Chapter 11: Feedback and PID Control Theory

I. Introduction

Feedback is a mechanism for regulating a physical system so that it maintains a certain state. Feedback works by measuring the current state of a physical system, determining how far the current state is from the desired state, and then automatically applying a control signal to bring the system closer to the desired state. This process is repeated iteratively to bring the system to the desired state and keep it there.

Feedback can be used very effectively to stabilize the state of a system, while also improving its performance: Engineers use feedback to control otherwise unstable designs; op-amps use feedback to stabilize and linearize their gain; and physicists use feedback to stabilize and improve the performance of their instruments.

A. Feedback in engineering

Feedback is ubiquitous in engineering. Its application has led to device features and machines which would not otherwise function. Here are few examples:

Climate control: A sensor measures the temperature and humidity in a room and then heats or cools and humidifies or dehumidifies accordingly.

Automobile cruise control: The car measures its speed and then applies the accelerator or not depending on whether the speed must be increased or decreased to maintain the target speed.

Highly maneuverable fighter jets: The F-16 Falcon fighter jet is an inherently unstable aircraft (i.e. the airframe will not glide on its own). The F-16 does fly because 5 onboard computers constantly measure the aircraft's flight characteristics and then apply corrections to the control surfaces (i.e. rudder, flaps, ailerons, etc…) to keep it from tumbling out of control. The advantage of this technique is that the aircraft has the very rapid response and maneuverability of a naturally unstable airframe, while also being able to fly.

B. Feedback in electronics:

Op-amps use feedback to achieve very high linearity and predictability for their closed-loop gain by sacrificing some of their extremely high open-loop gain.

Another common application of feedback in electronics is in precision, fast- response power supplies. Constant current and constant voltage power supplies which have a high degree of stability use feedback to regulate their current or their voltage, by measuring the current and voltage across a precision shunt resistors and then using feedback to automatically correct for any deviations from the desired output. Feedback also allows the power supply to adjust its voltage or current very quickly and controllably in response to a change in load.

C. Feedback in physics

Feedback has become a familiar tool for experimental physicists to improve the stability of their instruments. In particular, physicists use feedback for precise control of temperature, for stabilizing and cooling particle beams in accelerators, for improving the performance of atomic force microscopes, for locking the optical frequency of lasers to atomic transitions, and referencing quartz oscillators to ground state atomic hyperfine microwave transitions in atomic clocks, to name just a few example.

Temperature control: Many delicate physics devices, such as crystals, lasers, RF oscillators, and amplifiers, require their temperature to be very stable in order to guarantee their performance. For example, the wavelength of diode lasers generally has a temperature dependence on the order of 0.2 nm/°C, but requires a stability of 10^{-6} nm for experiments.

Stochastic cooling: In a particle accelerator, the transverse momentum spread of particles must be reduced to a minimum. The reduced momentum spread increases the particle density, or beam luminosity, and consequently the probability of collisions with a similar counter-propagating particle beam in the detector area. Stochastic cooling works by measuring the transverse positions and momenta of the particles as they pass through a section of the accelerator, and then applying appropriate momentum kicks to some of the particles at other points in the accelerator ring to reduce the overall transverse momentum spread. The process is repeated until the momentum spread is sufficiently reduced. The 1984 Nobel Prize in Physics was awarded in part to Simon van der Meer for his invention of stochastic cooling which contributed to the discovery of the W and Z bosons (weak force mediators) at CERN.

Atomic force microscope: An atomic force microscope uses a very sharp tip (just a few nanometers in size at the very tip) which is scanned back and forth just a few nanometers above the surface to be imaged. Instead of scanning the tip at a constant height above the surface, which could lead to the tip actually running into a bump on the surface, the microscope uses feedback to adjust the tip height such that the force (from the surface atoms) on the tip is constant.

Laser locking: Many experiments in atomic and optical physics require lasers which have a very stable optical frequency. The optical frequency of the laser is locked by measuring the optical frequency difference between the laser and an atomic transition and using feedback to set this difference to a constant value. Lasers can be routinely stabilized with feedback to better than 1 MHz out of 3×10^{14} Hz (about 1 part per billion), though stabilities close to 1 Hz have been reported after heroic efforts.

