often $\ln L^*$ is used). To find the optimal parameters $\hat{q}_i$ that best fit the data, we maximize L^* (or $\ln L^*$) with respect to each q_i:

$$\frac{\partial(\ln L^*)}{\partial q_i} = 0$$

This gives us estimates $\hat{q}_1, \ldots, \hat{q}_m$. We then use these $\hat{q}_i$ to calculate expected probabilities $N\hat{P}_k$ and perform a χ^2 test as before. The degrees of freedom will now be $df = \rho - 1 - m$ (since m parameters were estimated).

The text gives a simplified likelihood for unbinned data, where z_k are individual measurements:

$$L = \prod_{k=1}^{N} \frac{dP}{dz}(z_k, q_1, \ldots, q_m)$$

For the radioactive decay example, $dP/dt = (1/\tau)e^{-t/\tau}$. $L = \prod_{k=1}^{N}(1/\tau)e^{-t_k/\tau} = (1/\tau)^N e^{-(\sum t_k)/\tau}$. $\ln L = -N \ln \tau - (\sum t_k)/\tau$. To find optimal $\hat{\tau}$: $\frac{\partial \ln L}{\partial \tau} = -N/\tau + (\sum t_k)/\tau^2 = 0$.

$$\implies \hat{\tau} = \frac{1}{N} \sum t_k = \bar{t}$$

The best estimate for τ is the sample mean of the decay times.

Go to Frame 14.

Frame 14

Kolmogorov-Smirnov Test (Cumulative Test)

This is another goodness-of-fit test. Given a sample z_i and a hypothesized (continuous) cumulative distribution function (CDF) $F(z) = \int_{-\infty}^{z}(dP/d\xi)d\xi$. We also form the **empirical cumulative distribution function (ECDF)** from the data, $f(z)$: $f(z) = k(z)/N$, where $k(z)$ is the number of sample points $z_i \leq z$. (The ECDF is a step function that increases by $1/N$ at each observed data point). The Kolmogorov-Smirnov (KS) test statistic D is the maximum absolute difference between $F(z)$ and $f(z)$:

$$D = \sup_{-\infty < z < \infty} |F(z) - f(z)|$$

(A sketch like OCR page 7, top, shows $F(z)$ as a smooth curve and $f(z)$ as a step function, with D being the largest vertical gap). If the hypothesized $F(z)$ describes the data well, D will be small. Kolmogorov showed that for large N:

$$\lim_{N \to \infty} P(D\sqrt{N} < d) = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2 d^2}$$

This distribution is tabulated. We choose a critical value d_c such that $P(D\sqrt{N} < d_c) = 1 - \alpha$. If our calculated $D\sqrt{N} > d_c$, we reject the hypothesis that $F(z)$ is the true CDF with significance α.

This test is often more powerful than χ^2 for continuous distributions, especially with small sample sizes, as it doesn't require binning.

End of Section.

9 Methods of Least Squares and Signal Recovery

Frame 1

We often want to find a model for an unknown function $z(t)$ where the function is assumed to have a specific form, linear in its parameters x_j:

$$z(t) = x_0 f_0(t) + x_1 f_1(t) + \cdots + x_m f_m(t)$$

Here, $f_j(t)$ are known basis functions (e.g., $1, t, t^2, \cos(\omega t)$), and x_j are the unknown parameters we want to find. This problem is "linear in the parameters" even if the functions $f_j(t)$ are non-linear in t.

We can use our knowledge of the Kalman filter. Let's write this problem in matrix form. If we have n measurements $z(t_i)$:

$$\boldsymbol{z} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{r}$$

where: $\boldsymbol{z} = \begin{bmatrix} z(t_1) \\ z(t_2) \\ \vdots \\ z(t_n) \end{bmatrix}$ is the vector of measurements. $\boldsymbol{x} = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_m \end{bmatrix}$ is the vector of parameters we want to estimate ($m+1$ parameters). $\boldsymbol{H}$ is the "design matrix" or "observation matrix" (size $n \times (m+1)$), whose elements depend on the known functions $f_j(t)$ evaluated at measurement times t_i:

$$\boldsymbol{H} = \begin{bmatrix} f_0(t_1) & f_1(t_1) & \dots & f_m(t_1) \\ f_0(t_2) & f_1(t_2) & \dots & f_m(t_2) \\ \vdots & \vdots & \ddots & \vdots \\ f_0(t_n) & f_1(t_n) & \dots & f_m(t_n) \end{bmatrix}$$

$\boldsymbol{r}$ is the vector of measurement errors.

What is the general objective of the "least squares" method? [a] To maximize the number of parameters m. [b] To find parameters $\boldsymbol{x}$ that make $\boldsymbol{H}\boldsymbol{x}$ as close as possible to $\boldsymbol{z}$ by minimizing the sum of squared errors. [c] To ensure the basis functions $f_j(t)$ are orthogonal.

Go to Frame 2.

Frame 2

Your answer was [a — b — c].

The correct answer is [b]. The method of least squares aims to find the parameter vector $\boldsymbol{x}$ such that the model predictions $\boldsymbol{H}\boldsymbol{x}$ are as close as possible to the actual measurements $\boldsymbol{z}$, by minimizing the sum of the squares of the differences (the residuals $\boldsymbol{r} = \boldsymbol{z} - \boldsymbol{H}\boldsymbol{x}$).

We use the Kalman filter framework. Recall the quadratic form $2J(\boldsymbol{x})$ that we want to minimize:

$$2J(\boldsymbol{x}) = (\boldsymbol{z} - \boldsymbol{H}\boldsymbol{x})^T \boldsymbol{R}^{-1} (\boldsymbol{z} - \boldsymbol{H}\boldsymbol{x})$$

where $\boldsymbol{R}$ is the covariance matrix of the measurement noise $\boldsymbol{r}$. To minimize $J(\boldsymbol{x})$, we set the derivative with respect to $\boldsymbol{x}$ to zero:

$$\frac{\partial(2J(\boldsymbol{x}))}{\partial \boldsymbol{x}} = \boldsymbol{0}$$

(This derivative is a vector of partial derivatives with respect to each x_j).

For simplicity, let's assume all measurement points are independent and have the same error variance σ^2. What does the measurement noise covariance matrix $\boldsymbol{R}$ become? (Hint: Identity matrix $\boldsymbol{I}$) ____________________

Go to Frame 3.

Frame 3

If all measurements $z(t_i)$ are independent and have the same error variance σ^2, then the measurement noise covariance matrix $\boldsymbol{R}$ becomes:

$$\boldsymbol{R} = \sigma^2 \boldsymbol{I}$$

where $\boldsymbol{I}$ is the identity matrix. Then $\boldsymbol{R}^{-1} = \frac{1}{\sigma^2}\boldsymbol{I}$.

The Kalman filter update equation for estimating a constant parameter vector $\boldsymbol{x}$ (where the "prediction" $\bar{\boldsymbol{x}}$ and its covariance $\boldsymbol{M}$ are based on prior knowledge) is:

$$\hat{x} = \bar{\boldsymbol{x}} + \boldsymbol{P}\boldsymbol{H}^T\boldsymbol{R}^{-1}(\boldsymbol{z} - \boldsymbol{H}\bar{\boldsymbol{x}})$$

The posterior covariance $\boldsymbol{P}$ is given by:

$$\boldsymbol{P}^{-1} = \boldsymbol{M}^{-1} + \boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{H}$$

If we have no prior information about the parameters $\boldsymbol{x}$ (before filtering/fitting), what can we assume about our initial estimate $\bar{\boldsymbol{x}}$ and its uncertainty (represented by $\boldsymbol{M}^{-1}$)? 1. $\bar{\boldsymbol{x}} =$ ________2. $\boldsymbol{M}^{-1} =$ ____________________

Go to Frame 4.

Frame 4

If we have no prior information about the parameters before the measurements: 1. Our initial estimate $\bar{\boldsymbol{x}} = \boldsymbol{0}$ (or any arbitrary value, as its influence will vanish). 2. Our initial uncertainty is infinite, so the "information" $\boldsymbol{M}^{-1} = \boldsymbol{0}$ (zero matrix).

