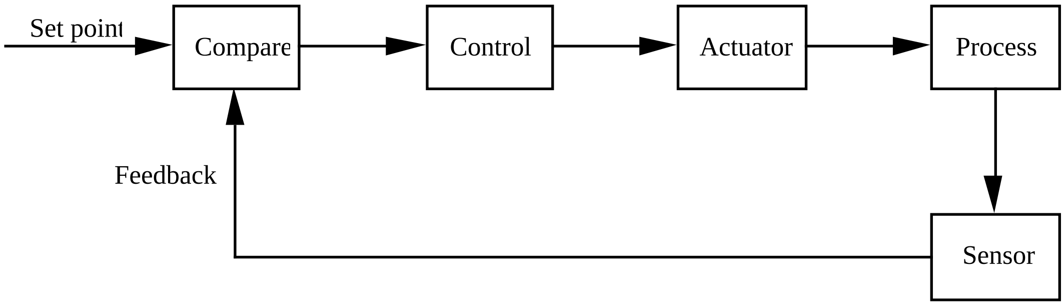
Computer Science 773

Robotics and Real-time Control

CONTROL THEORY

The name "control theory" usually denotes the mathematical treatment of closed-loop, continuous, linear systems including computing elements intended for the exercise of control over the rest of the system. This diagram shows the archetypal control system :



The aim of control theory is to identify the function which must be computed by the controller in order to make the system work as required. This involves something more or less equivalent to solving the equations of motion of the whole system, which are usually differential equations.

The restriction to linear systems is imposed to make the calculations tractable. While the equations of motion can be solved for some sorts of non-linearity in a single-loop system of the sort shown in the diagram, such solutions become much harder to obtain with more complex, and more typical, multiple-loop systems. The assumption of linearity was also convenient when analogue computing devices were used to compute the control function. Effective linear analogue devices - adders, constant multipliers, integrators, and differentiators - are comparatively easy to construct; a few particular non-linear functions, which happen to describe convenient electrical properties of some materials, can be computed, but a general non-linear function can only be implemented as a piecewise linear approximation.

Unfortunately, the linearity which simplifies the mathematics isn't always realistic : many physical devices and phenomena are not linear. Control theory can nevertheless be applied effectively to many problems by linearising the systems. Provided that all functions which describe the behaviour of the process are smooth and deviate only a little from linearity over the working range of the system variables, it may be an adequate approximation to replace them by linear approximations which satisfactorily reproduce the behaviour of the system in the working region. The result is a control system which works when the process is in its normal running state, but not necessarily under other conditions. It is typically necessary to start the process without the control system, engaging the control system only when the process variables have attained values in the normal working ranges.

It is convenient to carry out the calculations using the Laplace transforms of the equations of motion, in which differentiation and integration are replaced by simple algebraic operations. Each item in the process can be characterised by its *transfer function*, which describes the ratio of its output to its input. The transfer function of a component of the controlled system can be determined by open-loop testing, in which the response of the component to certain standard input patterns is measured. Then, given the transfer functions of the plant components and knowing the desired behaviour of the whole system, the function which must be implemented by the controller can be calculated.

As well as the control function, the analysis yields information on the stability of the complete system. Instability can be a serious problem, particularly in complex systems for which the behaviour is described by differential equations of high order, and in systems in which there are time delays. It is exceedingly important to identify conditions under which a system is unstable. With this knowledge, it may then be possible to eliminate the instability by making changes to the system, or at least to ensure that the unstable conditions are well away from the normal working regime.

DIGITAL CONTROL SYSTEMS.

Control theory was derived assuming that all variables were continuous, which was an appropriate assumption for the analogue computing machinery with which it was first implemented. The simple way to use a digital computer in a control system is to simulate as closely as possible the analogue computing elements in the traditional controllers. With luck, no further change will be needed : in particular, we will be able to keep the old control theory, at least for the time being. In fact, though, moving to digital systems has introduced two sorts of discontinuity, and these need investigation if we are to exploit the digital methods to the full.

Discontinuous number representation.

Instead of representing process measurements by continuously varying quantities like voltages, we now use binary numbers. One value can no longer change smoothly to another; instead, we must deal with a sequence of discretely different numbers. In practice, the number representation itself is not normally a problem. As most physical processes are continuous in nature, analogue-to-digital and digital-to-analogue converters are needed to link the control system to the process being controlled, but converters for most routine applications are now cheap, precise, and reliable. Once digitised, there is no problem with precision : indeed, the digital computations can easily be made more precise than their analogue counterparts, and they are not subject to drift.

Discontinuous time.

A more significant change is the move from continuous computation by analogue devices to the pattern of repeated sampling followed by computing characteristic of digital techniques. Even if a processor is dedicated to a particular control function, it cannot pay attention to the process variable at every instant, as the computation takes a small but non-zero time to complete. In practice, we would prefer to sample each input measurement as infrequently as possible, so that we don't need to dedicate a processor to each input line. How can we guarantee that we don't miss anything significant ?

