ELECTRONOTES 219

Newsletter of the Musical Engineering Group

1016 Hanshaw Road, Ithaca, New York 14850

Volume 23, Number 219 November 2013

FEEDBACK REVISITED – GAIN DUE TO FEEDBACK

INTRODUCTION

Most readers here automatically embrace the general notions of feedback, and use it reflexively in many designs. For this reason, we may well not take too much issue with the inexact uses of the term (and with misconceptions) when they are bantered about in various media. So engineers use certain standard structures rather automatically. When the media (and pundits) wants to sound scientific, they use the term and trot out various canonic examples (too often innately flawed or embellished to a bogus status).

Currently we see a discussion of positive feedback as an element in understanding climate change (formerly called "Global Warming"). In an attempt to explain the reported global temperature increases from about 1970 to about 2000, man-made CO_2 is not enough. Hence a positive feedback mechanism involving water vapor was postulated to amplify the CO_2 effect by a factor of 3. This is almost certainly quite ridiculous, and anyway, since about 2000 the temperatures have leveled off or declined. All I want to discuss here, however, is the misunderstanding of the term "positive feedback" and hence a review here. Here are three important results to investigate and explain:

(1) A positive feedback is not a blow-up or run-away going on forever. We could have and often do have a positive feedback of gain less than 1, and the result is an amplification to a larger steady state level – not the level we have without that feedback of course. This new level might be, or might NOT be of significance or concern.

(2) A feedback should be first assumed to be linear. Calculations of gains and losses are exactly the same as what we often do with linear systems automatically – the feedbacks just

being part of a "transfer function". It is however perfectly possible for a feedback to go beyond linear regions. Such instances may lead to blow-ups, and extrapolations from the linear results DO NOT apply. They may go to infinity – in theory. But if they are physical systems, something else must step in to limit the output which would otherwise correspond to infinite energy.

(3) The analogies cited in the media are likely to mislead. When a favorite example of positive feedback – an atomic bomb – is cited, that is alarming of course. One timid neutron causing a fission and resulting in two or three neutrons, etc., etc. An explosion. Do we need more convincing that positive feedback is dangerous?

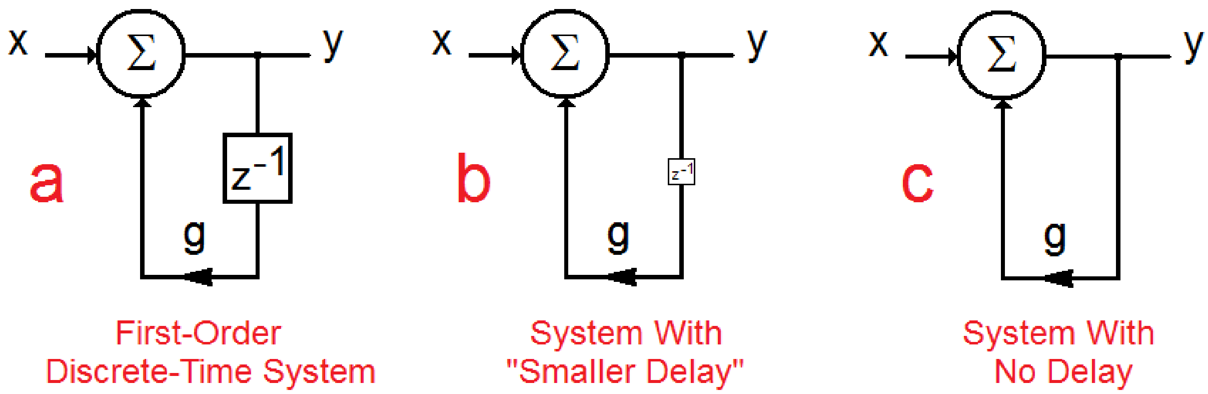


Fig. 1

LINEAR SYSTEM VIEW

Here we first take the standard viewpoint of linear systems. Fig. 1a shows the familiar first-order recursive structure where an output (y) is the sum of an input (x) plus a delayed version of the output multiplied by a gain g. In standard notation:

$$Y(z) = X(z) + gz^{-1}Y(z) \tag{1}$$

which results in a standard transfer function H(z) = Y(z)/X(z) of:

$$H(z) = \frac{1}{1 - gz^{-1}} \tag{2}$$

Working from this, we see that that H(z) has a pole at z=g, and the magnitude of the transfer function (the frequency response) is:

$$|H(z)| = [\ H(e^{j\omega T})H(e^{-j\omega T})\ ]^{1/2} = \left[\frac{1}{1 - 2gCos(\omega T) + g^2}\right]^{1/2} \tag{3}$$

where ω is the angular frequency and T is the time delay (reciprocal of the sampling frequency), and we have used $z^{-1} = e^{-sT} = e^{-j\omega T}$. To this point we note that g can be either negative or positive, but must be within the limits $-1<g<+1$ for stability.

