25 October 2013 – 21 October 2018
Note for course 12741: Data Management
by Matteo Pozzi

# Uncertainty, noise, noise propagation and estimation

## 1. Uncertainty

Let us assume we are monitoring the physical quantity $q$. When we measure this quantity, we get value $y$. If the sensor is perfect, $y$ is equal to $q$, and the measure gives us perfect information on $q$. However, sensors are inevitably affected by errors and unpredictable fluctuation in their behavior. We can model the relationship between $q$ and $y$ as:

$$y = q + n \tag{1}$$

where we can call $n$ as "error" or "noise". $n$ models the uncertainty we have in the behavior of the sensor. Of course, if we deterministically knew $n$ we could derive: $q = y - n$. However, usually we do not know $n$, and we treat it as a random variable, i.e. an uncertainty quantity. A classic assumption is to model $n$ as a zero mean Gaussian (aka "normal") variable, with standard deviation $\sigma$. We can call $\sigma$ the "noise level". Consequently, the measure $y$ is distributed as a Gaussian with mean $q$ and standard deviation $\sigma$. As an example, suppose we are measuring a table of length $q = 3\text{m}$, and the standard deviation is $\sigma = 0.5\text{m}$. Figure 1 reports the distribution of $y$.

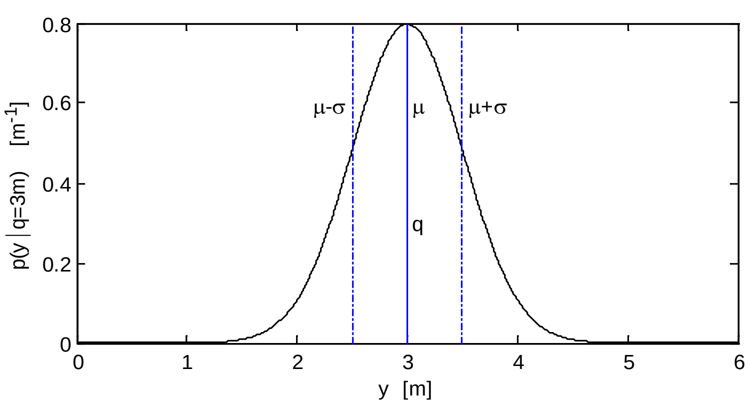


Figure 1: probability distribution of the measure $y$, when the $q = 3\text{m}$.

Formally, the distribution of $n$ is referred to as $p(n)$. If the distribution of variable $x$ is Gaussian, we indicate it as $\mathcal{N}(x, \mu, \sigma^2)$, where $\mu$ is the mean, and $\sigma^2$ the variance. So the distribution of noise $n$ and measure $y$ are defined by:

$$p(n) = \mathcal{N}(n, 0, \sigma^2), \quad p(y|q) = \mathcal{N}(y, q, \sigma^2) \tag{2}$$

We use notation $p(y|q)$ for the distribution of measure $y$. This indicates the conditional probability: the probability of having measure $y$ when the length of the table is $q$. We say that Figure 1 reports the conditional probability of $y$ when $q = 3\mathrm{m}$.

The goal of data processing is to estimate the physical quantity $q$ from the sensor measure $y$. As we see in Figure 1, the basic idea is that: if $\sigma$ is low, we assume the sensor being precise and the measure $y$ being close to the true quantity $q$, while if $\sigma$ is high the sensor is not reliable and measure $y$ may be far from $q$.

Suppose, for example, that the reading of the sensor is $y = 2.7\mathrm{m}$. Because of the uncertainty in the sensor, we cannot infer $q$ without uncertainty. We can treat $q$ as a random variable itself. We refer to its distribution as $p(q|y = 2.7\mathrm{m})$. The simplest model is to conclude that the distribution of $q$ has mean equal to the measured value $y$, and standard deviation equal to that of the sensor:

$$p(q|y) = \mathcal{N}(q, y, \sigma^2) \tag{3}$$

Figure 2 reports the corresponding distribution.

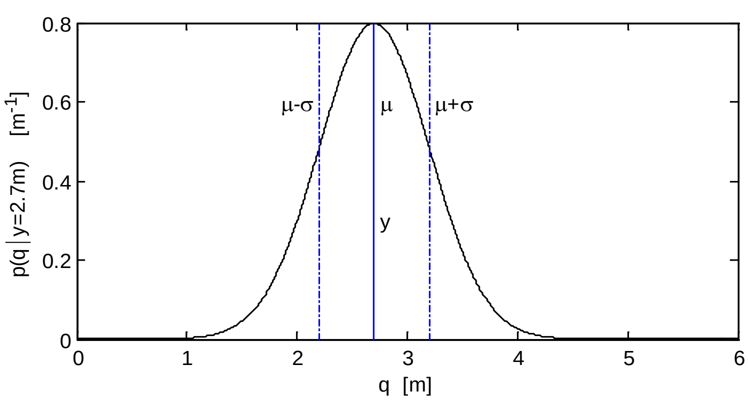


Figure 2: probability distribution of the quantity $q$, when $y = 2.8\mathrm{m}$.