Atomic clocks: In an atomic clock, the frequency of an RF oscillator (a quartz crystal for example) is compared to that of a ground state atomic hyperfine microwave transition (6.8 or 9.2 GHz). The frequency difference is measured and the frequency of the RF oscillator is corrected by feedback. The process is constantly repeated to eliminate any drift in the frequency of the RF oscillator. Atomic fountain clocks can achieve accuracies in the range of 1 part in 10^{15}, and plans are underway to construct optical atomic clocks with accuracies and stabilities of about 1 part in 10^{18}.

II. Feedback

In this section we introduce the main elements of a generic feedback model.

A. System

Consider a simple system characterized by a single variable ***S***. Under normal conditions the system has a steady state value of $\boldsymbol{S=S_0}$ which may vary and drift somewhat over time due to the variation of environmental variables ***v*** which we cannot measure or are unaware of. We possess a mechanism for measuring the state of the system as well as a control input ***u*** with which we can use to modify the state S of the system. In summary, the system has the following functional form ***S(u; v; t)***. We will make the final assumption that S is monotonic with *u* in the vicinity of S_0 (i.e. that the plot of *S* vs. *u* does not have any maxima or minima, and that *dS/du* is either always positive or always negative).

Figure 11.1 shows a conceptual schematic of the relationship between the system, the variables *u* and *v*, and the measurement of the system state *S*.

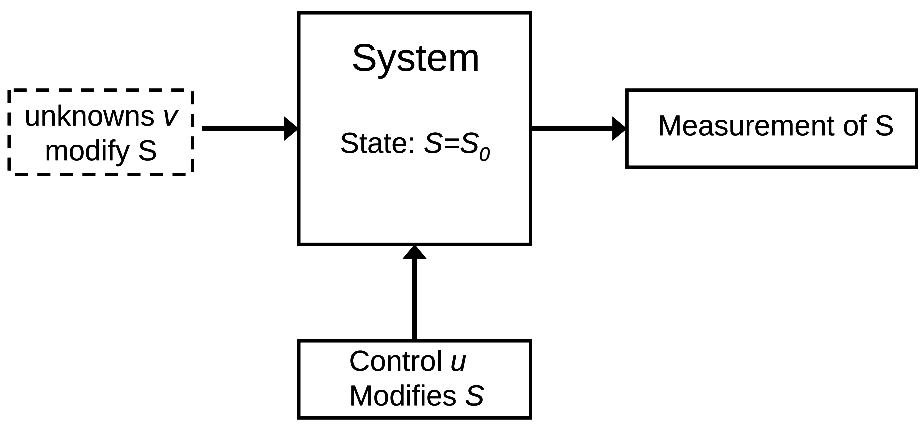


Figure 11.1: Conceptual schematic of system

B. Objective

Our objective is to set or lock the state of the system to a desired value $S=S_d$ and keep it there without letting it drift or vary over time, regardless of variations in the environmental variables *v*.

C. Feedback model

We will set or lock the state of the system to $S=S_d$ with the following procedure (see also figure 11.2):

1. Measure the state S of the system.

2. Determine how far the system is from its desired set point by defining an error variable, $e=S-S_d$.

3. Calculate a trial control value $\boldsymbol{u=u(e)}$.

4. Feed the calculated control value, $u(e)$, back into the control input of the system S.

5. The state of the system changes in response to the change in the control value.

6. Return to step.

If we repeat this feedback cycle indefinitely with an appropriately calculated control value $u(e)$, then the system will converge to the state $S=S_d$ and remain there even under the influence of small changes to other variables (i.e. v) which influence the value of the state S.

This feedback model can be adapted to include several state variables and several feedback variables.