Here we want to concentrate on a constant input. We want to see how the gain g, still chosen to be stable, amplifies or attenuates this DC signal. This we do by evaluating the frequency response at DC, which is $\omega = 0$ in equation (3), or more simply z=1 in equation (2). Both give a DC gain of 1/(1-g).

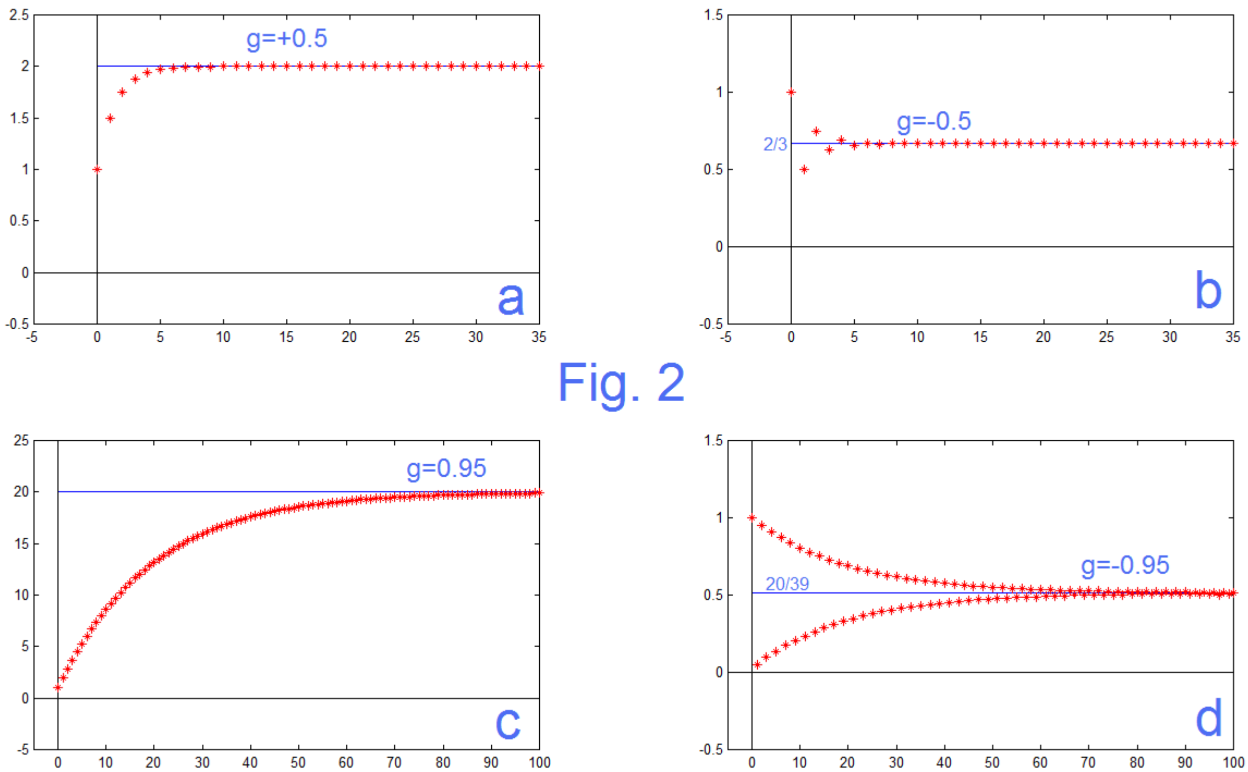


Fig. 2

A way of verifying this result is to use simulation. We simply assume a step for the input in Fig. 1a and iterate the system for a while until it seems to converge. Fig. 2 shows some example results. These cases show values of g in between -1 and +1, according to the linear systems requirement that the pole be inside the unit circle. Indeed we see that in all four examples, we have the results of each iteration (the red stars) being attracted toward and converging on the result of equation (3), the blue line. Note that the result can be very large (approaching infinity) as g approaches +1, and the result approaches 0.5 as g approaches -1. While we may never have looked at this in exactly this way, there is nothing new or unexpected here.