Substituting these into the Kalman equations: $\boldsymbol{P}^{-1} = \boldsymbol{0} + \boldsymbol{H}^T(\frac{1}{\sigma^2}\boldsymbol{I})\boldsymbol{H} = \frac{1}{\sigma^2}\boldsymbol{H}^T\boldsymbol{H}$. So, the covariance matrix of the estimated parameters is:

$$\boldsymbol{P} = \sigma^2(\boldsymbol{H}^T\boldsymbol{H})^{-1}$$

And the estimate $\hat{x}$ (with $\bar{\boldsymbol{x}} = \boldsymbol{0}$):

$$\hat{x} = \boldsymbol{P}\boldsymbol{H}^T\boldsymbol{R}^{-1}\boldsymbol{z} = \sigma^2(\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T(\frac{1}{\sigma^2}\boldsymbol{I})\boldsymbol{z}$$

$$\implies \hat{x} = (\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T\boldsymbol{z}$$

This is the standard **least squares solution** for the parameter vector $\hat{x}$. So, all we need is the design matrix $\boldsymbol{H}$ (from our model functions $f_j(t)$) and the measurements $\boldsymbol{z}$. We also get the covariance matrix $\boldsymbol{P}$ of our estimated parameters "for free."

The quantity $2J(\hat{x})/\sigma^2$ (where $2J(\hat{x})$ is the minimized sum of squares) is distributed as $\chi^2(n - (m+1))$, where n is the number of data points and $m+1$ is the number of parameters. This allows us to check the goodness-of-fit.

Go to Frame 5.

Frame 5

Example: Linear Fit Suppose we want to fit measurements z_i taken at times t_i to a linear function: $z(t) = x_0 t + x_1$ (so $f_0(t) = t, f_1(t) = 1$). Our parameter vector is $\boldsymbol{x} = \begin{bmatrix} x_0 \\ x_1 \end{bmatrix}$. The design matrix $\boldsymbol{H}$ for n measurements is:

$$\boldsymbol{H} = \begin{bmatrix} t_1 & 1 \\ t_2 & 1 \\ \vdots & \vdots \\ t_n & 1 \end{bmatrix}$$

Then: $\boldsymbol{H}^T\boldsymbol{H} = \begin{bmatrix} \sum t_i^2 & \sum t_i \\ \sum t_i & n \end{bmatrix}$. $(\boldsymbol{H}^T\boldsymbol{H})^{-1} = \frac{1}{n\sum t_i^2 - (\sum t_i)^2} \begin{bmatrix} n & -\sum t_i \\ -\sum t_i & \sum t_i^2 \end{bmatrix}$. $\boldsymbol{H}^T\boldsymbol{z} = \begin{bmatrix} \sum t_i z_i \\ \sum z_i \end{bmatrix}$. Plugging these into $\hat{x} = (\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T\boldsymbol{z}$ gives the familiar formulas for the slope x_0 and intercept x_1 of a linear least squares fit:

$$\hat{x}_0 = \frac{n\sum t_i z_i - (\sum t_i)(\sum z_i)}{n\sum t_i^2 - (\sum t_i)^2}$$

$$\hat{x}_1 = \frac{(\sum z_i)(\sum t_i^2) - (\sum t_i)(\sum t_i z_i)}{n\sum t_i^2 - (\sum t_i)^2}$$

Go to Frame 6.

Frame 6

Peeling Peaks in a Spectrum (Photoemission Example)

Consider photoemission spectroscopy where we might have two overlapping emission peaks. We know the expected shapes (profiles) of individual peaks (e.g., Gaussian or Lorentzian). The kinetic energy E_{kin} of an emitted electron is related to photon energy $h\nu$, binding energy E_V, and work function ϕ: $E_{\text{kin}} = h\nu - E_V - \phi$. Let a be the center energy of a peak and σ its width (for Gaussian). Gaussian profile: $f(E_i) = \exp\left[-\frac{(E_i-a)^2}{2\sigma^2}\right]$ Lorentzian profile: $f(E_i) = \frac{\gamma/2}{(E_i-a)^2+(\gamma/2)^2}$ (The OCR uses $(\gamma/2)^2$ in the numerator by mistake, it should be proportional to γ).

A common, more general shape is the Voigt profile, which is a convolution of a Gaussian and a Lorentzian. If we have a measured spectrum z_i (counts at energy E_i) that is a sum of a background B and two known peak shapes $f_1(i)$ and $f_2(i)$ with unknown amplitudes A_1, A_2:

$$z_i = B + A_1 f_1(i) + A_2 f_2(i) + r_i$$

This is a linear least squares problem if B, A_1, A_2 are the parameters to be found, and $f_1(i), f_2(i)$ (and a constant function for B) are the basis functions. The parameters $\boldsymbol{x} = \begin{bmatrix} B \\ A_1 \\ A_2 \end{bmatrix}$. The design matrix $\boldsymbol{H}$ would have columns $[1, f_1(i), f_2(i)]$.

The solution $\hat{x} = (\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T\boldsymbol{z}$ gives estimates for B, A_1, A_2. The covariance matrix $\boldsymbol{P} = \sigma^2(\boldsymbol{H}^T\boldsymbol{H})^{-1}$ gives the uncertainties in these estimates and their correlations. For example, P_{12} would be the covariance between the estimates of A_1 and A_2.

When are the estimates for A_1 and A_2 likely to be highly correlated (large $|P_{12}|$ or correlation coefficient ρ_{12})? [a] When the peaks f_1 and f_2 are far apart and distinct. [b] When the peaks f_1 and f_2 are very close and overlap significantly.

Go to Frame 7.

Frame 7

Your answer was [a — b].

The correct answer is [b]. If the peaks f_1 and f_2 are very close and overlap significantly (i.e., $f_1 \approx f_2$ in the overlap region), then the columns of $\boldsymbol{H}$ associated with f_1 and f_2 become nearly linearly dependent. This makes the matrix $\boldsymbol{H}^T\boldsymbol{H}$ ill-conditioned (determinant close to zero), leading to large off-diagonal terms in its inverse $\boldsymbol{P}$. The term $\sum f_1(i)f_2(i)$ will be large. If peaks are well-separated, $\sum f_1(i)f_2(i) \approx 0$, making P_{12} small (low correlation). The text states that if peaks are well resolved, the off-diagonal term p_{12} is small: $\sum f_1 f_2 \approx 0$. If peaks are partially merged ($f_1 \approx f_2$), then $\sum f_1^2 \sum f_2^2 - (\sum f_1 f_2)^2 \to 0$, making p_{ij} large, and $\rho_{12} \approx -1$ (negative correlation meaning if A_1 is overestimated, A_2 tends to be underestimated to compensate).

Go to Frame 8.

Frame 8

Measuring System Response to Periodic Perturbation (Phase Detector / Lock-in Amplifier)

(The OCR shows a complex circuit diagram for a phase detector, which is essentially a lock-in amplifier. The principle is to measure how a system responds to a sinusoidal excitation at a specific frequency.)