To summarize: when sensor reads value $y$, we model our knowledge of $q$ as a Gaussian distribution, with mean $y$ and standard deviation $\sigma$. In other words, we can use the sensor reading $y$ as our *best guess* on $q$, but we are aware of not being sure about $q$, and the probability distribution model our uncertainty. The smaller the noise level $\sigma$, the more confident we are on $q$.

## 2. Uncertainty propagation

Suppose we combine the measures of many sensors for estimating some physical quantities. We should understand how the uncertainties of the measures affect our estimation.

As an example, consider Figure 3. $y_U$ and $y_L$ indicate the measures of strain in the Upper and Lower edges of a cantilever, respectively. We model each measure following the approach of Eq.1:

$$y_U = q_U + n_U, \quad y_L = q_L + n_L \tag{4}$$

where $q_U$ and $q_L$ is the actual strain the Upper and Lower edges of a cantilever and $n_U$ and $n_L$ are the corresponding errors. The axial strain $q_A$ (the strain at the middle of the cantilever section) is defined as the average between the strains at the edges. The curvature $q_C$ is defined by the difference divided by the cantilever thickness $h$:

$$q_A = \frac{q_U + q_L}{2}, \quad q_C = \frac{q_U - q_L}{h} \tag{5}$$

We can estimate $q_A$ and $q_C$ by applying to the measures the same functions:

$$z_A = \frac{y_U + y_L}{2}, \quad z_C = \frac{y_U - y_L}{h} \tag{6}$$

$z_A$ and $z_C$ can be intended as Virtual Sensors devoted to the measure of the axial strain and curvature respectively.

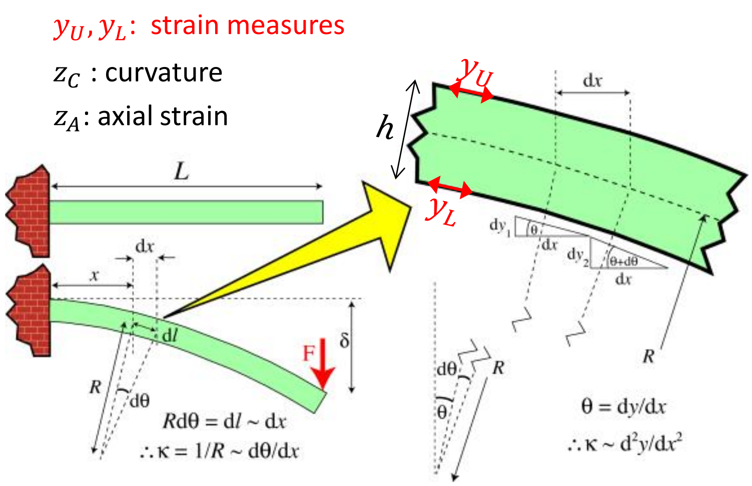


Figure 3: layout of a scheme of 2-sensors set-up on the cantilever
http://www.doitpoms.ac.uk/tlplib/beam_bending/beam_deflection.php]

We can use directly the measurements of the virtual sensors to estimate $q_C$ and $q_A$. The question is: what is the corresponding precision? To answer this question we can combine Eqs.4-6:

$$\begin{cases} z_A = \frac{q_U + q_L}{2} + \frac{n_U + n_L}{2} = q_A + \frac{n_U + n_L}{2} \\ z_C = \frac{q_U - q_L}{h} + \frac{n_U - n_L}{h} = q_C + \frac{n_U - n_L}{h} \end{cases} \tag{7}$$

This shows that $z_A$ is just the actual axial strain $q_A$ plus a noise term depending on the sensors' noise, and so it is for $z_C$. We can simply express this as:

$$z_A = q_A + n_A, \quad z_C = q_C + n_C \tag{8}$$

where we have defined the noise of the virtual sensors as:

$$\begin{cases} n_A = \frac{n_U + n_L}{2} \\ n_C = \frac{n_U - n_L}{h} \end{cases} \tag{9}$$