CLASSICALLY UNSTABLE

What happens if g is +1, or exceeds +1; and what if g is -1, or more negative than -1? Fig. 3, analogous to Fig. 2, shows these examples. These cases are usually part of any linear systems studies. All four cases here are unstable, Fig. 3b at least is unstable in being an oscillator. Fig. 3a is a linear ramp, Fig. 3c is an exponential ramp, and Fig. 3d is an exponentially increasing oscillator. All of these are understood in terms of the poles reaching the unit circle, or moving outside. Here it is interesting that Fig. 3a ramps while Fig. 3b merely oscillates. Both have poles exactly on the unit circle (at z=+1 and z=-1 respectively). The difference is that the input used to drive these is a step, which also has a pole at z=1, so combined, Fig. 3a has a second-order pole at z=+1, which blows up. Note that we don't need any fancy math – it is clear from Fig. 1a and the parameters chosen that the sequences of Fig. 3 result.

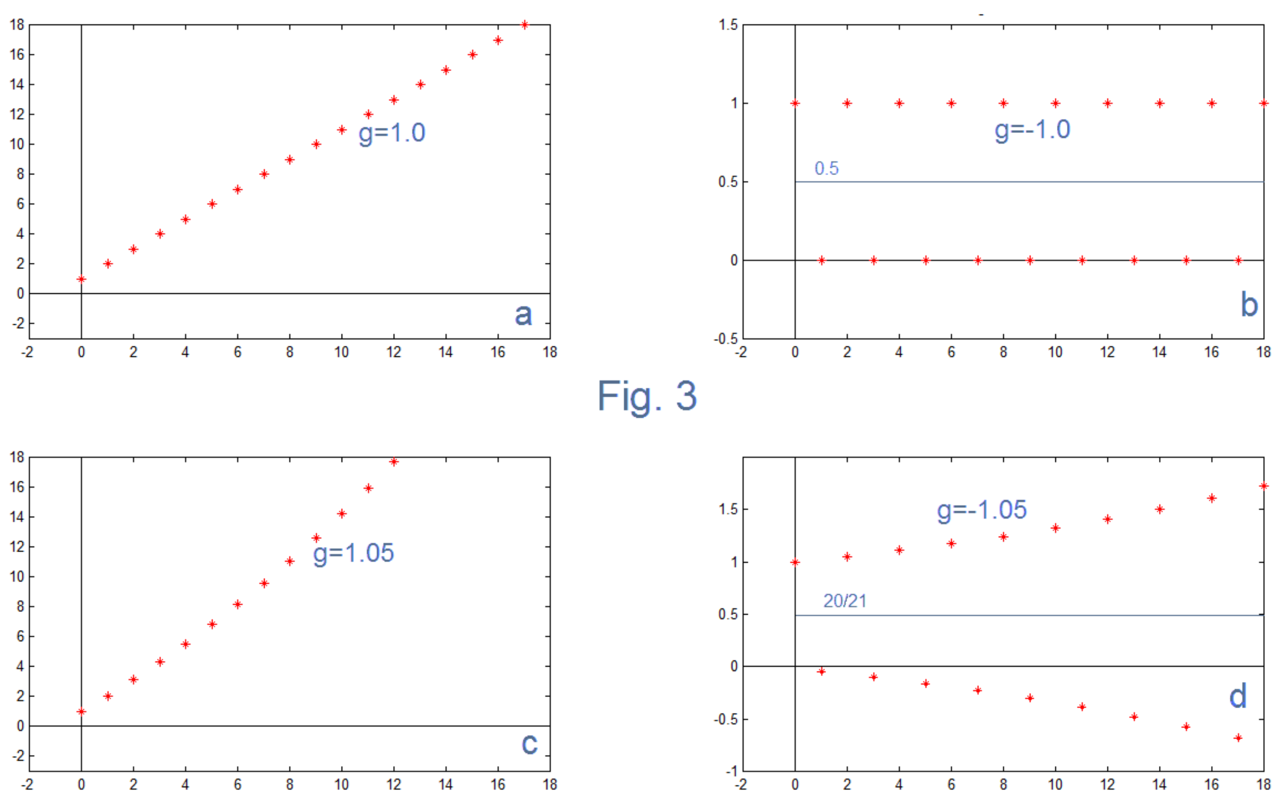


Fig. 3

REMOVING THE DELAY

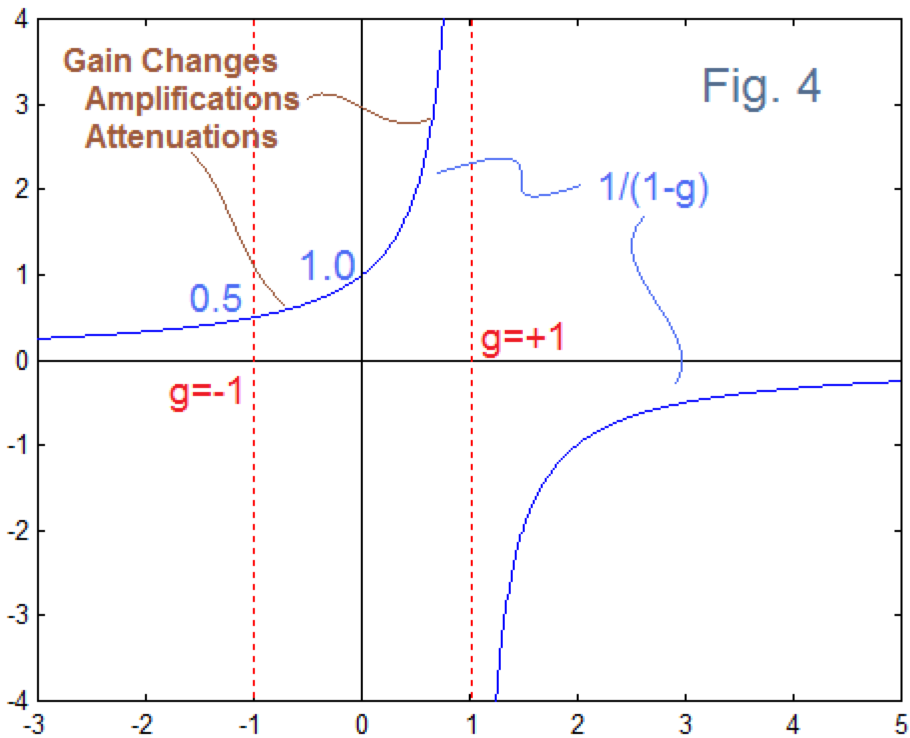
So far, we have done nothing that the reader is not likely to have seen many times before. We want next to remove the delay from the system. This is indicated in Fig. 1c. We have also included a BOGUS case of Fig. 1b where we suggest getting rid of the delay gradually by making it smaller and smaller. But there is no actual difference between Fig. 1a and Fig. 1b – just a reminder. Recall that:

$$z^{-1} = e^{-sT} = e^{-j\omega T} \qquad (4)$$

so in order to make z "go away" we want $z^{-1}=1$, which is possible if either ω=0 (a property of the input) or if T=0 (a property of the system). Here we are mainly interested in T → 0 as this is what we were proposing, to make the delay go away. In this case, equation (2) becomes:

$$H(z) = \frac{1}{1-g} \qquad (5)$$

which is the same as the DC gain of the system with delay. This had to happen because we said that ω→0 had the same effect as T→0 in equation (4). Fig. 4 reminds us of what equation (5) really looks like. So it is clear that equation (5) does blow up at g=1, but otherwise, the z-plane having disappeared with z, the stability properties are not very evident at this point.



In fact, the region inside the dashed red lines is still being thought of as related to our familiar interior of the unit circle. While equation (5) has to be true, the question becomes: will it leads to solutions that are the result of an inherent convergence? The solutions in Fig. 2 clearly are attracted to the blue-line results. Those in Fig. 3 are not – or the blue line does not even show on that scale. We thus see nothing new about setting T→0 except the possibility that equation (5) might mislead us. But is it still the result that the cases of interest are -1 < g < +1? Is it the case that the feedback may be positive or negative, restricted to a magnitude of just less than 1, and correspond to a gain change? In particular, for sure, positive feedbacks of less than 1 are amplifications, and not run-away blow-ups. But we also need to consider values of g with magnitudes greater than 1.

OBEYING THE EQUATION

Still, equation (5) seems “right” and must mean something. In fact, for values of g outside $-1 < g < +1$ it tells us some results corresponding to unstable equilibrium. Consider for example g = +2. In this case, 1/(1-g) = -1 (in general, refer to Fig. 4). In this case, if the input is +1, and the output is already -1, the amount fed back is -2, and the output is 1 – 2 = -1. This is a solution to equation (5).

