Using the "TransferFunction" Shiny App

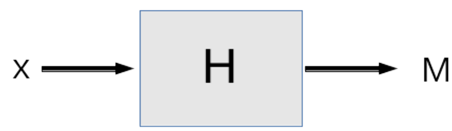
The information in the next section is general information describing sensor response to a varying input. The diagrams there are particularly relevant to the calculation of the response in this ShinyApp. The following provides some guidance on using the app itself.

The plots show the assumed input as a red line. The calculated response uses, for first-order response, a simple exponential with time constant $\tau = 1$. This response in each case is displayed as the blue line described as "M1" in the legends. For second-order response for the system displayed, which is a damped harmonic oscillator, the resulting lines are labeled "M2" and plotted as green lines. This second-order sensor is characterized by two parameters, taken to be omega (ω) and gamma γ, and they can be controlled by two sliders in the sidebar panel. The natural undamped oscillation angular frequency if specified by ω and the ratio of the amount of damping to the critical amount is specified by γ, so that $\gamma = 1$ specifies critical damping and smaller values can oscillate and overshoot while larger values respond slowly. The slider controlling the frequency applies only to repetitive waveforms like the sine or square waves. All these sliders are set using base-10 logarithms to provide a better range than is possible with a linear slider, and the values actually used are then included in the titles of the plots.

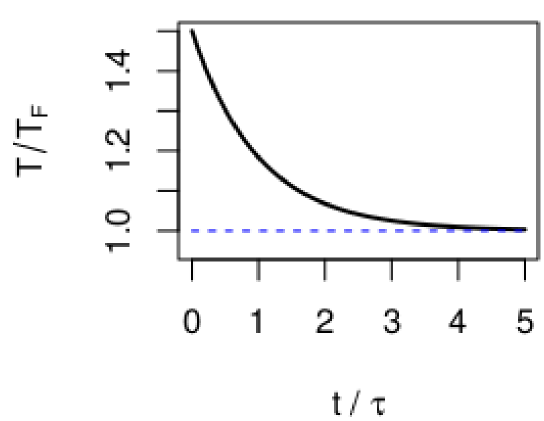
The remainder of this document contains general information on the dynamic response of sensors and how that response can be represented by transfer-function diagrams. The R code used to generate the response is a very simple Euler integration based on the procedure suggested by the diagrams. The results have been checked against analytical solutions in some cases and match those solutions well, but the intent of these displays is to illustrate general features of the response to various input signals. These results are not reliable enough for quantitative use. The numerical integrations could easily be improved using, e.g., Runge-Kutta methods.

Transfer functions and the transfer-function diagram

A sensor as used in a measuring instrument produces some signal M that is related to a measurand x. *Calibration* of a sensor consists of determining the static relationship between M and x, often plotted as a transfer curve showing the relationship, as discussed in he section on static response. A sensor can be depicted generically using a diagram like the following, where $H_I(x)$ will be called the transfer function (although in much of the instrumentation literature that term is reserved for the Laplace transform of the transform as used here, so the subscript I here distinguishes this transfer function from that Laplace transform).[1]



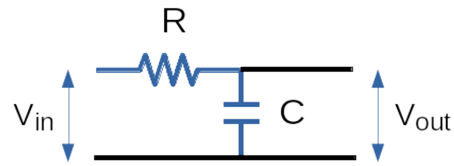
The transfer curve characterizes this relationship (M given x) under static conditions, but we are often interested in applications where the measurand x varies in time. The simple-box description of the sensor still applies, but now the output M will depend not only on the present value of x but also on its past history. As an example, consider a liquid-in-glass thermometer immersed at time $t_0 = 0$ into a bath of fluid having temperature T_F. If the thermometer reads T_0 before immersion, it will start to move toward T_F after immersion and will display a temperature $T(t)$ that moves toward T_F over some period of time. The rate at which heat is transferred to the thermometer depends on the temperature difference $(T(t) - T_F)$, often in a linear relationship, so the displayed temperature will move faster at first and gradually move more slowly toward T_F as the temperature difference becomes small.