Again, if we knew actual sensors to be perfect, $n_U$ and $n_L$ would be zero and $n_A$ and $n_C$ would be zero as well. Consequently, virtual sensors' measure would be perfect estimation of $z_A$ and $z_C$. However, generally noises $n_U$ and $n_L$ are random variables, and $n_A$ and $n_C$ turn out to be random variables as well. Let us assume both $n_A$ and $n_C$ are zero-mean Gaussian variables:

$$p(n_U) = \mathcal{N}(n_U, 0, \sigma_U^2), \quad p(n_L) = \mathcal{N}(n_L, 0, \sigma_L^2) \tag{10}$$

where $\sigma_U$ and $\sigma_L$ are the standard deviation of the Upper and Lower sensors respectively. This is still an incomplete representation of the sensors' noise, as we have to model any interaction among the noises at the two edges. The simplest assumption is that or independence, that is related to that of the noise being uncorrelated. In a nutshell, we assume that when one sensor overestimates the physical quantity, by introducing a positive error, we do not tend to think that the other sensor overestimates as well (nor we tend to think that other sensor underestimates): the two noises are independent quantities living in different worlds that do not communicate one to the other. Under these assumptions, the noises of the virtual sensors are still Gaussian, zero-mean:

$$p(n_A) = \mathcal{N}(n_U, 0, \sigma_A^2), \quad p(n_C) = \mathcal{N}(n_L, 0, \sigma_C^2) \tag{11}$$

We have to figure out what these variances are. It turns out that these can be computed as:

$$\sigma_A^2 = \frac{\sigma_U^2 + \sigma_L^2}{4}, \quad \sigma_C^2 = \frac{\sigma_U^2 + \sigma_L^2}{h^2} \tag{12}$$

The general formula is the following. Let us focus our attention on cantilever curvature only. According to Eq.5, we can re-write the curvature as:

$$z_C = \frac{1}{h} y_U - \frac{1}{h} y_L \tag{13}$$

and the corresponding noise, according to Eq.9, can be expressed as:

$$n_C = \frac{1}{h} n_U - \frac{1}{h} n_L \tag{14}$$

or, in other form:

$$n_C = \alpha_{C,U} n_U - \alpha_{L,U} n_L \tag{15}$$

that is the form of a linear function, where $\alpha_{C,U} = \frac{1}{h}$, $\alpha_{L,U} = \frac{1}{h}$.

The general formula is derived in the Appendix at the end of this document. The result is that the variance of the error in the virtual sensor is:

$$\sigma_C^2 = \alpha_{C,U}^2 \sigma_U^2 + \alpha_{L,U}^2 \sigma_L^2 = \frac{\sigma_U^2 + \sigma_L^2}{h^2} \tag{16}$$

For example, if the standard deviations are $\sigma_U = 150\mu\varepsilon$ and $\sigma_L = 150\mu\varepsilon$, where $\mu\varepsilon$ reads "microstrain", is equal to $10^{-6}$ and is the usual unit for expressing strain, and $h = 0.1\mathrm{m}$, we derive from Eq.16 that $\sigma_C^2 = 4.5 \cdot 10^6 \frac{\mu\varepsilon^2}{\mathrm{m}^2}$, and hence that $\sigma_C = 2121 \frac{\mu\varepsilon}{\mathrm{m}}$. Similarly, we can derive the variance and standard deviation of the virtual sensor measuring the axial strain as:

$$\sigma_A^2 = \frac{\sigma_U^2 + \sigma_L^2}{4} = 11250\mu\varepsilon^2 \quad \sigma_A = 106\mu\varepsilon \tag{17}$$

It should be noted that, as we assume $\sigma_U = \sigma_L$, it follows that $\sigma_A^2 = \frac{\sigma_U^2}{2}$ and $\sigma_A = \frac{\sigma_U}{\sqrt{2}}$.

3. Noise for one sensors

A sensor records quantity $q$, so that $y_i$ indicates the measure at time $t_i$, affected by noise $n_i$. Eq. 1 can be re-written as:

$$y_i = q_i + n_i \tag{18}$$

where $q_i = q(t_i)$. When we process the sensor data, we have to assume a model for the noise. The simplest assumption is that the noise for each time is zero mean and with variance $\sigma^2$:

$$\forall i \quad p(n_i) = \mathcal{N}(n_i, 0, \sigma^2) \tag{19}$$

According to this equation, the noise level is constant, as variance $\sigma^2$ does not depend by time. However, to model completely the noise, we also have to assume if and how noise at one time is related to that at another. The simplest assumption is that the noise for each time is independent to that at another:

$$\forall i,j \quad n_i \perp n_j \tag{20}$$

This is the model for what is called the white noise.