While we can choose an input and output to meet equation (5) we may be disappointed to find that iterations of the equation, for general initializations, do not converge to that solution – in fact it seems that they repel away from the solution. Fig. 5 shows such a situation corresponding to g=2.

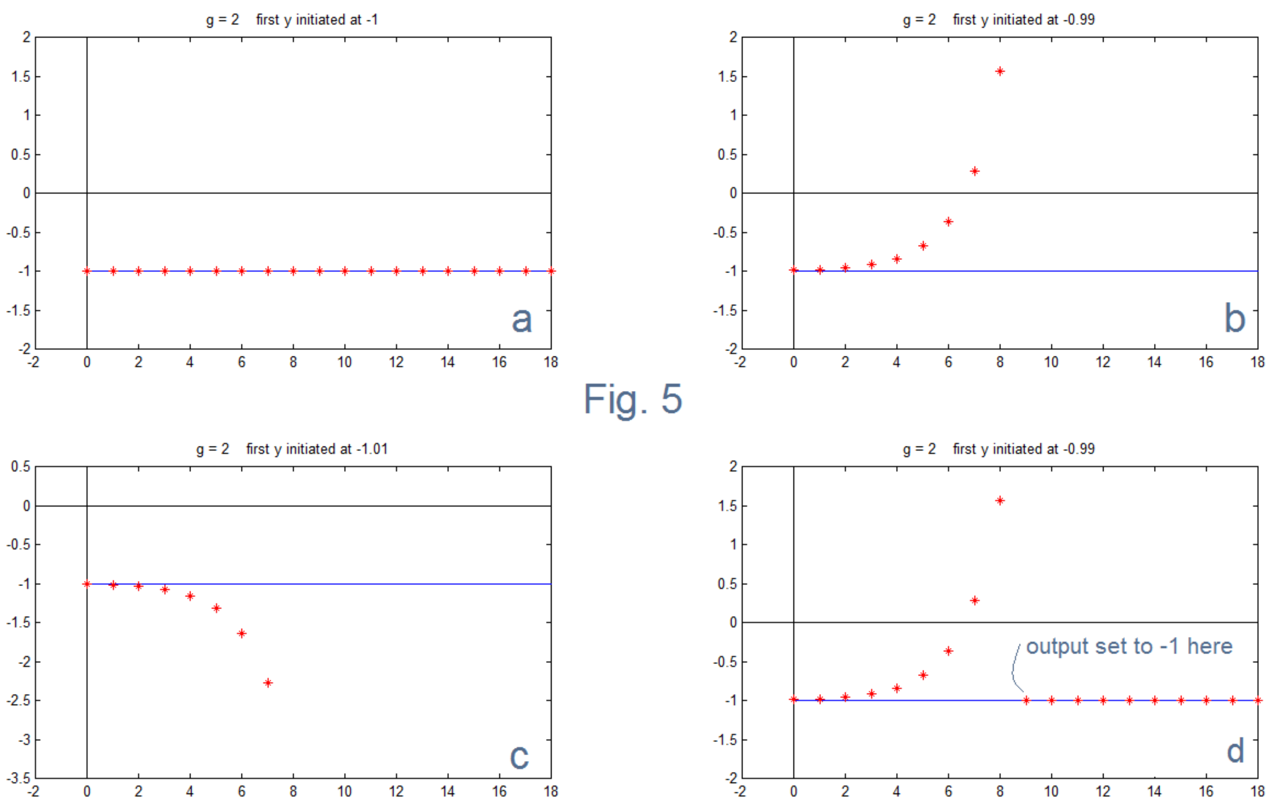


Fig. 5

In Fig. 5a, we see what happens when we set the very first output to the exact answer of equation (5) - that is, to y=-1. It sticks there – it kind of had to. So if initializing to -1 works, surely it would also work if we used something close to -1, such as -0.99 or -1.01. Fig. 5b and Fig. 5c show that this is not the case. In fact, the output repels rapidly from the equation (5) solution. Note also that the discrepancy from the solution we were hoping for is “moving away” according to the displacement of the first choice – not wandering about and possibly running into the correct solution and sticking there. We can artificially force a return to the solution of -1 by simply letting the system run a bit, and then overriding the current value with -1, and it sticks there (at iteration 9 in Fig. 5d). Similar results are obtained for negative feedback more negative than -1. For example, for g=-2, a solution of y=1/3 sticks.