The **Sampling Theorem** helps. Any signal can be represented (by Fourier analysis) as the sum of a spectrum of signals of different frequencies. If the highest frequency in the signal is f, then the Sampling Theorem asserts that it is possible to reconstruct the whole signal from a sequence of samples taken at a frequency of $2f$ (or higher). If we know, then, that the characteristic times of a system are in the range 5 seconds and upwards, we might guess that sampling every one or two seconds would give us enough information for control purposes.

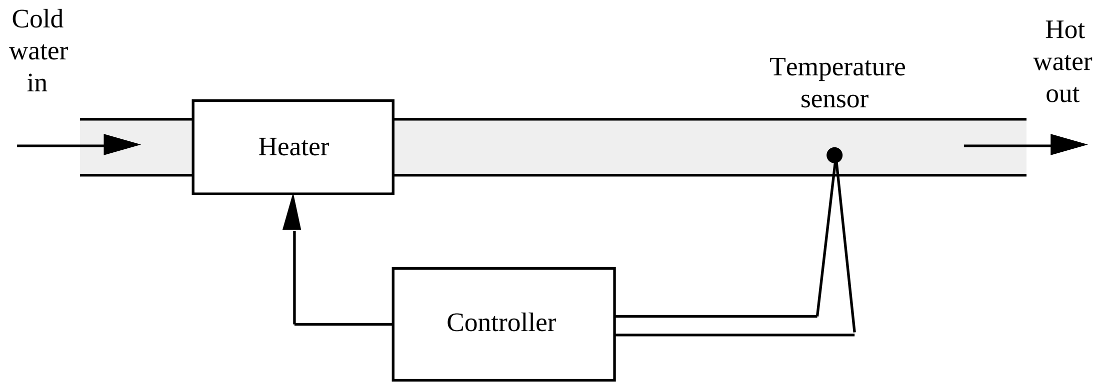
Unfortunately, our guess might be wrong. If we have a nice quiet system and environment, and there really are no awkward high-frequency resonances in the system, we would probably get away with it; but the Sampling Theorem is based on the highest frequency component which is in the signal, not the highest frequency in which we happen to be interested, so high-frequency noise of significant amplitude could wreck the control system.

To avoid such difficulties, we can use a low-pass filter to cut down the high frequency component of the signal, but such a filter will also cut down any high-frequency components of the real control signal. The result is to slow down the sudden changes, so sharp steps in the signal are to some degree smoothed, and therefore necessarily delayed. In practice, it is commonly possible to choose parameters so that a satisfactory compromise is achieved, but it's a lot better to eliminate the noise at its source than to rely on filters.

Of course, the less frequently we sample the behaviour of the system, the further we move away from the continuous model on which control theory is based. To remedy this defect, we need something like traditional control theory but based on sampled systems. Methods have been developed which parallel the Laplace transform techniques using instead the z-transform, which is similar to the Laplace transform, but with integration replaced by summation. The result is a little more clumsy, but works.

TIME LAGS AND FREQUENCY EFFECTS.

Here's a simple example which illustrates my earlier remark on the effect of time delays in a system. Consider this simple but realistic apparatus for providing a stream of water at a controlled temperature :



Suppose that the system starts from cold. The sensor detects very cold water, so the controller turns the heater full on. Assuming the simplest "ideal" case, and sensible system design, this will in fact make the water too hot - but the temperature sensor will not find out until some time later, when the flow of water through the system brings the overheated water to the sensor. The sensor now detects very hot water, so the controller will turn off the heater, and cold water will again be fed onwards towards the sensor. This cycle could continue for ever, and we'll never get a stable system.

The reason is obvious : the sensor should be much closer to the heater - ideally, inside the heater so that it can detect the high temperature as soon as possible. This is not always possible for a number of reasons, such as physical access to the system, electrical interference between the heater coils and the sensor, and so on. (Even apart from that, it isn't a perfect answer unless the water within the heater is constantly mixed so that the temperature is homogeneous - think about it.) In practice, things are not quite so bad, because the significant thermal capacity of the stationary parts of the heater smooth out the switches between heating and cooling, in effect turning the heating on and off slowly rather than instantaneously.

Ordinary control theory is not very good at handling time lags, particularly long ones, but with a digital controller one can devise a programme which takes into account the time lag and controls the temperature with reasonable success. This takes us back to the principle of computer control : to find the control programme which makes the system behave as we would like.

Time lags do not need to be long to have unfortunate consequences. The consequences of a change in any part of the system take a finite time to travel through the whole control loop, and if that time is of the order of the period of one of the system's natural frequencies (which are themselves the result of internal feedback loops), trouble is possible. The effective time lag can be affected by electrical capacitance and inductance, and by mechanical inertia and physical time delays of the sort we have just described, so a real system may have many potential sources of danger. One of the preoccupations of control theory is therefore the identification of instabilities in the system, and the approach to this is in effect to determine the behaviour of the system at all frequencies, and then by some means to find frequency ranges in which unstable behaviour is found. It may then be possible to compensate for these instabilities by adjusting the controller programme, or to avoid them by properly managing the system.

Alan Creak,
February, 1995.