Figure 4 shows an example of noisy data. Blue line reports function $q(t)$, while black dots shows the measures, generated by adding white noise, as in the assumptions of Eqs.19-20.

In real life, an agent has access only to the measures $y$s, and she has to estimate $q(t)$ from them. The simplest approach would be to estimate quantity $q_i$, at time $t_i$ using only measure $y_i$, collected at the same time. Following Eq.3, we get:

$$p(q_i|y_i) = \mathcal{N}(q_i, y_i, \sigma^2) \tag{21}$$

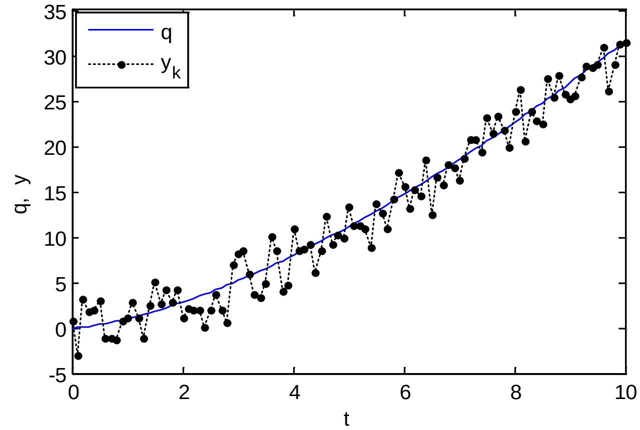


Figure 4: example of measures affected by white-noise

More often, we make some additional assumption on the time variation of $q(t)$. For example, if we assume that $q(t)$ varies very slowly, we can use measures collected before and after $t_i$ to estimate $q_i$, as we usually do in regression.

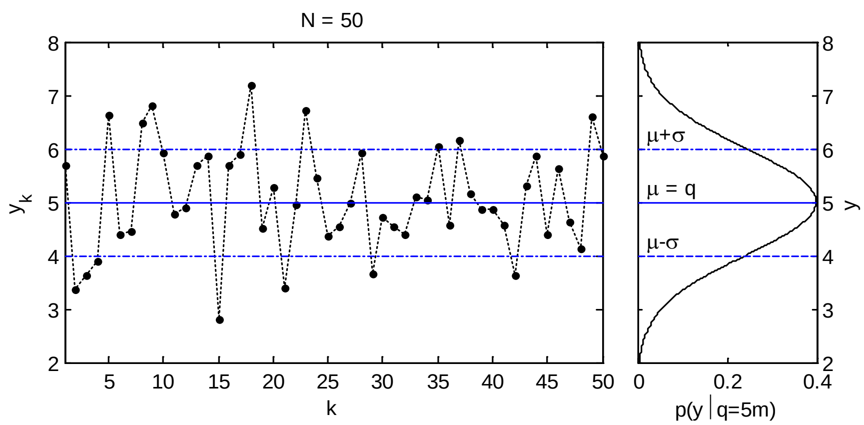


Figure 5: example of measures of a constant, affected by white-noise

## 4. Estimating a constant physical quantities and noise level

Suppose we are measuring $N$ times a constant quantity $q$, and that the measures are affected by white noise, with variance $\sigma^2$. For example, $q = 5\mathrm{m}$ is the length of a table and $\sigma = 1\mathrm{m}$ the standard deviation of a (very bad) sensor. Figure 5 reports an example for $N = 50$ measures. Measures can be intended as samples drawn from the distribution $p(y|q) = \mathcal{N}(y, q, \sigma^2)$, as in Eq.2.

Given all this set of the measures, what if the best estimate of $q$? It is the sample mean $\bar{y}$, defined as the arithmetical mean of all measures:

$$\bar{y} = \frac{1}{N}\sum_{k=1}^{N} y_i \tag{22}$$

The sample mean $\bar{y}$ can be intended as the sum of $q$ plus a small noise, which we expect to be much smaller than that affecting the measures.

Formally, we can define the distribution of $\bar{y}$ as:

$$p(\bar{y}|q) = \mathcal{N}(\bar{y}, q, \bar{\sigma}^2) \tag{23}$$

where the variance $\bar{\sigma}^2$ is related to that of the raw measures as:

$$\bar{\sigma}^2 = \frac{\sigma^2}{N} \tag{24}$$

For example, if $N = 100$ then $\bar{\sigma} = \sigma/10 = 0.1\text{m}$. Figure 6 shows how the distribution of the sample mean is affected by the number of measures.