Let the system be excited sinusoidally. Since the system is linear (assumed), it responds with a harmonic signal at the same frequency, but the amplitude and phase of the response will be different from the input. The amplitude and phase of the response contain information about the quantity being measured. Let the reference signal (related to excitation) be $(Hx)_{\text{ref}} = x_0 \sin(\omega t + \delta)$. We make periodic measurements $z(t_n)$ at times $t_n = n\Delta t$. The design matrix $\boldsymbol{H}$ has elements $H_{ni} = \sin(\omega t_n + \delta_i)$ if we are fitting multiple phase components, or just one column if x_0 is the only parameter and δ is known. The estimate for the amplitude x_0 is:

$$\hat{x}_0 = (\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T\boldsymbol{z} = \frac{\sum z(t_n)\sin(\omega t_n + \delta)}{\sum \sin^2(\omega t_n + \delta)}$$

As $\Delta t \to 0$ and the sum becomes an integral over one period $T_p = 2\pi/\omega$: $\sum \sin^2(\omega t_n + \delta)\Delta t \to \int_0^{T_p} \sin^2(\omega t + \delta)dt = T_p/2$. $\sum z(t_n)\sin(\omega t_n + \delta)\Delta t \to \int_0^{T_p} z(t)\sin(\omega t + \delta)dt$.

$$\hat{x}_0 \approx \frac{2}{T_p}\int_0^{T_p} z(t)\sin(\omega t + \delta)dt$$

This is cross-correlation. It extracts the component of $z(t)$ that is in phase with the reference $\sin(\omega t + \delta)$. Random noise in $z(t)$ tends to average to zero in this integral. The error (variance) of this estimated amplitude is $P = \sigma^2(\boldsymbol{H}^T\boldsymbol{H})^{-1}$. If σ^2 is the variance of $z(t_n)$ per unit time (i.e. $R = \sigma^2\Delta t$ is constant), then $P \approx \frac{R}{T_p/2} = \frac{2R}{T_p}$. The longer we measure (larger T_p), the smaller the error in the amplitude estimate.

Go to Frame 9.

Frame 9

Lock-in Detection (More Detail)

A lock-in amplifier measures a signal that is "locked in" to a specific reference frequency and phase. Suppose our signal of interest is $A(t) = A_\omega \cos(\omega_0 t)$ (this is the reference signal). The input signal from our experiment is $z(t) = x_0 \sin(\omega t + \delta) + \text{noise}(t)$. We want to find x_0 and δ, assuming $\omega = \omega_0$.

We use two reference signals, phase-shifted by 90° (in quadrature): $\text{Ref}_0(t) = R_0 \sin(\omega_0 t)$ $\text{Ref}_{\pi/2}(t) = R_0 \cos(\omega_0 t)$ (which is $R_0 \sin(\omega_0 t + \pi/2)$)

The input $z(t)$ is multiplied by each reference and then low-pass filtered (integrated): $Y_0 = \text{LPF}[z(t) \cdot \text{Ref}_0(t)]$ $Y_{\pi/2} = \text{LPF}[z(t) \cdot \text{Ref}_{\pi/2}(t)]$ $z(t)\text{Ref}_0(t) = x_0 R_0 \sin(\omega_0 t + \delta) \sin(\omega_0 t) = \frac{1}{2} x_0 R_0 [\cos(\delta) - \cos(2\omega_0 t + \delta)]$. After LPF, the $2\omega_0$ term vanishes: $Y_0 = \frac{1}{2} x_0 R_0 T \cos(\delta)$ (where T might be an integration time factor). Similarly: $z(t)\text{Ref}_{\pi/2}(t) = x_0 R_0 \sin(\omega_0 t + \delta) \cos(\omega_0 t) = \frac{1}{2} x_0 R_0 [\sin(\delta) + \sin(2\omega_0 t + \delta)]$. After LPF: $Y_{\pi/2} = \frac{1}{2} x_0 R_0 T \sin(\delta)$.

From these two outputs $(Y_0, Y_{\pi/2})$, how can we find the amplitude x_0 and phase δ of the signal component at ω_0? 1. Amplitude x_0: ______________ 2. Phase δ: ______________

Go to Frame 10.

Frame 10

From $Y_0 = \frac{1}{2} x_0 R_0 T \cos(\delta)$ and $Y_{\pi/2} = \frac{1}{2} x_0 R_0 T \sin(\delta)$: 1. To find amplitude x_0: $Y_0^2 + Y_{\pi/2}^2 = (\frac{1}{2} x_0 R_0 T)^2 (\cos^2 \delta + \sin^2 \delta) = (\frac{1}{2} x_0 R_0 T)^2$. So, $\frac{1}{2} x_0 R_0 T = \sqrt{Y_0^2 + Y_{\pi/2}^2}$.

$$x_0 = \frac{2\sqrt{Y_0^2 + Y_{\pi/2}^2}}{R_0 T}$$

2. To find phase δ:

$$\tan \delta = \frac{\sin \delta}{\cos \delta} = \frac{Y_{\pi/2}}{Y_0} \implies \delta = \arctan\left(\frac{Y_{\pi/2}}{Y_0}\right)$$

This technique allows extraction of very small signals at a specific frequency, even if they are buried in much larger noise, because the noise (if uncorrelated with the reference) will average to zero after multiplication and low-pass filtering.

Adjustable Phase (FOO - Phase-Locked Loop related?) If the reference signal has an adjustable phase φ: $\text{Ref}_\varphi = R \sin(\omega t + \varphi)$. Then the LPF output $Y_{\varphi F} = \frac{1}{2} x_0 R T \cos(\delta - \varphi)$. We can adjust φ until $Y_{\varphi F}$ is maximized. This occurs when $\delta - \varphi = 0 \implies \varphi = \delta$. At this point, $Y_{\varphi F}^{\max} = \frac{1}{2} x_0 R T$, from which x_0 can be read directly.

The text ends with "What are we actually modulating?" It explains that in lock-in detection, we modulate some stimulus $s(t) = \bar{s} + A \sin(\omega t)$ and measure the system's response $V(s)$. The lock-in extracts the component of $V(s)$ that varies at ω, which is proportional to $A \frac{dV}{ds}|_{\bar{s}}$. This is useful for measuring derivatives or small changes.

End of Section.

10 Stability of Feedback Loops

Frame 1

An optimal measurement system often involves a specific structure adapted to the dynamics of the observed system. It includes a variable amplification factor for the innovation, $K(t)$. This is essentially a feedback loop. Recall the innovation term:

$$\text{Innovation} = z - Hx$$

(using x for the estimate $\hat{x}$ here, z for measurement, H for observation matrix/factor). A real measurement system (a sensor of a certain order) often has a constant amplification factor K and is universal, but not strictly optimal (it might have transients, errors, offsets, i.e., systematic errors). We aim to maintain the principle of a feedback loop with appropriate damping for stability.

Go to Frame 2.

Frame 2

Example: Electrical Measurement System (Voltage Follower / Unity Gain Amplifier)

We want a system that is both fast and stable. Let's construct a voltage follower. (A diagram like OCR page 1, top, showing an op-amp with its output connected directly to its inverting input, and the input signal U_{vh} applied to the non-inverting input. Output is U_{izh}). The transfer function for an op-amp (from the previous chapter) can be complex, e.g.:

$$H_{OL}(s) = \frac{\alpha(1 + \frac{2\xi s}{\omega_0})}{1 + \frac{2\xi s}{\omega_0} + \frac{s^2}{\omega_0^2}}$$

(where H_{OL} is the open-loop gain, α might be A_{DC}). Let H_{OZ} denote the open-loop transfer function (often just $H(s)$ or $G(s)$ in control systems). Let H_{ZZ} denote the closed-loop transfer function. For a negative feedback system where the output x is related to an error signal $e = z - x$ (if $H = 1$ in $z - Hx$) by $x = H_{OZ} \cdot e$, or more generally, if $x = H_{OZ}(z_{in} - x_{feedback})$ and $x_{feedback} = x$ for unity feedback: The relationship given is $H_{OZ}(s)(z - x) = x$. This implies $(H_{OZ}(s) + 1)x = H_{OZ}(s)z$. So, the closed-loop transfer function $H_{ZZ}(s) = x/z$ is:

$$H_{ZZ}(s) = \frac{H_{OZ}(s)}{1 + H_{OZ}(s)}$$

What happens to $H_{ZZ}(s)$ if $H_{OZ}(s) \to -1$? ______________________________________

Go to Frame 3.

Frame 3

If the open-loop transfer function $H_{OZ}(s) \to -1$, then the denominator $1 + H_{OZ}(s) \to 0$. This makes the closed-loop gain $H_{ZZ}(s) \to \infty$, which signifies **instability**. The system output can become very large even for small inputs, or it can oscillate.