A linear relationship between the derivative and the temperature difference would be represented by an equation like $\dot{T} = (T_F - T(t))/\tau$ where τ is a constant that determines the rate at which the sensor responds to its environment and the dot over $\dot{T}$ indicates a time derivative. This equation has an exponential solution, such that $T(t)$ approaches T_F but never reaches that value, as shown in the figure to the right. This is characteristic behavior for many sensors and electronic components of sensors.

[1] The Laplace transform is not treated in this module, but readers interested in more details about transfer functions will want to learn this approach because it is a very powerful method for solving differential equations with specified boundary conditions.

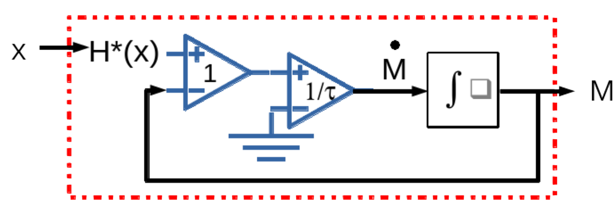
Another example of first-order response is an "RC" circuit, consisting of a resistor R and a capacitor C as shown in the figure to the right. The same equations apply to this circuit, with time constant $\tau = \mathrm{RC}$. Functionally, a capacitor can be regarded as an integrator because the voltage across it equals the time integral of the current (I(t)) that passes through it divided by the capacitance. The same current passes through the resistor and produces a voltage across it of I(t)R, so equating the current passing through the resistor and through the capacitor gives $\dot{V}_{out} = (V_{in} - V_{out})/(CR)$ which has the same form as the equation for the derivative of temperature, $dT(t)/dt$, in the precedng example if the time constant τ is equal to RC. All electronic components have some resistance and some capacitance associated with their output impedance, so this is a common factor influencing dynamic response. Generalizing, any sensor exhibiting first-order dynamic response can be characterized by



$$\frac{dM(t)}{dt} = \frac{[H_I^*(x(t)) - M(t)]}{\tau} \tag{1}$$

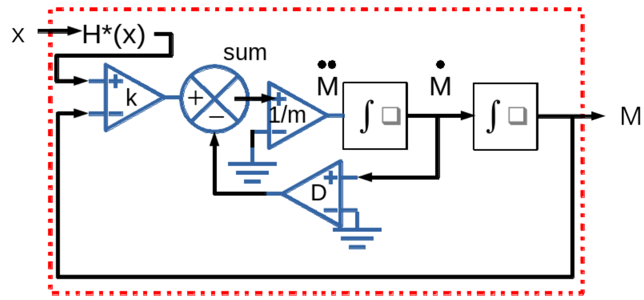
where $H^*(x)$ represents the static transfer function or the transfer curve, giving the output value that would be produced by a steady measurand with the value x.

The diagram at the right, based on this equation, is a functional depiction of a first-order dynamic system having time constant τ. It shows how the output M from a sensor having first-order dynamic response depends on the measurand x. The red box is the complete instrument transfer function H_I. However, x and M both generally vary with time. The triangular symbols represent amplifiers producing an output that is the difference between the two inputs multiplied by the indicated gain (1 or $1/\tau$ in these two examples). The box with an integration symbol integrates the input, in this case the time derivative of the measured quantity as given by the preceding equation, to produce the measurement. However, that measurement is also fed back to the left amplifier, which produces an output determined by the difference between $H^*(x)$ and M. This diagram is a representation equivalent to the preceding differential equation, the RC circuit diagram, or the diagram with the generic transfer function H_I. The transfer function depicted above is a low-pass filter because fluctuations having frequencies large compared to $1/\tau$ will be attenuated while those with frequencies small compared to $1/\tau$ will appear unaffected in the output.