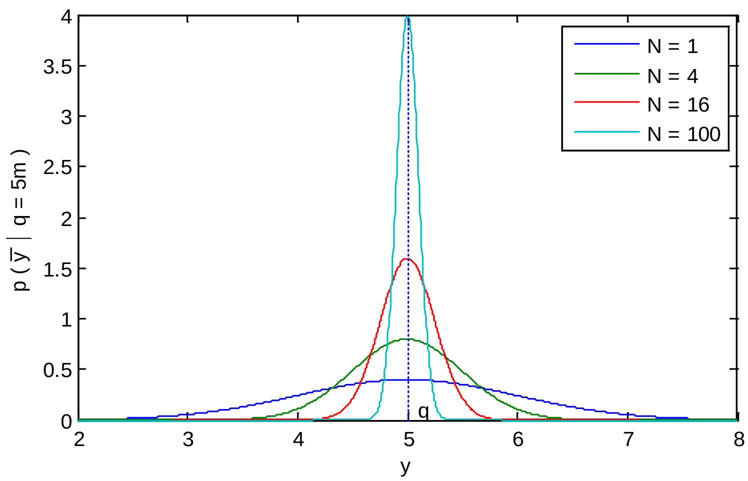


Figure 6: distribution of the sample mean, for an increasing number of measures.

Eq.24 can be derived by the equations for error propagation, as reported in the Appendix. The basic idea behind this result is the following. With only one measure, the error can be large, either positive or negative. With many measures, some will be affected by positive errors, some by negative errors, and these errors will be cancels out by the average.

The previous equations allow us to define the error level in the estimation, when we fix the error level $\sigma$ of the measures. However, in many applications we do not know what the noise level is, and we want to estimate it from the measures. Intuitively, this has to be possible. Consider again Figure 5. Looking at the measures, it is possible to estimate not only the mean value, but also how disperse the data are scattered around the mean. To do so, suppose for example to know the value of $q$ which is the mean $\mu$ of the distribution: we could estimate the noise variance as:

$$\sigma^2 \cong \frac{1}{N}\sum_{k=1}^{N}(y_k - \mu)^2 \tag{25}$$

However, we do not know the value of $q$, but we can use the sample mean $\bar{y}$, obtained by Eq.22, to estimate it. We obtain sample variance $s^2$:

$$s^2 = \frac{1}{N-1}\sum_{k=1}^{N}(y_k - \bar{y})^2 \tag{26}$$

Note that the value $N-1$ in the denominator (which increases the estimate, respect to the use of $N$) compensates for the approximate estimation of $\bar{y}$. Figure 7 reports again the graph of Figure 5, but also shows in red the sample mean $\bar{y}$ and the sample standard deviation $s$. After 50 measures, $\bar{y}$ is close to $q$, and $s$ is a close estimate of $\sigma$.

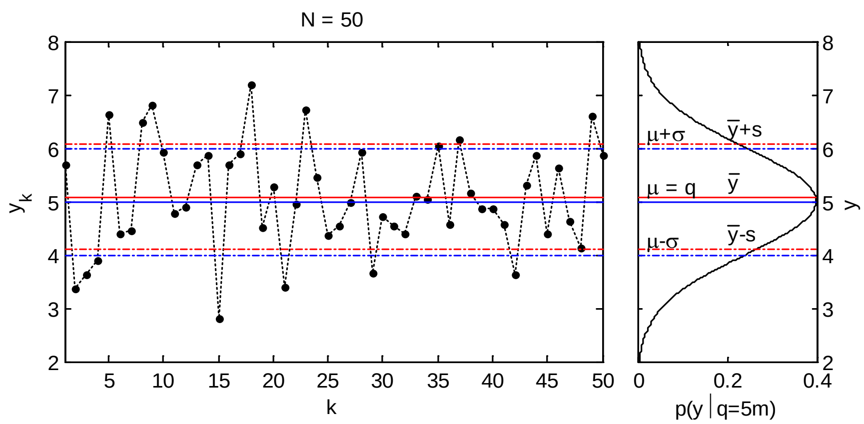


Figure 7: estimation of mean and standard deviation after 50 measures.