We can write the frequency response $H_{OZ}(i\omega)$ in polar form:

$$H_{OZ}(i\omega) = |H_{OZ}(i\omega)|e^{i\phi(\omega)}$$

Instability occurs if $H_{OZ}(i\omega) = -1+0i$. This means $|H_{OZ}(i\omega)| = 1$ AND $\phi(\omega) = \pm\pi$ radians ($\pm 180°$). We want to find the angle (phase) ϕ for which this instability occurs. The text uses $\tan\phi = \frac{\text{Im } H(i\omega)}{\text{Re } H(i\omega)}$. If $H(i\omega) = -1$, then Im is 0 and Re is -1. This indeed implies $\phi = \pm\pi$.

Go to Frame 4.

Frame 4

For our unity gain amplifier example (voltage follower), the ideal closed-loop gain H_{ZZ} is 1. The open-loop gain of a typical op-amp A_{DC} is very large (e.g., 10^6) at DC ($\omega \to 0$). $H_{OZ}(i\omega \to 0) \approx A_{DC}$. Then $H_{ZZ}(i\omega \to 0) = \frac{A_{DC}}{1+A_{DC}} \approx 1$. At very high frequencies ($\omega \to \infty$), the op-amp's open-loop gain $|H_{OZ}(i\omega \to \infty)| \to 0$ (due to its internal low-pass filter characteristic). Then $H_{ZZ}(i\omega \to \infty) = \frac{0}{1+0} = 0$.

The stability is checked by examining the behavior of the loop gain $L(i\omega) = H_{OZ}(i\omega)$ (for unity feedback). We are concerned about frequencies where the phase shift $\phi(\omega)$ is $\pm 180°$. At these frequencies, if the magnitude of the loop gain $|H_{OZ}(i\omega)| \geq 1$, the system is unstable. Let $f = |H_{OZ}(i\omega)|$. The text considers a scenario of iterative propagation around the loop. If an initial signal is z: 0th pass (output): $x_0 = H_{OZ} \cdot z = fe^{i\phi}z$. If this is fed back, it becomes $fe^{i\phi}z$. The error signal is $z - fe^{i\phi}z$. The text seems to be looking at the value x returning after multiple passes if the loop were opened and then closed: $x_0 = z$ (input) $x_1 = H_{OZ}x_0 = fz$ (if phase is temporarily ignored for magnitude argument) $x_2 = H_{OZ}(-x_1) = -fx_1 = -f^2z$ (if it's fed back negatively) The text states: $x = f \cdot z$ (effective signal after one pass with magnitude f) $z - (fz)$ (if this is an error signal, the text writes $z + fz = (1+f)z$ which is confusing here) Then $z - (-(1+f)fz) = z + fz + f^2z = z(1+f+f^2)$. This series $z(1+f+f^2+\cdots+f^n)$ sums to $z\frac{1-f^{n+1}}{1-f}$ for $f \neq 1$. This sum converges if $|f| < 1$.

The stability criterion is that at frequencies where the phase shift $\phi(\omega) = \pm 180°$ (so $e^{i\phi} = -1$), we must have the magnitude of the open-loop gain $|H_{OZ}(i\omega)| < 1$ for stability. This is the condition for "damping."

Go to Frame 5.

Frame 5

Nyquist Stability Criterion and Diagram

A powerful tool for analyzing stability is the **Nyquist diagram**. This is a plot of the open-loop frequency response $H_{OZ}(i\omega)$ in the complex plane, as ω varies from 0 to ∞. (A diagram like OCR page 2, bottom, is shown. It plots Im(H_{OZ}) vs. Re(H_{OZ})). The critical point for stability is $(-1, 0i)$ in this complex plane. The Nyquist stability criterion states (simplified version for stable open-loop systems): The closed-loop system is stable if and only if the Nyquist plot of $H_{OZ}(i\omega)$ does **not** encircle the point $(-1, 0i)$.

From the Nyquist plot, we define two important stability margins:

- **Gain Margin (GM):** At the frequency where the phase $\phi(\omega) = -180°$, if the gain magnitude $|H_{OZ}(i\omega)| = A_{180}$, the gain margin is $1/A_{180}$ (or $-20\log_{10} A_{180}$ in dB). We need $A_{180} < 1$ (so GM ¿ 1, or positive in dB) for stability.
- **Phase Margin (PM):** At the frequency where the gain magnitude $|H_{OZ}(i\omega)| = 1$, if the phase is ϕ_1, the phase margin is $180° - |\phi_1|$. We need PM ¿ 0 for stability.

The diagram in OCR page 2 shows a "Nyquist diagram" with a region marked "Vecje dušenje" (More damping) further away from the -1 point, and "Nestabilna točka" (Unstable point) near -1.

The text notes that for an optimal 2nd order system, with $u = \omega/\omega_0$, the damping factor $\xi = 1/\sqrt{2}$. Go to Frame 6.

Frame 6

The text then re-examines the closed-loop magnitude squared for a second-order system (from previous chapter): If $H_{OL}(s) = \frac{A_0}{1+\frac{2\xi s}{\omega_0}+\frac{s^2}{\omega_0^2}}$ (a simple model for op-amp open loop gain, where A_0 is DC gain). The closed-loop transfer function for unity gain feedback ($H_{ZZ} = H_{OL}/(1 + H_{OL})$):

$$\left(\frac{x}{z}\right)_{ZZ} = \frac{H_{OL}}{1 + H_{OL}}$$

If H_{OL} is very large, $(x/z)_{ZZ} \approx 1$. The text then plots $M^2 = |x/z|^2$ where x/z is the closed-loop gain for a system that is NOT necessarily unity feedback, but $H_{CL}(s) = \frac{1+2\xi s/\omega_0}{1+2\xi s/\omega_0+(s/\omega_0)^2}$ (This looks like a specific filter form, possibly a bandpass or a general second order system with a zero). Let $s = i\omega$ and $u = \omega/\omega_0$.

$$M^2(u) = \left|\frac{x}{z}\right|^2 = \frac{1 + (2\xi u)^2}{(1 - u^2)^2 + (2\xi u)^2} = \frac{1 + 4\xi^2 u^2}{(1 - u^2)^2 + 4\xi^2 u^2}$$

The text simplifies this to (perhaps assuming $\xi = 1/\sqrt{2}$ for the "optimal" system):

$$M^2(u) = \frac{1 + 2u^2}{1 + u^4}$$

(This specific simplification holds if $4\xi^2 = 2$, so $\xi^2 = 1/2 \implies \xi = 1/\sqrt{2}$). A plot of M^2 vs $\log u$ is shown (OCR page 3, top). It shows a peak if ξ is small.

To find the maximum of $M^2(u) = \frac{1+2u^2}{1+u^4}$, we set $d(M^2)/du = 0$. This leads to $-1(1+2u^2)4u^3 + (1 + u^4)4u = 0$. $4u[(1 + u^4) - u^2(1 + 2u^2)] = 0$. $1 + u^4 - u^2 - 2u^4 = 0 \implies 1 - u^2 - u^4 = 0$. Let $y = u^2$. $y^2 + y - 1 = 0$. $y = u^2 = \frac{-1\pm\sqrt{1-4(1)(-1)}}{2} = \frac{-1\pm\sqrt{5}}{2}$. Since $u^2 > 0$, we take $u^2 = (\sqrt{5}-1)/2 \approx (2.236-1)/2 = 0.618$. So $u = \sqrt{0.618} \approx 0.786$. At this u, $M^2(u \approx 0.786) \approx 1.62$. This is a peak (resonance) if the system is underdamped. For good damping (to avoid excessive peaking), the condition $M_{OZ} \leq 1.3$ is mentioned (where M_{OZ} seems to be the peak value of M).

Go to Frame 7.