The above diagram can be generalized to higher-order dynamic response. An example is the response of a wind vane to the direction of the wind. (This is actually a much more complex topic and the following is simplified and not realistic except as an example of a second-order dynamic system. For a real wind vane, the restoring force becomes non-linear for any but very small deflection angles, varying wind speed changes the restoring torque, and there are aerodynamic contributions to the drag term and other complicating factors.) In this simplification, two factors control how the vane will respond to the direction of the wind when the wind speed remains constant. First, a difference between the orientations of the vane and the wind will produce a torque proportional to that difference, with proportionality constant k. However, such a system will tend to oscillate about the correct orientation, like a weight suspended on a rubber band. Therefore a wind vane will have some resistance to motion built in to its mounting structure, and this resistance provides damping of the oscillation.

A graphic representation of the transfer function of this system is shown in the figure to the right, where the amplifier with gain D provides the damping effect of this feedback. The new symbol introduced in this figure is the circle-with-X representing summation of the inputs, with signs applied as indicated. This shows a damped forced harmonic oscillator, where k is analogous to a spring constant but in this case is the proportionality constant between the restoring torque and the angle between the orientation of the wind and of the wind vane. The symbol m in the amplifier with gain $1/m$ is analogous to mass for the loaded spring but here represents the moment of inertia of the wind vane. The response to changes in wind direction is determined by the relative magnitudes of the parameters (k, m, D): If D is very small, the vane tends to oscillate too much, while a large value of D causes the vane to respond too slowly. "Critical damping," for which the vane just moves to a new equilibrium position without overshooting, occurs when $D = 2\sqrt{mk}$, and the natural oscillation without damping has an angular frequency equal to $\sqrt{k/m}$.



In contrast to the first-order transfer function, which serves as a low-pass filter, the response of a second-order transfer function can lead to erroneous large-amplitude fluctuations near the natural frequency if the damping is insufficient. For large damping, the transfer function again serves as a low-pass filter.

The governing differential equation can be determined from this and other similar transfer-function diagrams, as follows: Start at the point in the diagram labeled as $\ddot{M}$ (or whichever term is the highest-order derivative). Working backward, incorporate each contribution into the differential equation to obtain, in the case of the preceding diagram,

$$\ddot{M} = \frac{1}{m}\left(k(H^*(x) - M) - D\dot{M}\right) \tag{2}$$

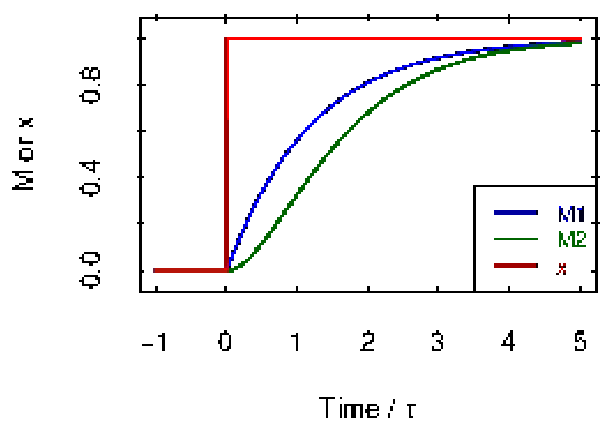
It is then possible to solve the resulting differential equations for specified values of the measurand as a function of time.

Another benefit of the characterizing the sensor with a transfer-function diagram is that it provides a structure easily incorporated, with initial values, into a numerical solution that gives the response to any input.[2] To demonstrate how solutions to these equations or the transfer-function diagrams relate to characteristics of instruments, the responses to some specific types of input will be shown in the next subsection.