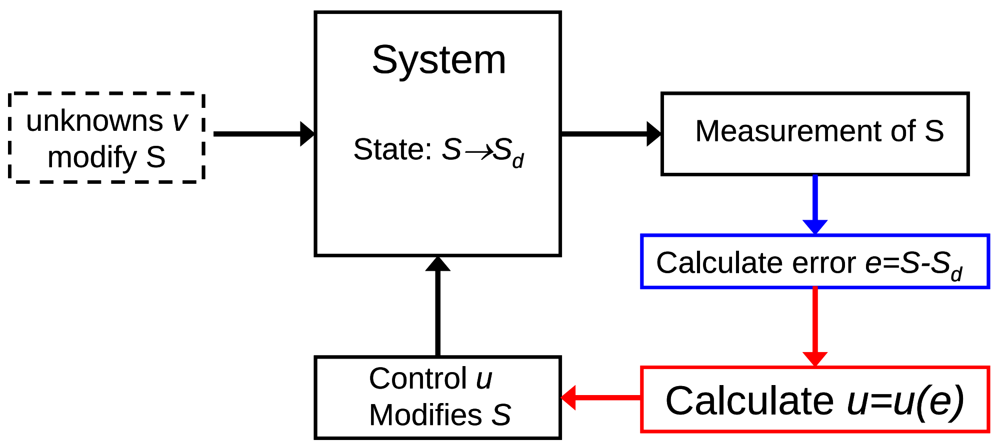


Figure 11.2: Conceptual schematic of system with the feedback loop.

In section III, we discuss a frequently used expression for calculating the feedback control variable $u(e)$.

III. PID Feedback Control

The most popular type of feedback stabilization control, $u(e)$, is Proportional-Integral-Derivative (PID) gain feedback. PID is very effective and easy to implement. The expression for $u(e)$ depends only on the error signal $e=S-S_d$ and is given by

$$u(e;t) = g_P e(t) + g_I \int_0^t e(t)dt + g_D \frac{d}{dt} e(t) \tag{11.1}$$

where g_P, g_I, and g_D are respectively the proportional, integral, and derivative gains. We also note that g_P, g_I, and g_D do not have the same units. We will assume for simplicity that g_P is dimensionless in which case $u(e)$ has the same units as S.

A. Time evolution of the system with PID feedback control

We are now in a position to calculate the time evolution of the system under the influence of feedback. Without feedback, the system would remain in the state S_0:

$$S_{no\ feedback}(t) = S_0 \tag{11.2}$$

S_0 may vary in time, but we will ignore this effect until part III.C.

In the presence of feedback, the state of the system at time $t+\Delta t$ (step 5) depends on the state of the system without feedback, S_0, which has been modified by the control input variable $u(e)$. We now make the following simplifying assumption that the control input variable, $u(e)$, "controls" or modifies the state of the system S through the process of addition. In this case, the system state variable S evolves according to the following equation:

$$S(t+\Delta t) = S_0 + u(e;t) \tag{11.3}$$

We can convert this equation to an integro-differential equation, if we assume that the system has a characteristic reponse time $\boldsymbol{\tau}$ (small). In this case, equation 3 becomes

$$S(t) + \tau \frac{d}{dt} S(t) = S_0 + g_P e(t) + g_I \int_0^t e(t)dt + g_D \frac{d}{dt} e(t) \tag{11.4}$$

B. Special case: pure proportional gain feedback

As a limiting case we consider pure proportional gain feedback ($g_I=0$ and $g_D=0$). We study this special case, because it is the basis for op-amp feedback and is also the simplest form of feedback. For $g_I=0$ and $g_D=0$, equation 4 becomes

$$S(t) + \tau \frac{d}{dt} S(t) = S_0 + g_P e(t) \tag{11.5}$$

We can solve this 1st order differential equation, for the **initial condition** $\boldsymbol{S(t=0)=S_0}$, with the same technique we used in chapter 3 (equations 17-21). After a little bit of integration and algebra, which is left as an exercise to the reader, we find the following solution:

$$S(t) = \left(S_0 - \frac{S_0 - g_P S_d}{1-g_P} \right) e^{-\left(\frac{1-g_P}{\tau}\right)t} + \frac{S_0 - g_P S_d}{1-g_P} \tag{11.6}$$

Equation 11.6 shows that the system will converge to the state $S=(S_0-g_PS_d)/(1-g_P)$ when feedback control is applied, so long as the exponential exponent is negative (i.e. $g_P<1$), otherwise S will diverge. We note that $g_P<0$ corresponds to negative feedback.

Figure 11.3 shows the response of a system for a dimensionless gain of $g_P=-10$ and state values $S_0=0.5$ and $S_d=1$, with time measured in units of τ (the system characteristic response time).