Frame 7

The text then returns to the closed-loop transfer function for unity feedback: $M^2 = \left|\frac{H}{1+H}\right|^2$, where $H = H_{OZ}(i\omega)$. Let $H = \xi_R + i\eta_I$ (using ξ_R for Re(H) and η_I for Im(H) to avoid confusion with damping ratio ξ).

$$M^2 = \frac{\xi_R^2 + \eta_I^2}{(1 + \xi_R)^2 + \eta_I^2}$$

This equation can be rearranged: $M^2((1 + \xi_R)^2 + \eta_I^2) = \xi_R^2 + \eta_I^2$ $M^2(1 + 2\xi_R + \xi_R^2 + \eta_I^2) = \xi_R^2 + \eta_I^2$ If $M^2 \neq 1$: $M^2 + 2M^2\xi_R + M^2\xi_R^2 + M^2\eta_I^2 = \xi_R^2 + \eta_I^2$ $\xi_R^2(M^2 - 1) + 2M^2\xi_R + \eta_I^2(M^2 - 1) + M^2 = 0$ $\xi_R^2+\frac{2M^2}{M^2-1}\xi_R+\eta_I^2+\frac{M^2}{M^2-1} = 0$ Completing the square for ξ_R: $\left(\xi_R + \frac{M^2}{M^2-1}\right)^2-\left(\frac{M^2}{M^2-1}\right)^2+\eta_I^2+\frac{M^2}{M^2-1} = 0$

$$\left(\xi_R + \frac{M^2}{M^2 - 1}\right)^2 + \eta_I^2 = \left(\frac{M^2}{M^2 - 1}\right)^2 - \frac{M^2}{M^2 - 1} = \frac{M^4 - M^2(M^2 - 1)}{(M^2 - 1)^2} = \frac{M^2}{(M^2 - 1)^2}$$

This is the equation of a circle in the H-plane (Re H, Im H): $(\text{Re}H - c_x)^2 + (\text{Im}H - c_y)^2 = r^2$ Center: $c_x = -\frac{M^2}{M^2-1} = \frac{M^2}{1-M^2}$, $c_y = 0$. Radius: $r = \left|\frac{M}{M^2-1}\right| = \frac{M}{|1-M^2|}$. These are called "Constant M circles" on the Nyquist diagram (or Nichols chart). The region $M > 1.3$ is marked as an area to avoid for good stability ("Prepovedano območje" - Forbidden area).

The diagram on OCR page 4 (bottom) shows these M-circles and a typical op-amp open-loop response. The diagram on OCR page 5 shows corresponding amplitude and phase plots (like a Bode plot), indicating "Amplitudna rezerva" (Gain Margin) and "Fazna rezerva" (Phase Margin). A phase margin of about 30° is mentioned at the point where gain crosses 0dB (i.e., $|H| = 1$).

This analysis helps determine if a feedback system will be stable and how much margin for stability it has. End of Section.

11 Measuring Frequency and Time: The Phase-Locked Loop (PLL)

Frame 1

We measure frequency by counting oscillations over a known time period. Since we can measure time very accurately (e.g., with atomic clocks), we can also measure frequencies accurately. It's often advantageous if a measurement of some quantity can be converted into a frequency measurement.

A key tool for working with frequencies and synchronizing oscillators is the **Phase-Locked Loop (PLL)**. What is the main purpose of a PLL, as described in the text? (Hint: It involves an oscillator in an observed system S and a model oscillator in system M). ____________________

Go to Frame 2.

Frame 2

The main purpose of a Phase-Locked Loop (PLL) is to **synchronize an oscillator in a model system M (typically a Voltage-Controlled Oscillator, VCO) with an oscillator (or signal) in an observed system S**. This is done via a feedback loop where the phase of the VCO (θ_M) is compared to the phase of the input signal (θ_S), and the difference is used to regulate the frequency (and thus phase) of the VCO in system M.

A schematic of a PLL includes (see OCR page 1, top diagram):

1. **Phase Detector (PD or FD):** Compares the input signal phase with the VCO's output phase.
2. **Low-Pass Filter (LPF):** Smooths the output of the phase detector.
3. **Voltage-Controlled Oscillator (VCO):** An oscillator whose output frequency is controlled by an input voltage.

The input signal $U_{\text{in}} \sin(\omega_0 t + \theta_S(t))$ goes into the PD. The VCO output $U_0 \cos(\omega_0 t + \theta_M(t))$ also goes into the PD. The PD output U_{FD} goes to the LPF, whose output U_{filter} controls the VCO. The VCO then tries to match its frequency ω_2 (and phase θ_M) to the input.

Go to Frame 3.

Frame 3

Phase Detector (PD)

Let the input signal (from system S, with noisy phase) be: $U_1 = U_{1,0} \sin(\omega_0 t + \theta_1(t))$, where $\theta_1(t) = \theta_S(t) + r(t)$ (true phase + noise). Let the VCO output signal (model system M) be: $U_2 = U_{2,0} \cos(\omega_0 t + \theta_2(t))$, where $\theta_2(t) = \theta_M(t)$ (VCO phase).

The phase detector's output voltage U_{FD} is proportional to the phase difference $\theta_e = \theta_1 - \theta_2$. A common way to implement a phase detector is using a multiplier followed by a low-pass filter. The multiplication gives: $U_1 U_2 = U_{1,0} U_{2,0} \sin(\omega_0 t + \theta_1) \cos(\omega_0 t + \theta_2)$ Using the trigonometric identity $\sin A \cos B = \frac{1}{2}[\sin(A - B) + \sin(A + B)]$: $U_1 U_2 = \frac{U_{1,0} U_{2,0}}{2}[\sin((\omega_0 t + \theta_1) - (\omega_0 t + \theta_2)) + \sin((\omega_0 t + \theta_1) + (\omega_0 t + \theta_2))]$ $U_1 U_2 = \frac{U_{1,0} U_{2,0}}{2}[\sin(\theta_1 - \theta_2) + \sin(2\omega_0 t + \theta_1 + \theta_2)]$

The LPF (either part of the PD or the main loop filter) removes the high-frequency term ($2\omega_0 t$). So the output U_{FD} (averaged over time) is:

$$U_{FD} = k\langle U_1 U_2 \rangle_{\text{low-pass}} = \frac{k U_{1,0} U_{2,0}}{2} \sin(\theta_1 - \theta_2)$$

Let $K_{FD} = \frac{kU_{1,0}U_{2,0}}{2}$. Then:

$$U_{FD} = K_{FD}\sin(\theta_e)$$

If the phase difference θ_e is small, what approximation can we make for $\sin(\theta_e)$? ______________

Go to Frame 4.

Frame 4

If the phase difference $\theta_e = \theta_1 - \theta_2$ is small, we can use the approximation $\sin(\theta_e) \approx \theta_e$. So, for small phase errors (when the PLL is nearly locked):

$$U_{FD} \approx K_{FD}\theta_e$$

The phase detector output is approximately linear with the phase error.

The text mentions two common types of digital phase detectors:

- **Type I: XOR gate.** Output is proportional to the phase difference (represented by the duty cycle of the XOR output). Disadvantage: cannot distinguish between leading and lagging phase, can lock onto harmonics. (Diagram on OCR page 2, top).
- **Type II: Edge-triggered (e.g., phase-frequency detector).** Can distinguish lead/lag. Gives a positive pulse if one signal leads, negative if it lags. (Diagram on OCR page 2, middle).

Go to Frame 5.

Frame 5

Loop Filter (Regulatory Filter)

The output of the phase detector U_{FD} is fed into a loop filter. Purpose:

1. To remove high-frequency components from the PD output (e.g., the $2\omega_0$ term).
2. To provide stability for the loop.
3. To determine the dynamic characteristics of the PLL (e.g., how quickly it locks, how well it tracks changes).

The output of the filter is $U_F = F(s)U_{FD}$, where $F(s)$ is the transfer function of the filter.