Here we have established a useful perspective relating to feedback gains and our well-known results concerning the stability of linear systems. They agree that for $-1 < g < +1$, the system is stable (including positive feedback up to +1). We see expected gain changes. For unstable cases $g<-1$ and $g > +1$, we have exponential blow-ups. We need to remind ourselves below that it is an error to consider a mathematical blow-up to correspond to reality in a corresponding physical system.

POSITIVE FEEDBACK LEADING TO BLOW-UP?

We have seen clearly that, in isolation, the mathematics leads to a conclusion that in discrete time, if the magnitudes of a feedback coefficient is greater than one, that an exponentially increasing solution occurs. We have tried to hint at the logical point that this mathematical result notwithstanding, in a real situation (physical), such a blow-up is of course impossible. That is, If the mathematical output corresponds to something real (matter, energy) we can't expect a process to run to an infinite amount.

What is the nature of the process or processes that limit the exponential growth? In particular, is it a feedback of the opposite sign that begins to take effect? Or does it just run out of "stuff" to grow with? Or is this related to the notion of a "governor" that we have learned about (kind of an artificial, intentionally imposed limit)? Let's look at some examples.

REGENERATION – RADIO RECEIVERS:

This is an early example of positive feedback in radio engineering. The "Regenerative Receiver" design, invented in 1914 by Edwin Armstrong, and a popular home receiver in the 1920's through 1930's. The name even sound good. This design used a single amplifying element (a tube) to first amplify the RF signal and then usually also the AF signal. Very large gains were possible at the expense of having to frequently "tweak" the feedback "tickler coil". This coil fed back some of the amplified signal to the main tuning coil. (In some commercial radios, this tickler coil was inside the main coil that could be rotated by a lever to run from perpendicular to parallel to the main windings.) But monkeying with it was a good part of the fun. Even as this design was replaced with the multi-tube super-heterodyne "five-tube-wonder" in the 1940's and 1950's (and beyond), the regenerative receiver was revived in the 1950's for hobbyist one-transistor circuits. Again, tweaking was half the fun. Clearly, this was good positive feedback. In our electronic music work, Moog's four-pole filter with its feedback was (and is still) similar.

So this is an example of amplification through positive feedback. It was certainly possible for there to be too much feedback, and an RF oscillation with unwelcome squeal. Not to mention that could broadcast to your neighbors! But it never blew up because its electrical power supply had limits – much as we are accustomed to circuit voltages being limited to being between power supply rails.

AN ATOMIC BOMB:

I guess no one generally cites the regenerative receiver as an example of positive feedback – too complicated to explain and put in perspective. But the notion of an atomic bomb is probably the second-most common example cited in popular presentations. Now, clearly, a bomb blows up – who could say anything different. Actually, it just runs out of fuel after producing a lot of energy in a very short time. My compost pile is similar**:** once rejuvenated it produces a lot of heat for a few days and then moves along to cooler phases.

Even as I compare my compost pile to an atomic bomb, I am mindful that we judge matter and energy from a human reference. We don't see them as similar except as both, through positive feedback, accelerate and then use all the "fuel" available for a particular mode or phase. Accordingly, from a human perspective, a compost pile (for many of us anyway) is a friendly object, and an atomic bomb is unfriendly almost beyond imagination.

If we cite a phenomenon as being due to positive feedback, compare the mechanism to an atomic bomb (obviously bad), and don't mention that it is really self-limiting, it is no wonder that people assume an equation: POSITIVE FEEDBACK= DISASTER.

THE SQUEALING PA SYSTEM

After saying the atomic bomb is the second-most cited example of positive feedback, what is the first? Well, it's probably the PA system where someone is trying to talk through loudspeakers and we get some measure of squeal that makes us cringe, expecting worse, or it may outright hurts our ears. Everyone knows you turn down the volume to cure this.

Although the basic explanation of what causes this (simply "feedback") is correct and has become more accurate over the years, there is still too often a belief that the sound comes out a speaker, "goes" back to the microphone, and around and around. Further it is often stated that the pitch of the oscillation is determined by distance between the speaker and the microphone – and this is wrong. But, as often happens, we don't see the evidence as getting in the way of a good story! And this bogus explanation is so easy to envision physically.