Response to some specific input functions

The first-order response is characterized uniquely by the time constant τ, so a single representative response to each input function characterizes all such sensors. In the plots that follow, the time constant will be 1 s. However, for a sensor having second-order response, the response is determined by the three terms shown in Eqn. (2) or as the gain factors in the transfer-function diagram, k, m, and D. Only their relative sizes are important, so the controlling factors can be reduced to two, $\omega = \sqrt{(k/m}$ (the natural angular frequence of the system) and $\gamma = D/(2\sqrt{mk})$, the ratio of the damping term to the critical value. In the following examples, only one choice will be shown to illustrate second-order response, that for critical damping and for which $k/m = 2$ or $\omega_0 = \sqrt{2}$. This corresponds to a natural period of $T = 2\pi/\omega_0 \approx 4.4$ s. Making this period of natural oscillation smaller increases the response time, so the responses shown below for a second-order sensor are only one representative example and the response for other natural periods will vary.

The step function



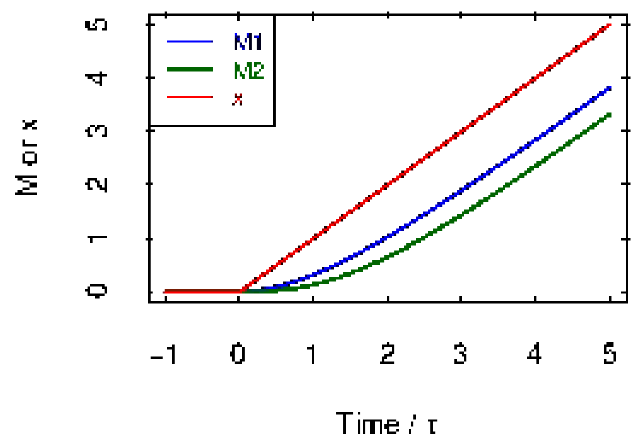
An abrupt change in the measurand to a new value, or step-function change, is shown as the red line in the next figure. The first-order response is an exponential, but the second-order response is slower in this case and has a different shape, although it is not dramatically different. The second-order response for a natural period of about 2.2 s is very close to the first-order response, and shorter natural periods give still faster response. However, the shapes of the response curves remain distinctly different

[2]For example, the following loop of code in R, with the three lines in the "for" loop corresponding to the steps $\ddot{M} \to \dot{M} \to M$ in the diagram, gives the response $M2$ to x, specified here as a step function at $t = 0$:

```
k <- 4; m <- 2; D <- 2 * sqrt(m * k)
N <- 10000; T <- 5; dt <- T / N ## loop spans 5 s in 10000 steps
x <- rep(1, N); t <- (1:N) * 5 / N
M2 <- Mdot <- rep(0, N)
for (i in 1:(N-1)) {
  H2 <- ((x[i] - M2[i]) * k - D * Mdot[i]) / m
  Mdot[i+1] <- Mdot[i] + H2 * dt
  M2[i+1] <- M2[i] + Mdot[i] * dt
}
```

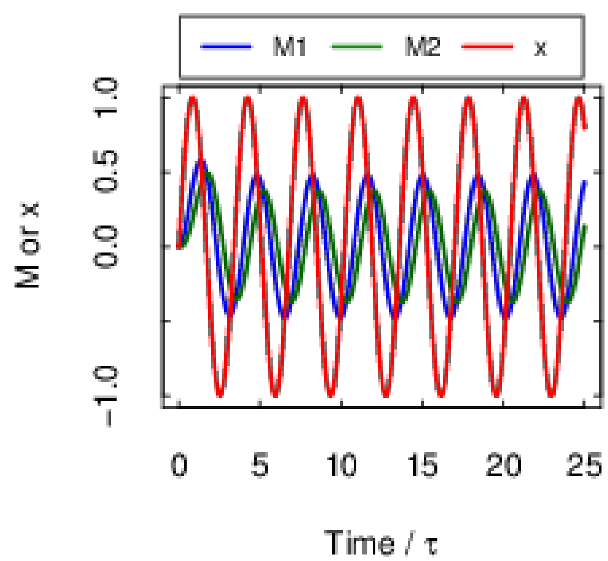
from the exponential shape of the first-order response. These response curves provide good guidance regarding how long it is necessary to wait before errors caused by time response become negligible in a particular application. Response like that shown is commonly seen; for example, temperature sensors on research aircraft often have first-order time constants of a few seconds and this imposes important limitations on the detection of spatial structures using such measurements.