Voltage-Controlled Oscillator (VCO) The VCO generates the output signal whose phase $\theta_2(t)$ (or $\theta_M(t)$) we want to match $\theta_1(t)$. Its instantaneous frequency $\omega_2(t)$ is controlled by the filter output voltage $U_F(t)$:

$$\omega_2(t) = \omega_0 + K_0U_F(t)$$

where ω_0 is the VCO's free-running (center) frequency and K_0 is the VCO gain (radians/sec per Volt). The phase $\theta_2(t)$ is the integral of the frequency deviation from the center frequency ω_0: $\dot{\theta}_2(t) = \omega_2(t) - \omega_0 = K_0U_F(t)$. (If $\theta_2(t)$ is defined as total phase $\int \omega_2(t')dt'$, then $\omega_0 t + \theta_2(t) = \int_0^t \omega_2(t')dt' + \theta_2(0)$ from the OCR. The derivative form for the phase deviation is more common for loop analysis). In the Laplace domain (assuming $\theta_2(0) = 0$ for the deviation): $s\theta_2(s) = K_0U_F(s) \implies \frac{\theta_2(s)}{U_F(s)} = \frac{K_0}{s}$. The VCO acts as an integrator for phase.

Go to Frame 6.

Frame 6

PLL Transfer Function

We want to find the overall transfer function of the PLL, typically relating the output phase $\theta_2(s)$ to the input phase $\theta_1(s)$, or the phase error $\theta_e(s)$ to $\theta_1(s)$. The loop components are:

1. Phase Detector: $\theta_e(s) = \theta_1(s) - \theta_2(s)$; $U_{FD}(s) = K_{FD}\theta_e(s)$ (linearized)

2. Filter: $U_F(s) = F(s)U_{FD}(s)$

3. VCO: $\theta_2(s) = \frac{K_0}{s}U_F(s)$

Substitute backwards: $\theta_2(s) = \frac{K_0}{s}F(s)K_{FD}\theta_e(s) = \frac{K_0 K_{FD} F(s)}{s}(\theta_1(s) - \theta_2(s))$. Let $G(s) = \frac{K_0 K_{FD} F(s)}{s}$ be the open-loop transfer function of the PLL. $\theta_2(s) = G(s)(\theta_1(s) - \theta_2(s))$. $\theta_2(s)(1 + G(s)) = G(s)\theta_1(s)$. The closed-loop transfer function $H_{PLL}(s) = \frac{\theta_2(s)}{\theta_1(s)}$ is:

$$H_{PLL}(s) = \frac{G(s)}{1 + G(s)} = \frac{\frac{K_0 K_{FD} F(s)}{s}}{1 + \frac{K_0 K_{FD} F(s)}{s}} = \frac{K_0 K_{FD} F(s)}{s + K_0 K_{FD} F(s)}$$

The transfer function for the phase error $\theta_e(s)/\theta_1(s) = (\theta_1 - \theta_2)/\theta_1 = 1 - H_{PLL}(s)$:

$$\frac{\theta_e(s)}{\theta_1(s)} = \frac{1}{1 + G(s)} = \frac{s}{s + K_0 K_{FD} F(s)}$$

For the PLL to track the input phase well, what do we want θ_e to be ideally? ________________

Go to Frame 7.

Frame 7

Ideally, for good tracking, we want the phase error θ_e to be zero or very small. This means we want $|\theta_e(s)/\theta_1(s)|$ to be small, which implies we want $|1 + G(s)|$ to be large, or $|G(s)|$ to be large, over the frequency range of interest.

The choice of filter $F(s)$ is crucial. A simple RC filter (first order LPF) may not be sufficient due to its long time constant τ (which can lead to poor damping if it's the only pole). The text mentions a "modified RC filter" with two parameters. A common PLL loop filter is a lead-lag filter or an active PI filter to provide a pole and a zero for better control over loop dynamics and stability. The OCR shows a passive lead-lag filter (img. credit: Ana Štuhec): $F(s) = \frac{1+\tau_2 s}{1+(\tau_1+\tau_2)s}$, where $\tau_1 = R_1 C, \tau_2 = R_2 C$.

Substituting this $F(s)$ into $\frac{\theta_e(s)}{\theta_1(s)} = \frac{s}{s+K_0 K_{FD} F(s)}$ gives (after algebra, from OCR):

$$\frac{\theta_e(s)}{\theta_1(s)} = \frac{s(1 + (\tau_1 + \tau_2)s)}{s(1 + (\tau_1 + \tau_2)s) + K_0 K_{FD}(1 + \tau_2 s)}$$

$$= \frac{s + (\tau_1 + \tau_2)s^2}{(\tau_1 + \tau_2)s^2 + s(1 + K_0 K_{FD}\tau_2) + K_0 K_{FD}}$$

This is a second-order system for the phase error. Comparing to the standard second-order form $s^2 + 2\xi\omega_n s + \omega_n^2$ in the denominator (after dividing by $\tau_1 + \tau_2$): Natural frequency: $\omega_n^2 = \frac{K_0 K_{FD}}{\tau_1+\tau_2}$ Damping factor $2\xi\omega_n = \frac{1+K_0 K_{FD}\tau_2}{\tau_1+\tau_2}$. The text notes that by choosing filter components, one can achieve low natural frequency (good noise filtering) and adequate damping (e.g., $\xi = 1/\sqrt{2}$ if $K_0 K_{FD}\tau_2 \gg 1$, leading to $2\xi\omega_n \approx \tau_2\omega_n^2 \implies \xi \approx \tau_2\omega_n/2$).

Go to Frame 8.

Frame 8

PLL Stability and Lock Range

We assumed $\theta_e \ll 1$ for linearization ($\sin\theta_e \approx \theta_e$). If the phase error is large, the dynamics are non-linear:

$$\ddot{\theta}_e + \frac{1 + K_0 K_{FD}\tau_2 \cos\theta_e}{\tau_1 + \tau_2}\dot{\theta}_e + \frac{K_0 K_{FD}}{\tau_1 + \tau_2}\sin\theta_e = \ddot{\theta}_1 + \frac{1}{\tau_1 + \tau_2}\dot{\theta}_1$$

(This is a generalized form if $\cos\theta_e$ term from differentiating $\sin\theta_e K_{FD}$ is kept).

Frequency Step (Lock-in Range / Pull-in Range) If there's a sudden frequency step in the input $\dot{\theta}_1 = \Delta\omega$ (so $\ddot{\theta}_1 = 0$). For a stationary solution, $\dot{\theta}_e = 0, \ddot{\theta}_e = 0$. The equation becomes: $\frac{K_0 K_{FD}}{\tau_1+\tau_2}\sin\theta_e = \frac{\Delta\omega}{\tau_1+\tau_2} \implies \omega_n^2 \sin\theta_e = \frac{\Delta\omega}{\tau_1+\tau_2}$. Since $|\sin\theta_e| \leq 1$, we must have $\omega_n^2 \geq |\frac{\Delta\omega}{\tau_1+\tau_2}|$. The maximum frequency step the PLL can lock onto (pull-in range) is $\Delta\omega_{\max} \approx K_0 K_{FD}$. (This is the DC loop gain).

Frequency Drift (Tracking Range) If the input frequency is drifting, $d(\Delta\omega)/dt = \ddot{\theta}_1 \neq 0$. For steady state tracking, $\dot{\theta}_e = 0, \ddot{\theta}_e = 0$: $\omega_n^2 \sin\theta_e = \ddot{\theta}_1 + \frac{\dot{\theta}_1}{\tau_1+\tau_2}$. The maximum rate of frequency change it can track is $d(\Delta\omega)/dt|_{\max} \approx \omega_n^2$.

How fast does the PLL lock? The "lock-in time" is roughly $2\pi/(\Delta\omega)_0$ where $(\Delta\omega)_0 = K_0 K_{FD}|F(i\Delta\omega)|$.

Go to Frame 9.

Frame 9

Applications of PLL

1. **Frequency Modulation (FM) Demodulation:** In FM radio, information $\Delta\omega(t)$ is encoded in frequency variations around a carrier ω_0: $\omega_{\text{signal}} = \omega_0 + \Delta\omega(t)$. If a PLL is locked to this signal, its VCO control voltage $U_F(t)$ must be such that $K_0 U_F(t) = \Delta\omega(t)$. So, $U_F(t) = \Delta\omega(t)/K_0$. The VCO control voltage $U_F(t)$ is the demodulated audio signal.