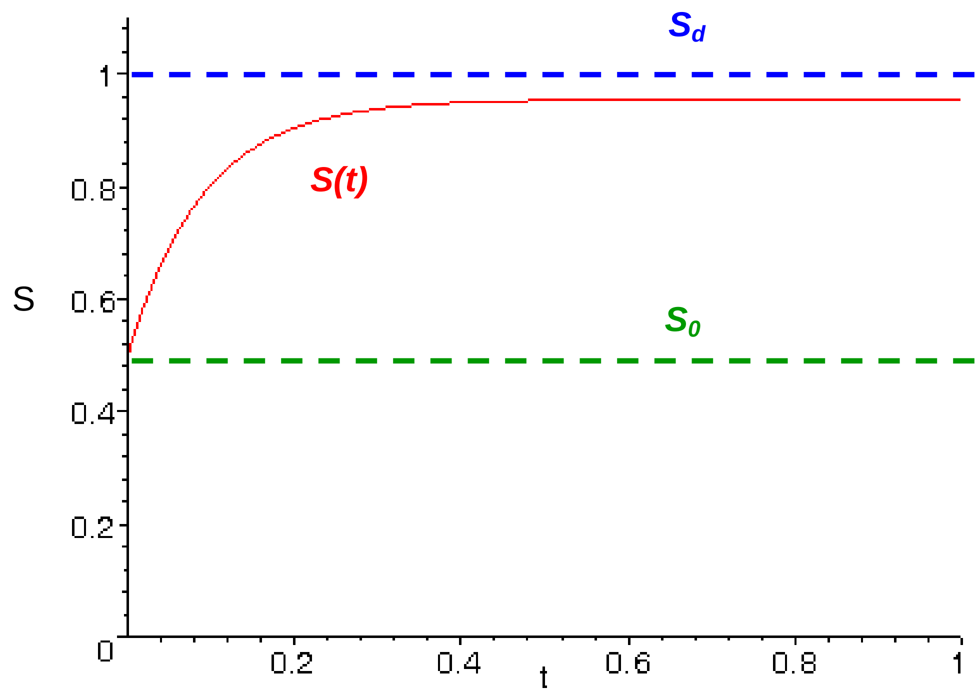


Figure 11.3: System response with pure proportional gain feedback control and parameters $g_P=-10$, $S_0=0.5$, and $S_d=1$. Time is measure in units of τ (the system characteristic response time).

As figure 11.3 makes clear, the system does not converge to the desired state $S=S_d$, though it does reach its final steady state value relatively quickly. If we restrict ourselves to negative feedback, then according to equation 10, the system will converge to the steady state value S_{ss} of

$$S_{ss} = \frac{S_0 - g_P S_d}{1-g_P} \tag{11.7}$$

Equation 11.7 indicates that the system can be made to converge to a steady state value S_{ss} which is arbitrarily close to $S=S_d$ just by increasing the gain. In fact, for infinite proportional gain (i.e $g_P \to -\infty$) the system does converge to $S_{ss} = S_d$: This is the limit in which op-amps feedback operates.

A note of caution: On its own, equation 7 is a little misleading since it would seem to imply that large positive feedback, $g_P \to -\infty$, would also produce $S_{ss} = S_d$. Of course, this is not true since according to equation 11.6, the system will never achieve a steady state, but instead will diverge forever.

C. Solution for PI feedback control

A large majority of PID feedback controllers are actually just PI controllers (i.e. proportional and integral gain, but no derivative gain), and so for simplicity we solve equation 11.4 without the derivative gain term ($g_D=0$). The inclusion of the derivative gain term is conceptually simple and follows the same treatment as PI feedback and is left as an exercise to reader. Derivative gain is used to improve the time response of the feedback, so that the system converges more quickly to its steady state value.

With the derivative gain term omitted, equation (11.4) becomes

$$S(t) + \tau \frac{d}{dt} S(t) = S_0 + g_P e(t) + g_I \int_0^t e(t) dt \tag{11.8}$$

We can convert this integro-differential equation to a 2nd order linear differential equation with constant coefficients by taking the time derivative of equation (11.8) to obtain

$$\frac{d}{dt} S + \tau \frac{d^2}{dt^2} S = g_P \frac{d}{dt} S + g_I S - g_I S_d \,, \tag{11.9}$$

where we employed the substitution $e(t)=S-S_d$. After combining similar terms, equation (11.9) becomes

$$\tau \frac{d^2}{dt^2} S + (1 - g_P) \frac{d}{dt} S - g_I S = -g_I S_d \tag{11.10}$$

Equation 11.10 is an inhomogeneous 2nd order differential equation with constant coefficients. The full solution to equation (7) is given by

$$S(t) = A_+ e^{\lambda_+ t} + A\_ e^{\lambda\_ t} + S_d \tag{11.11}$$

$$\lambda_\pm = \frac{(g_P - 1) \pm \sqrt{(g_P - 1)^2 + 4 g_I \tau}}{2\tau} \tag{11.11a}$$

The first two terms of equation 11.11 represent the homogeneous solution to equation 11.10, while the 3rd term is the inhomogeneous solution to the equation (it does not depend on the initial condtions). A_+ and A_- are constants to be determined from the initial conditions.