## 5. Noise propagation for many sensors

To conclude this note, let us consider the case of more than one sensor. As in Part 2, we focus on the application with 2 strain sensors on a cantilever beam. Suppose the two strain gauges are interrogated with a sampling period of 1 minute. The blue line reports strain $q_U(t)$ at the upper edge, and red line reports strain $q_L(t)$ at the lower edge. Measures of two sensors are defined by:

$$\begin{cases} y_{1,k} = q_{U,i} + n_{1,k} \\ y_{2,k} = q_{L,i} + n_{2,k} \end{cases} \tag{27}$$

and noise are modeled as zero mean, independent for each time and sensor. As in Part 1, we can assume $\sigma_U = 150\mu\varepsilon$ and $\sigma_L = 150\mu\varepsilon$ for the standard deviation of sensor 1 and 2 respectively.

Simulated measures are reported in Figure 8.

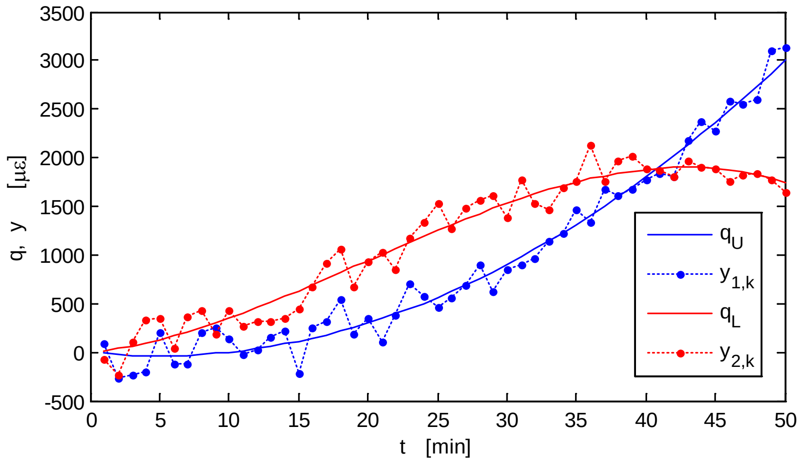


Figure 8: measures of strain gauges on the cantilever.

Figure 9 reports $q_A(t)$ and $q_C(t)$, as derived by Eq.5, and the outcomes of the virtual sensors $z_A$ and $z_C$, as derived by Eq.6.

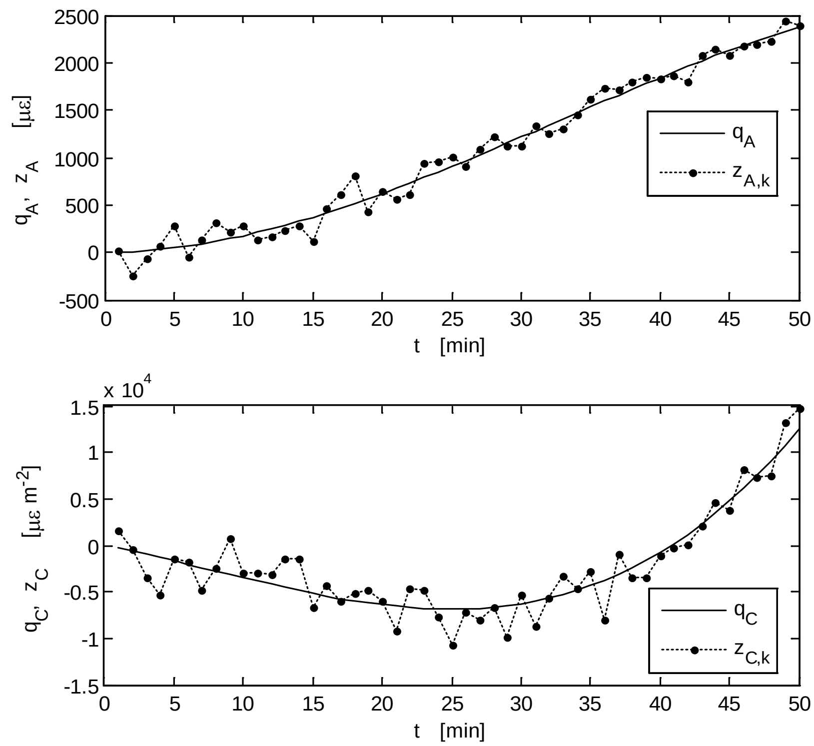


Figure 9: actual axial strain and corresponding outcomes of the virtual sensor (upper graph), and curvature and corresponding outcomes of the virtual sensor (lower graph).

As expressed by Eq.8, it is apparent that the virtual sensors follow axial strain and curvature, and they are affected by a zero mean noise with variance $\sigma_A = 106\mu\varepsilon$ and $\sigma_C = 2121\,\frac{\mu\varepsilon}{\mathrm{m}}$, as computed in Part 2.