Books are written on feedback theory and they involve a complex set of graphs and equations that describe an overall feedback path involving amplitude and phase changes, usually depending on frequency. We understand oscillation occurring (by accident, or intentionally designed) due to a magnitude of 1 and a phase of 2π around a loop. This is harder to explain to a general audience than just the sound looping around.

The sound signal as picked up by a microphone is a mixture of the desired sound (the person speaking), other man-made and natural sounds (some electrical), and the sound returning from the loudspeaker. Each frequency component has the "opportunity" to see if it is worthy of a sustained oscillation. When the amplifier is set to a low volume, likely no component makes the grade. It should be understood that the feedback path involves frequency-dependent properties of everything in the loop (amplifiers, microphones, speakers) and this includes the acoustical "frequency response" or resonances of the room. This is not flat of course. As the gain is turned up (so that everyone can hear) eventually a candidate oscillation goes over the top.

The actual squeal is in general not a pure sinusoidal oscillation, but includes a lesser amount of harmonic distortion. The oscillation builds up gradually (it may seem sudden) and overdrives against some limit (as loud as the system can produce). It is never a good thing. It does not blow up like an atomic bomb, nor is it as inoffensive (to many of us) as a compost pile. It is somewhere in between. But, in this example, even as we may not be clear on the origin of the oscillation, most of us appreciate that it can not become loud beyond bounds. [Arthur C. Clarke has a similar delightful story which must have amused him even as he surely understood he was stretching the science – which is after all what science fiction usually is.]

The idea that the squeal of a PA system is due to acoustic resonances and not the separation of a microphone and speaker is easily demonstrated by moving one or the other so that the distance changes. Same pitch. We also may see, in the more professional systems, that the amplifier involves an "equalizer" and it is possible to push down sliders so that certain frequency ranges are attenuated. In this way, we can compensate for a particular peak, and increase the volume until the next tallest resonance comes into play, or hopefully, the sound level is agreeable.

OTHER POSITIVE FEEBACKS

An engineer can often quantify as well as qualify feedback systems. Such descriptions, for simple circuits or mechanical objects, involve mathematics that corresponds closely to real physical quantities (voltage, momentum, surface area, that sort of thing). Often as well, these engineering concerns are linear. Happy days.

It is also possible to consider feedbacks for descriptions in earth sciences, biology, medicine, and even psychology or economics. These systems may be and likely are highly complex and non-linear as well. In such a case, a mathematical description in terms of feedback may not correspond well to easily measurable quantities, not be based on the correct observable parameters, and is probably no more than a limited model. Indeed, many may have multiple and competing feedback components.

Climate science issues offer a wide variety of supposed feedback issues which are probably worth looking at individually; although in conglomeration, very little of much use or surety is there. We do have the fact of very long term, relatively impressive stability of climate to suggest an overall net negative feedback. Two issues are frequently cited.

First there is the supposed positive feedback of "albedo". For example, snow is white and white reflects sunlight making it colder, and thus, more snow. In reverse, snow or ice melted is darker, and dark adsorbs sunlight making it warmer, and thus, less snow. Both seem possible, and thus we immediately understand, since the earth has some ice and some bare land/sea (and always has), that neither can be a complete story.

The second point that is sometimes tried is to suggest that while the observed changes in atmospheric CO_2 are not enough (by a factor of 3) to account for an observed warming trend, that there must be a positive feedback (as an amplification in the sense above). This involves water vapor, a known (more powerful and much much more prevalent gas than CO_2). Possibly the most logical argument in support of this idea is "We can't think of anything else"! The idea is that the CO_2 causes some warming which leads to more evaporation, and more (water vapor) greenhouse warming. Two possible things wrong with this are (1) why does not (did not) water vapor already positively feed back on itself, and (2) why did not this mechanism exist prior to about 1970 when a warming trend began (ending about 2000). So once again, the simple feedback story is incomplete and/or just wrong. For one thing, clouds and weather itself need to be accounted for.