Equation 11.11 shows that the system will converge to the state $S=S_d$ when feedback control is applied, so long as λ_+ and λ_- are negative (i.e. negative feedback), otherwise S will diverge (exactly the opposite of what we want to accomplish with feedback).

If we choose $S(t=0)=S_0$ and $dS(t=0)/dt=0$ as our initial conditions, we can calculate the constants A_+ and A_-. After a little bit of algebra, we find that

$$A_+ = -\left(\frac{\lambda_-}{\lambda_- - \lambda_+}\right)(S_d - S_0) \tag{11.12a}$$

$$A_- = \left(\frac{\lambda_+}{\lambda_- - \lambda_+}\right)(S_d - S_0) \tag{11.12b}$$

In figure 11.4, the behavior of the system under PI feedback control is plotted for several different parameters configurations.

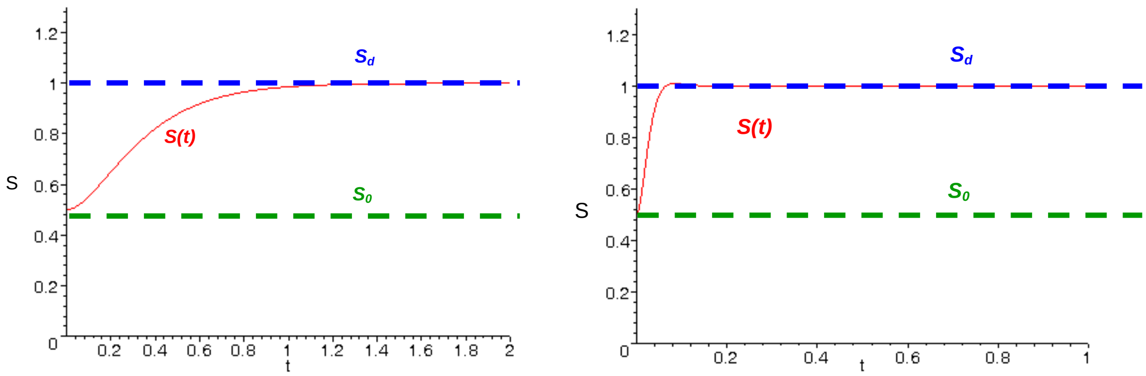


Figure 11.4: Time-evolution of a generic system with PI control feedback for $S_0=0.5$ and $S_d=1$. For the left hand plot the gain parameters are $g_P=-10$, $g_I=-30$, while for the right hand plot the parameters are $g_P=-100$, $g_I=-4000$. The small overshoot in the left hand plot is due to a small imaginary part in the exponential exponent of equation 11.11(a).

The primary purpose of integral gain is to provide essentially infinite gain at DC (0 Hz), which guarantees that $S_{ss}=S_d$, as can be seen in figure 11.4. Figure 11.4 also shows that **the larger the gain, the faster the correction time of the feedback control loop.**

D. Fourier space analysis of noise suppression

One of the primary objectives of feedback is to make the system insensitive to noise on the system state S, so that the system state stays locked to $S=S_d$ regardless of external influences.

In the absence of corrective feedback, external noise will cause the system state to deviate from $S=S_0$. External noise at a frequency ω will cause the system state to oscillate around its natural steady state such that $S=S_0+S_N\cos(\omega t)$, where S_N is the amplitude of the oscillations. Following the standard Fourier space recipe of chapter 3, we replace $\cos(\omega t)$ with $\exp(i\omega t)$, and then take the real part at the end of our calculations. In essence, we must re-solve equations 11.8 with the following modification:

$$S_0 \rightarrow S_0 + S_N e^{i\omega t} \tag{11.13}$$

Using the substitution of equation 11.13, equation 11.10 becomes

$$\tau \frac{d^2}{dt^2} S + (1 - g_P) \frac{d}{dt} S - g_I S = i\omega S_N e^{i\omega t} - g_I S_d \tag{11.14}$$