2. **Frequency Synthesizer:** A PLL can generate precise multiples or fractions of a reference frequency. If a frequency divider by m (denoted $\div m$) is inserted in the feedback path between the VCO output and the phase detector: The PD now compares θ_1 (reference) with θ_2/m (VCO phase divided by m). In lock, $\theta_1 = \theta_2/m \implies \omega_1 = \omega_2/m$. So, the VCO output frequency $\omega_2 = m \cdot \omega_1$. If m can be programmed (e.g., $m = 2^n$ with flip-flops), we can synthesize many frequencies from one stable reference. (Diagram on OCR page 5, bottom).

Go to Frame 10.

Frame 10

Quartz Clock

Quartz clocks use the extreme stability of quartz crystal oscillations as a time base. The weakness of mechanical oscillators is damping. A piezoelectric crystal (like quartz) deforms when a voltage is applied, and conversely, generates a voltage when deformed. If a quartz crystal "tuning fork" oscillates, it produces an AC voltage at its resonant frequency. This signal can be used as the input to a PLL or an amplifier circuit that uses a magnet to "excite" the fork, providing positive feedback to sustain oscillations (effectively canceling damping). A common crystal frequency is $\nu_0 = 32768$ Hz $= 2^{15}$ Hz. Why is this specific frequency chosen? [a] It's the highest frequency quartz can oscillate at. [b] It's outside the human hearing range and easily divisible by 2. [c] It's the most stable frequency for quartz.

Go to Frame 11.

Frame 11

Your answer was [a — b — c].

The correct answer is [b]. 32768 Hz is ultrasonic (inaudible) and is a power of $2(2^{15})$, making it very easy to divide down to 1 Hz using a chain of 15 binary dividers (flip-flops). This 1 Hz signal can then drive the second hand of a clock or a digital counter.

This shows how precise frequency control and measurement are fundamental to accurate timekeeping. End of Section.

12 Measuring Small Displacements

Frame 1

Resistive Strain Gauges (Uporovni lističi)

The resistance R of a wire is given by:

$$R = \rho \frac{l}{S}$$

where ρ is the resistivity, l is the length, and S is the cross-sectional area.

If we stretch the wire, its length l increases, and its cross-sectional area S decreases. Its resistivity ρ might also change due to stress (piezoresistive effect). The fractional change in resistance dR/R can be related to the fractional changes in these quantities:

$$\frac{dR}{R} = \frac{dl}{l} + \frac{d\rho}{\rho} - \frac{dS}{S}$$

For many materials, when stretched by dl/l: The fractional change in area is $dS/S \approx -2\mu(dl/l)$, where μ is **Poisson's ratio** (typically around 0.3 for metals). The negative sign indicates area decreases when length increases. The fractional change in resistivity (piezoresistive effect) can be related to strain: $d\rho/\rho = G_p(dl/l)$, where G_p is related to the "specific piezoelectricity" or piezoresistive coefficient. (Note: the text uses "piezoelektričnost" but for resistance change, "piezoresistivity" is more accurate).

Substituting these, the OCR text approximates:

$$\frac{dR}{R} \approx \frac{dl}{l}(1 + G_p + 2\mu)$$

The term $(1 + G_p + 2\mu)$ is often called the **gauge factor (GF)**. For simple metals, G_p might be small, and $GF \approx 1 + 2\mu \approx 1.6$. For semiconductor strain gauges, GF can be much larger. The text simplifies this to $\frac{dR}{R} \approx 3\frac{dl}{l}$ as a rough approximation for some materials if $(G_p + 2\mu) \approx 2$.

How is this change in resistance typically measured accurately? [a] Using a simple ohmmeter. [b] Using a Wheatstone bridge circuit. [c] By measuring the current for a fixed voltage.

Go to Frame 2.

Frame 2

Your answer was [a — b — c].

The correct answer is [b]. A **Wheatstone bridge** is commonly used to measure small changes in resistance accurately. (Diagram on OCR page 1, middle, shows a Wheatstone bridge with four resistors R_1, R_2, R_3, R_4, an excitation voltage U_0, and the output voltage ΔU across the bridge). The output voltage of a bridge is:

$$\Delta U = U_0 \left(\frac{R_2}{R_1 + R_2} - \frac{R_4}{R_3 + R_4} \right)$$

(assuming standard bridge configuration). The OCR uses a specific formula $\Delta U = U_0 \frac{R_2R_3 - R_1R_4}{(R_3+R_4)(R_1+R_2)}$. This is the differential voltage.

Suppose we have $R_3 = R_4 = R$, and $R_1 = R$ (reference) and $R_2 = R + \Delta R$ (the strain gauge, where $\Delta R \ll R$). Then $\Delta U = U_0 \frac{(R+\Delta R)R - R \cdot R}{(R+R)(R+R+\Delta R)} \approx U_0 \frac{R\Delta R}{2R \cdot 2R} = U_0 \frac{\Delta R}{4R}$. Using $\Delta R/R \approx 3(\Delta l/l)$ from before:

$$\Delta U \approx U_0 \frac{3}{4} \frac{\Delta l}{l}$$

If N such strain gauges are used effectively (e.g., in a full bridge configuration to maximize sensitivity or account for temperature), the sensitivity can be increased. The text suggests $\Delta U \approx \frac{3}{4} N U_0 \frac{\Delta l}{l}$, meaning the output is proportional to the strain $\Delta l/l$.

Go to Frame 3.

Frame 3

Capacitive Sensor

The capacitance C of a parallel plate capacitor is given by:

$$C = \frac{\epsilon_0 \epsilon_r S_{\text{area}}}{x}$$

where ϵ_0 is the permittivity of free space, ϵ_r is the relative permittivity of the dielectric, S_{area} is the plate area, and x is the distance between the plates. (Diagram on OCR page 1, bottom, shows capacitor plates with variable separation x).

If we change the distance x by a small amount dx, the capacitance changes by dC:

$$dC = \frac{dC}{dx} dx = -\frac{\epsilon_0 \epsilon_r S_{\text{area}}}{x^2} dx = -C_0 \frac{dx}{x_0}$$

where C_0 is the capacitance at separation x_0. The rate of change of capacitance is $\frac{dC}{dt} = -C_0 \frac{1}{x_0} \frac{dx}{dt}$.

Consider a circuit where this variable capacitor $C(t)$ is in series with a resistor R and a DC voltage source U_0. The current is I. The voltage across the capacitor is U_C. (Diagram on OCR page 2, top). In steady state (no change in x), $I = 0$. Charge only flows when C changes or U_C changes. The text assumes small displacements, so $U_i \approx 0$ (voltage across resistor is small compared to U_0 if current is small). This leads to $U_0 = -U_C$ (if U_0 is connected with opposite polarity to how U_C is defined, or it's a misunderstanding of the circuit). This approximation seems problematic for deriving a transfer function.

Let's use the standard AC analysis for a sensor. If U_C is the voltage across the capacitor $C(x(t))$ and U_i is the voltage across R. If U_0 is the total DC bias and we are interested in AC signals due to dx/dt: $I = \frac{d(CU_C)}{dt} = C\frac{dU_C}{dt} + U_C \frac{dC}{dt}$. If $U_C \approx U_0$ (constant DC bias across C when dx/dt is slow), then $dU_C/dt \approx 0$. $I \approx U_0 \frac{dC}{dt} = -U_0 C_0 \frac{1}{x_0} \frac{dx}{dt}$. The output voltage, if taken across R, would be $U_i = IR = -U_0 R \frac{C_0}{x_0} \frac{dx}{dt}$. The output voltage U_i is proportional to the velocity dx/dt. This is a differentiating sensor for displacement.