Appendix: error propagation

*with 2 variables*

Let us start with two random variables: $x$ and $y$. Mean value for these quantities are given by:

$$\begin{cases} \mu_x = \mathbb{E}[x] = \int x\, p(x) dx \\ \mu_y = \mathbb{E}[y] = \int y\, p(y) dy \end{cases} \tag{A1}$$

where $\mathbb{E}$ is the statistical expectation, and variance by:

$$\begin{cases} \sigma_x^2 = \mathbb{V}[x] = \mathbb{E}[(x-\mu_x)^2] \ = \int (x-\mu_x)^2 p(x) dx \\ \sigma_y^2 = \mathbb{V}[y] = \mathbb{E}\left[\left(y-\mu_y\right)^2\right] = \int \left(y-\mu_y\right)^2 p(y) dy \end{cases} \tag{A2}$$

where $\mathbb{V}$ indicates variance.

Let us $z$ define a linear function of $x$ and $y$:

$$z = f(x,y) = \alpha_x x + \alpha_y y + c \tag{A3}$$

where $\alpha_x$ and $\alpha_y$ and $c$ are known values. The mean value of $z$ can be derived by making use of the linearity of the expectation. Using Eq.3, we obtain:

$$\mu_z = \mathbb{E}[z] = \mathbb{E}\left[\alpha_x x + \alpha_y y + c\right] = \alpha_x \mathbb{E}[x] + \alpha_y \mathbb{E}[y] + c\mathbb{E}[1] = \alpha_x \mu_x + \alpha_y \mu_y + c \tag{A4}$$

Note that, according to Eq.4, $\mu_z = f\left(\mu_x, \mu_y\right)$. In other words, we can obtain the mean of $z$ by applying to the mean values ($\mu_x$ and $\mu_y$) the same linear function defining $z$ in Eq.A3.

The variance of $z$ is defined as:

$$\sigma_z^2 = \mathbb{V}[z] = \mathbb{E}[(z-\mu_z)^2] \tag{A5}$$

By using Eqs.3-4, and using again the linearity of expectation, we obtain:

$$\begin{aligned} \sigma_z^2 &= \mathbb{E}\left[\left(\alpha_x x + \alpha_y y + c - \alpha_x \mu_x - \alpha_y \mu_y - c\right)^2\right] \\ &= \mathbb{E}\left[\alpha_x^2 (x-\mu_x)^2 + \alpha_y^2 \left(y-\mu_y\right)^2\right] \\ &= \alpha_x^2 \mathbb{E}[(x-\mu_x)^2] + \alpha_y^2 \mathbb{E}[(x-\mu_x)^2] = \alpha_x^2 \sigma_x^2 + \alpha_y^2 \sigma_y^2 \end{aligned} \tag{A6}$$

The corresponding relation for standard deviation is

$$\sigma_z = \sqrt{\alpha_x^2 \sigma_x^2 + \alpha_y^2 \sigma_y^2} \tag{A7}$$

Examples:

a) if $z = 2x$, then $\alpha_x = 2, \alpha_y = 0, c = 0,\ \mu_z = 2\mu_x\,, \sigma_z^2 = 4\sigma_x^2\,,\ \sigma_z = 2\sigma_x$ .
b) if $z = 2x + 3$, then $\alpha_x = 2, \alpha_y = 0, c = 3,\ \mu_z = 2\mu_x + 3\,, \sigma_z^2 = 4\sigma_x^2\,,\ \sigma_z = 2\sigma_x$ .

c) if $z = x + y$, then $\alpha_x = 1, \alpha_y = 1, c = 0,\ \mu_z = \mu_x + \mu_y\ , \sigma_z^2 = \sigma_x^2 + \sigma_y^2\ ,\ \sigma_z = \sqrt{\sigma_x^2 + \sigma_y^2}\ .$

d) if $z = x - y$, then $\alpha_x = 1, \alpha_y = -1, c = 0,\ \mu_z = \mu_x - \mu_y\ , \sigma_z^2 = \sigma_x^2 + \sigma_y^2\ ,\ \sigma_z = \sqrt{\sigma_x^2 + \sigma_y^2}\ .$

e) if $z = 3x - 2y + 5$, then $\alpha_x = 3, \alpha_y = -2, c = 5,\ \mu_z = 3\mu_x - 2\mu_y + 5\ , \sigma_z^2 = 9\sigma_x^2 + 4\sigma_y^2\ , \sigma_z = \sqrt{9\sigma_x^2 + 4\sigma_y^2}\ .$

Example (a) shows that, when we scale a variable by factor $\alpha$, e.g. converting kilometers in meters, the standard deviation is scaled by the same factor, but the variance is scaled by the square of that factor (as clear from Eq.6).