The ramp function



The "ramp" variation in the measurand is that where the measurand increases at a constant rate, as shown by the red line in the figure at the right. The sensor output in this case, after a transient period, follows the slope of the measurand but with a delay. For the first-order case, the effective delay is equal to the time constant τ, so this is often a useful way to determine that time constant. In the case of aircraft-borne temperature sensors, this will result in a difference between soundings flown upward and flown downward, and the time constant of the sensors can be determined from that difference.

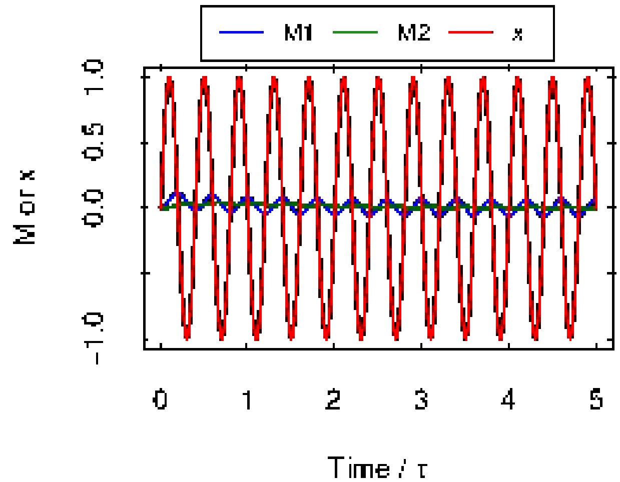
Low-frequency sine wave



The next figure shows a measurand that varies as a sine wave with frequency arbitrarily selected to be 0.29 Hz (i.e., a period of 3.4 s). This is long in comparison to the assumed first-order time constant ($\tau = 1$ s) but still below the natural frequency of the second-order system. Both response curves show attenuated response and a time lag that introduces a significant time lage between the input and response signals. For lower frequencies, the attenuation and phase lag are both reduced, but even at 0.1 Hz important attenuation and phase lag remain evident in the response curves. There is some transient response at the start of this and the next plot because the calculation is started with the measurement and its derivative both set to zero, but the transient response no longer has much effect after a few cycles of the sine wave.

High-frequency sine wave

The figure at the right shows the response for a 2.5 Hz sine wave. Here the attenuation for both response signals is strong and the phase lag approaches 90°. This and the previous figure show that both transfer functions are acting as low-pass filters by attenuating fluctuations that are fast compared to their characteristic response times. It is worth noting, however, that these aren't very good filters because the attenuation changes only slowly over a large frequency range and the transfer function introduces substantial lag over a similar range.



Some consequences for making measurements at high frequency

If good time response is needed for a particular measurement, much attention must be devoted to minimizing the effects of time lags and phase lags. Time lags introduced by effective RC delays in signal lines and instrument components are often difficult to minimize. Another concern when sampling time-series measurements is to choose an appropriate sample rate. If a system samples at a frequency f, it is not possible to detect sine-wave components that are faster than $f/2$, called the Nyquist frequency. Furthermore, higher-frequency components can be "aliased" to appear as contaminating contributions at lower resolved frequencies. To avoid this, it is best to remove components above the Nyquist frequency by filtering (using filters having better cutoff characteristics than the dynamic systems illustrated here). General guidance is to sample fast enough to give a Nyquist frequency significantly above the highest frequency of interest and then filter at or below the sample frequency to eliminate higher-frequency components that might influence the resolved frequency range.