The text derives a transfer function $H(s) = U_i(s)/x(s)$. Starting from $U_i + RC\dot{U}_i - \frac{C_0}{x_0} \dot{x} U_0 R = 0$ (from OCR, a more complete circuit analysis). With $\tau = RC_0$: $U_i + \tau s U_i = \frac{U_0}{x_0} \tau s x(s)$

$$H(s) = \frac{U_i(s)}{x(s)} = \frac{U_0}{x_0} \frac{\tau s}{1 + \tau s}$$

What kind of filter does this transfer function represent? [a] Low-pass filter [b] High-pass filter [c] Band-pass filter

Go to Frame 4.

Frame 4

Your answer was [a — b — c].

The transfer function $H(s) = \frac{U_0}{x_0} \frac{\tau s}{1+\tau s}$ represents a **high-pass filter** [b]. At low frequencies ($s = i\omega \to 0$), $H(s) \to 0$. The τs in numerator dominates $1 + \tau s$ behavior. At high frequencies ($s = i\omega \to \infty$), $H(s) \to \frac{U_0}{x_0} \frac{\tau s}{\tau s} = \frac{U_0}{x_0}$. For low frequencies, where $1 + \tau s \approx 1$, $H(s) \approx \frac{U_0}{x_0} \tau s$. Since s in

Laplace domain corresponds to differentiation, this sensor acts as a differentiator for low-frequency displacement changes (output proportional to velocity).

The text then shows an improved circuit using an op-amp in an integrator configuration (OCR page 2, bottom) to get a more direct relationship. Here $I = \dot{e} = C_0 \dot{U}_0$ where U_0 is the input voltage (not bias). The output $U_i(t) = -\frac{U_0}{x_0} C_0 R_f \dot{x}(t)$ (if $U_C = U_0$ is constant across the sensing capacitor $C_0(x)$ and R_f is feedback resistor). This setup also acts as a differentiator.

Go to Frame 5.

Frame 5

Piezoelectric Sensor

This sensor is also based on a capacitor, but the capacitor is filled with a **piezoelectric material**. A piezoelectric material generates an electric charge/voltage when mechanically stressed, and conversely, deforms when an electric field is applied. Examples: Quartz (SiO_2), $PbTiO_3$, $SrTiO_3$ (Perovskite structure ABX_3). (Diagram of crystal structure on OCR page 3, middle). In equilibrium, a symmetric crystal may have no net dipole moment. When stressed (e.g., compressed in the z direction), the equilibrium positions of ions (like Ti^{4+} in the diagram) shift, creating a net electric polarization $\vec{P}$. This polarization induces a surface charge $q = \int \vec{P} \cdot d\vec{S}$.

The polarization $\vec{P}$ is related to the mechanical stress tensor $\boldsymbol{T}$ by the piezoelectric tensor $\boldsymbol{d}$:

$$P_i = d_{ijk} T_{jk}$$

(summation over repeated indices implied). T_{jk} is the stress tensor. This is often written in a reduced matrix notation (Voigt notation) because T_{jk} is symmetric (6 independent components): $P_i = d_{im} T_m$ ($m = 1, \ldots, 6$). The piezoelectric tensor d_{im} can have up to 18 independent components. Crystal symmetry reduces this number.

The text notes that the more symmetric the crystal, the fewer distinct d_{ijk} components. If a crystal is deformed along axis j, and plates are on axis i: $e_i = P_i S_{\text{area}} = d_{ij} \sigma_j S_{\text{area}} = d_{ij} F_j$. (charge e_i, stress σ_j, Force F_j). If we assume a simple case where stress is primarily along one axis (e.g., d_{33} is often the largest component for compression perpendicular to main faces): $e = d_{33} F$.

Go to Frame 6.

Frame 6

A piezoelectric crystal acts as both a sensor (generates charge/voltage from force/displacement) and an actuator (deforms under applied voltage). The text considers a setup where the piezo crystal is used as a displacement sensor, possibly connected via a coaxial cable. (Diagram on OCR page 4, top, shows a piezo element, likely compressed by $x_0 + dx$, connected by a coax). The generated charge is $e = d_{33} F$. If force F is due to displacement x via a spring-like behavior ($F = kx$), then $e \propto x$. The change in charge $\dot{e} = d_{33} \dot{F}$. If $F = ES\frac{dx}{l_0}$ (stress $E\cdot$ strain, S=area, l_0=thickness): $\dot{F} = ES\frac{\dot{x}}{l_0}$. So $\dot{e} = d_{33} ES \frac{\dot{x}}{l_0}$. This $\dot{e}$ is a current I_P. This current flows into a circuit consisting of the crystal's own capacitance C_P, cable capacitance C_K, and load resistance R. The voltage measured is U_i. The circuit equation is: $I_P = \dot{e} = (C_P + C_K)\dot{U}_i + U_i/R$. $d_{33} ES \frac{\dot{x}}{l_0} = (C_P + C_K)\dot{U}_i + U_i/R$. Let $\tau = R(C_P + C_K)$. In Laplace domain: $d_{33}\frac{ES}{l_0} s x(s) = (\tau s + 1) U_i(s)$. The transfer function $H(s) = U_i(s)/x(s)$ is:

$$H(s) = \frac{d_{33} ES/l_0}{\tau s + 1} s = K \frac{\tau_p s}{\tau s + 1}$$

(where K and τ_p are constants related to material properties). This is again a high-pass (or band-pass if $\tau_p \neq \tau$) differentiating behavior for low frequencies. The text notes that cable properties vary, so an op-amp buffer/amplifier circuit is often used to standardize the response (see OCR page 4, bottom diagram). This leads to a modified transfer function that is less dependent on the specific cable.

Go to Frame 7.

Frame 7

Inductive Sensor

We measure displacement by arranging for the measurand (the object whose displacement x is measured) to cover or uncover a magnetic material μ_r, which changes the inductance L of a coil. The inductance of a coil with N turns, area S_{area}, length l_{coil} is approximately:

$$L = \mu_0 \mu_r \frac{N^2 S_{\text{area}}}{l_{coil}}$$

The text considers a coil where a part of its core (x) is filled with material μ_r and part ($l_{coil} - x$) is air ($\mu_r = 1$). (Diagram on OCR page 5, middle, shows a coil with a movable magnetic core). The total inductance is $L = L_1 + L_2$: $L_1 = \mu_0 \frac{N^2 S_{\text{area}}}{l_{coil}^2}(l_{coil} - x)$ (air part, assuming $N(l-x)/l$ turns cover air) $L_2 = \mu_0 \mu_r \frac{N^2 S_{\text{area}}}{l_{coil}^2} x$ (magnetic material part, assuming Nx/l turns cover material)

$$L(x) = \mu_0 \frac{N^2 S_{\text{area}}}{l_{coil}^2}(l_{coil} - x + \mu_r x) = \mu_0 \frac{N^2 S_{\text{area}}}{l_{coil}^2}(l_{coil} + (\mu_r - 1)x)$$

The voltage across the inductor U_L is related to the current I by $U_L = -\frac{d\Phi}{dt} = -\frac{d(LI)}{dt}$. If current I is constant (DC): $U_L = -I\frac{dL}{dt} = -I\frac{dL}{dx}\frac{dx}{dt} = -I\frac{dL}{dx}\dot{x}$. From $L(x)$ above: $\frac{dL}{dx} = \mu_0 \frac{N^2 S_{\text{area}}}{l_{coil}^2}(\mu_r - 1)$. So, $U_L \approx -\mu_0 I \frac{N^2 S_{\text{area}}}{l_{coil}^2}(\mu_r - 1)\dot{x}$. The output voltage is proportional to velocity $\dot{x}$.

If the inductor is part of an AC circuit (e.g., with a resistor R and AC source U_0), then $U_0 = IR + L\frac{dI}{dt} + I\frac{dL}{dt}$. For AC current I_s (Laplace domain), the transfer function $H(s) = U_L(s)/x(s)$ (if U_L is output) would involve sLI_s and $s\frac{dL}{dx}xI_s$. The text shows $H(s) = U_L(s)/x(s) \approx -\mu_0 \frac{N^2 S}{l^2}(\mu_r - 1)Is \propto s$. This indicates a differentiating behavior for displacement x.

This type of sensor measures displacement by its effect on inductance. End of Section.