Example (b) shows that adding a constant affects the mean value, but not the variance, as the constant does not add any uncertainty to $z$.

It is interesting to compare case (c) to (a). Suppose, in case (c), $\sigma_y^2 = \sigma_x^2$. One may think that this is equivalent to case (a), so that $\sigma_z = 2\sigma_x$. This is not the case: as we can easily check, $\sigma_z = \sqrt{2}\sigma_x = 1.41\sigma_x < 2\sigma_x$. This is due to the fact that we assume errors on $x$ and $y$ to be independent, so there is chance that they cancels out, while in (a) the error on $x$ is amplified by factor 2.

Consider case (d), again when $\sigma_y^2 = \sigma_x^2$. One may think that the variance $\sigma_z^2$ should be zero, as $z$ is defined as the difference between $x$ and $y$, and the errors may cancel out. Actually, the errors can be symmetrically positive or negative, and the variance is exact the same as in case (c).

With *$m$ variables*

We now consider a set of $m$, labeled $x_1, x_2, \dots, x_m$. Mean value and variance for each variable are defined as:

$$\forall k = 1,2, \dots, m \quad \mu_k = \mathbb{E}[x_k] \quad \sigma_k^2 = \mathbb{V}[x_k] \tag{A8}$$

Let us define $z$ as a linear combination of these variables:

$$z = \sum_{k=1}^{m} \alpha_k x_k + c \tag{A9}$$

Mean value and variance of $z$ are derived as:

$$\mu_z = \sum_{k=1}^{m} \alpha_k \mu_k + c \qquad \sigma_z^2 = \sum_{k=1}^{m} \alpha_k^2 \sigma_k^2 \tag{A10}$$

as can be proved by the same analysis leading to Eqs. A4-5.

*Application to virtual sensors and their noise*

Suppose we have $m$ sensors and $y_{k,i}$ indicates the output of sensor $k$ at time $t_i$, measuring quantity $q_{k,i}$. We assume the measures to be composed by actual quantity and noise $n_{k,i}$:

$$y_{k,i} = q_{k,i} + n_{k,i} \tag{A10}$$

We define a virtual sensor by linearly combining the measures of all these sensors. $z_i$ indicates the reading of the virtual sensor at time $t_i$, and it is defines as:

$$z_i = \sum_{k=1}^{m} \alpha_k y_{k,i} + c \tag{A11}$$

where coefficients $\alpha_k$ and $c$ are assume to be known. By using Eq.A10 in Eq.A11, we get:

$$z_i = \sum_{k=1}^{m} \alpha_k q_{k,i} + c + \sum_{k=1}^{m} \alpha_k n_{k,i} \tag{A12}$$

that we can re-write as:

$$z_i = q_{z,i} + n_{z,i} \tag{A13}$$

In short, we can claim that, at time $t_i$, the virtual sensor measures quantity $q_{z,i}$, with noise $n_{z,i}$, defined as:

$$\begin{cases} q_{z,i} = \sum_{k=1}^{m} \alpha_k q_{k,i} + c \\ n_{z,i} = \sum_{k=1}^{m} \alpha_k n_{k,i} \end{cases} \tag{A14}$$

As appears from Eq.A14, the quantity is just a linear combination of those measured by the $m$ actual sensors, according to the same function relating $y$s to $z$, according to Eq.A11. Now, let us assume that the noise affecting each actual sensors is zero mean, with variance defined by $\sigma_k^2$:

$$\forall k, i: \;\; \mathbb{E}\left[n_{k,i}\right] = 0 \;\;\; \sigma_k^2 = \mathbb{V}\left[n_{k,i}\right] \tag{A15}$$

In other words, we assume the variance to be constant, the same for all times. By using Eq.A10, we conclude that:

$$\mathbb{E}\left[n_{z,i}\right] = 0 \;\;\; \sigma_z^2 = \mathbb{V}\left[n_{z,i}\right] = \sum_{k=1}^{m} \alpha_k^2 \sigma_k^2 \tag{A16}$$

So we can read Eq.A13 in this way: the virtual sensor records quantity $q_z$, and it is affected by a zero-mean error, with variance $\sum_{k=1}^{m} \alpha_k^2 \sigma_k^2$.

*Application to normal variables*

Previous equations are general, and can be applied to every type of random variable, including normal variables. That case is particular easy, as mean and variance are the parameters of the distribution.