Equation 11.14 has the same homogeneous solution as equation 11.10, but the inhomogeneous solution, $S_{ih}(t)$, differs and is given by the following expression

$$S_{ih}(t) = \frac{i\omega}{(1 - g_P) i\omega - \tau\omega^2 - g_I} S_N e^{i\omega t} + S_d \tag{11.15}$$

We see that the noise term is present in the inhomogeneous solution, but with an additional factor modifying the amplitude of the noise. In the case of negative feedback the modulus of this suppression factor, which we will call A_N, is always less than unity and is given by the following expression

$$A_N = \frac{\omega}{\sqrt{(\tau\omega^2 + g_I)^2 + (1 - g_P)\omega^2}} \tag{11.16}$$

The plot in figure 11.5 shows the dependence of the suppression factor, A_N, on frequency for different feedback schemes. The plot shows that a combination of proportional and integral control gives the best suppression of noise, except in the vicinity of the "resonant" frequency $\omega = \sqrt{-g_I / \tau}$. The high frequency drop-off of the suppression factor is not due to feedback but simply the natural response time τ of the system which also suppresses noise.

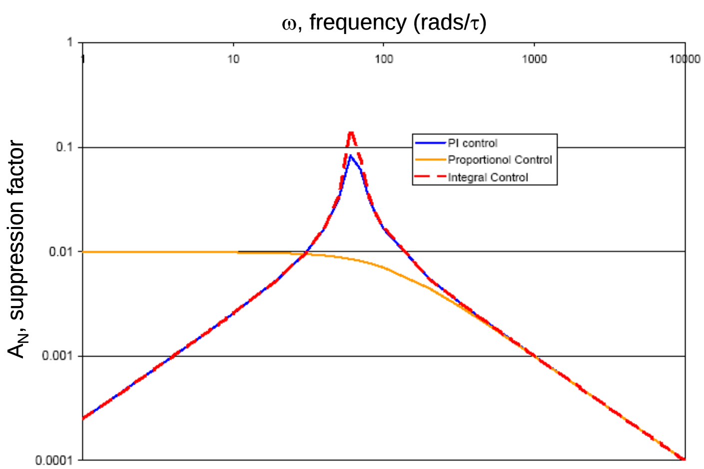


Figure 11.5: Comparison of the suppression factor, A_N, for different feedback schemes. The feedback control loop parameters are g_P*=-100* and g_I*=-4000*.

IV. Reality

In practice, feedback is not quite as straightforward as presented in the previous section.

A. Gain vs, Frequency

In the theoretical treatment of part III, we assumed that the proportional gain was independent of frequency. In practice, gain will generally fall off at higher frequencies due to natural low-pass RC filtering in an amplifier and the larger circuit.

As an example, figure 11.6 shows a plot of the open-loop gain of an op-amp as function of frequency, which has a clear drop-off in gain at higher frequencies.

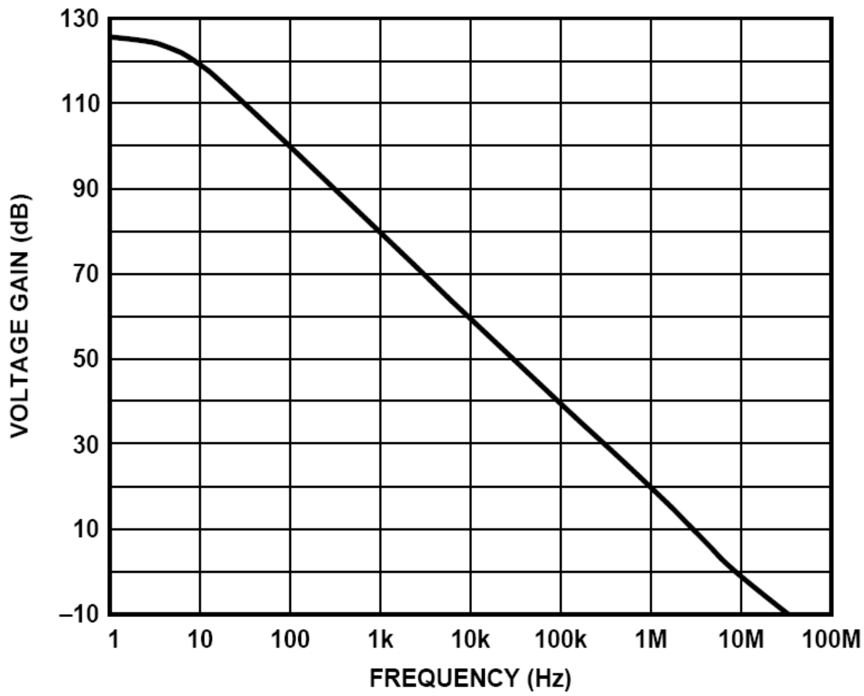


Figure 11.6: Open-loop gain of the OP27 op-amp (Analog Devices OP27 datasheet revision F, p. 10 (2006)).

B. Phase shifts and positive feedback

The natural or stray RC filtering of an amplifier not only rolls off the gain at high frequencies, but also introduces a $-\pi/2$ phase shift. If the feedback loop has a second stray unintentional RC filter present (for example, the natural time response of the system), then a second $-\pi/2$ phase shift is introduced. If the feedback gain is larger than 1 at the frequency at which the total accumulated phase is $-\pi$, then the feedback loop goes into positive feedback which causes the state of the system to diverge or sometimes oscillate out of control.

C. Stray RC positive feedback compensation

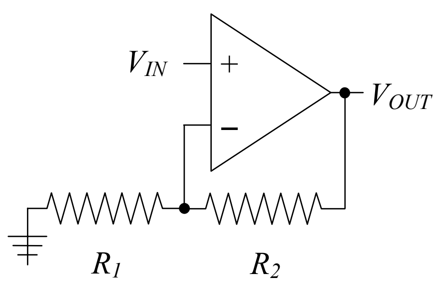
One way to avoid having the system go into positive feedback is to purposely introduce an additional RC low-pass filter into the feedback loop. If this RC filter has a f_{3dB} frequency which is sufficiently smaller than the frequency at which the positive feedback occurs then the attenuation of the filter can bring the gain below 1 when the $-\pi$ phase shift occurs. This way the feedback loop will no longer go into positive feedback above a certain frequency (of course there will not be any noise suppression or feedback action above this frequency either).

Design Exercises:

Design exercise 11-1: Consider an LED facing a photodiode in a manner similar to what you did last week in lab 10. Design a circuit which will maintain a constant optical power incident on the photodiode, even in the presence of external fluctuations in the room lighting. You should use the op-amp circuit of figure 10.10 and use PI feedback control to stabilize the intensity of the LED. Your circuit should be able to provide at least 10 mA at ~2 V to the LED.

Design exercise 11-2: **The non-inverting op-amp amplifier with finite open-loop gain.**

In this exercise you will NOT use the op-amp golden rules to solve the problem, unless explicitly indicated.
Consider the non-inverting op-amp amplifier in the circuit below.



In the following parts, V_+ and V_- refer respectively to the voltages at the non-inverting and inverting terminals of the op-amp. The open-loop gain of the op-amp is A.

a. Write down the fundamental op-amp relation between V_+, V_-, V_{out}, and A when no feedback is present (i.e. when R_1 and R_2 are not present).

b. Assuming that the V_- input draws no current, derive an expression for V_- in terms of V_{out}, R_1, and R_2.

c. Obtain an expression for V_{out} in terms of V_{in}, A, R_1, and R_2, and determine the gain G of the amplifier. Calculate V_{out} in the limit of $A \to +\infty$. Calculate V_{out} in the limit $A \to -\infty$ and comment on its physical meaning.

d. Suppose that the open-loop gain, A, is not very constant with frequency and changes by ΔA between frequencies f_1 and f_2. Derive an expression for the resulting relative variation in the amplifier gain $\Delta G/G$ in terms of the relative variation in the open-loop gain $\Delta A/A$. Calculate $\Delta G/G$ and $\Delta A/A$ for $A=10^6$, $\Delta A=10^5$, $R_2=100$ kΩ, and $R_1=10$ kΩ.

e. Most op-amps feature a significant drop-off in their open-loop gain at frequencies above ~10 Hz. The drop-off follows a well established curve given by Af=constant, where the constant is called the gain-bandwidth product (f is frequency in Hz). The gain-bandwidth product of the OP-27 is 8 MHz. On a same log-log graph, plot the open-loop gain A vs frequency and the closed loop gain G vs. frequency (for $R_2=100$ kΩ and $R_1=10$ kΩ) for an OP-27-based non-inverting amplifier.