THE FUEL LIMITS

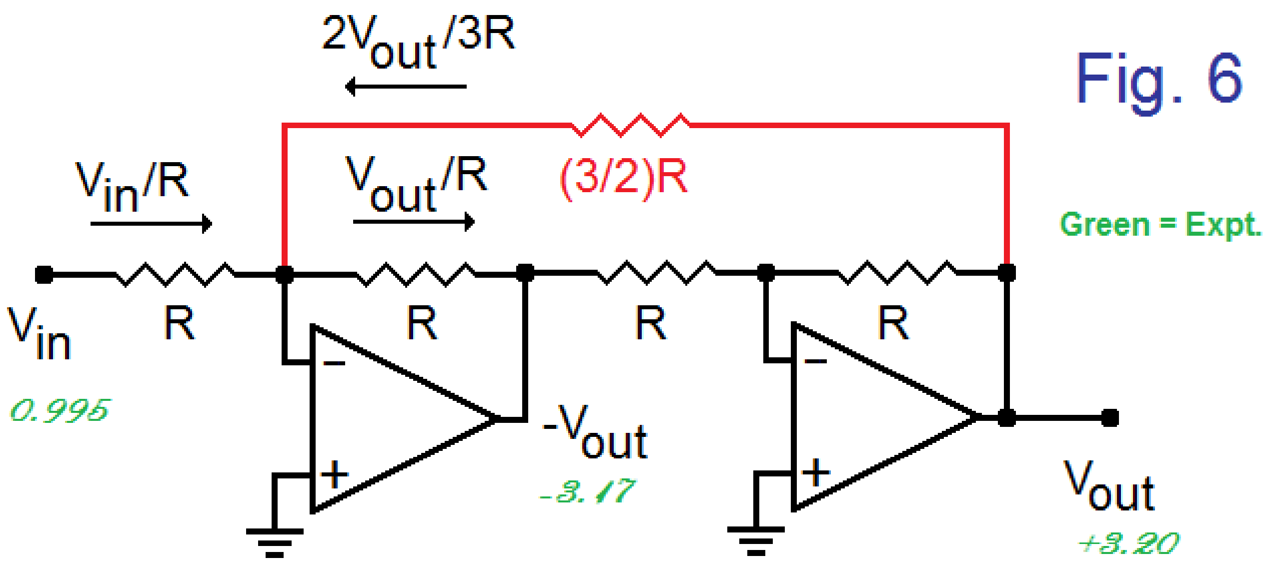
We have stated the obvious – that math is math and can run to infinity but reality is limited in the amount of "stuff" that is available. This limit may be a hard limit such as the maximum voltage we expect out of a circuit based on the power supply voltages. If we try to make a signal too big, we get "clipping" or saturation. If the signal is too large because of the amplification of a positive feedback less than 1, it still just clips. Or we can have a very large positive feedback, and intentionally drive an output to power supply rails – such as our much used Schmitt trigger circuits.

Another type of limiting is exemplified by "soft limiting" where the output is cut back, non-linearly as a set-point is approached. Various types of special effects circuits used in music are of this type. They generate harmonic distortion (they generate harmonics!) in a gradual manner as the signal gets larger.

An interesting system in climate science relates to the heat-trapping effects of CO_2. The amount of heating due to this "greenhouse effect" is limited first because of the fact that all heating is finite. Secondly, it is limited because if temperature tries to rise, various limiting (thermostatting) mechanisms will kick in to obey the 2nd Law of Thermodynamics. That is, nature "invents" (must invent) mechanisms to remove excess heat. Thirdly, the heat trapping is limited to photons that have a wavelength that can be trapped by CO_2. And fourthly, the effect has been known for a century or so to be soft-limited – logarithmic in fact. While CO_2 molecules are extremely rare in the atmosphere (1 in 2500, a puny penny in $25, a puny penny in 50 rolls of pennies!) there are already enough so that most photons in the CO_2's bandwidth are already intercepted as they try to escape to outer space. Suppose you have a baseball team and you find that your league has no limit on the number of outfielders you can field. So you place a large number on a 10 foot grid. Not many balls dropping. Still you double the number of outfielders yet again. Very few additional balls caught. Certainly not twice as many. This sort of logarithmic effect is well known in many instances in physics.

We speak of this fuel limit to make the point about limited stuff. At times, we engineer in a limit to flow, and sometimes call this an example of negative feedback. Nothing is more classic in this regard than the centuries old steam engine "governor". In this instance, a mechanism is attached to the engine shaft which rotates with it, and masses in the mechanism then move away from the shaft by centrifugal force, opposed by springs (and/or gravity). At some point as it speeds up (possibly from the very start), as the masses spin yet further away, the steam supply is reduced. Some may feel that this is not a valid form of negative feedback because it may be non-linear. Indeed, probably we want no feedback at the very start of the spinning (fast start-up), and more feedback as the rotation reaches the desired value, along with minimal lags when the load changes.

SOME FAMILIAR OP-AMP CIRCUITS

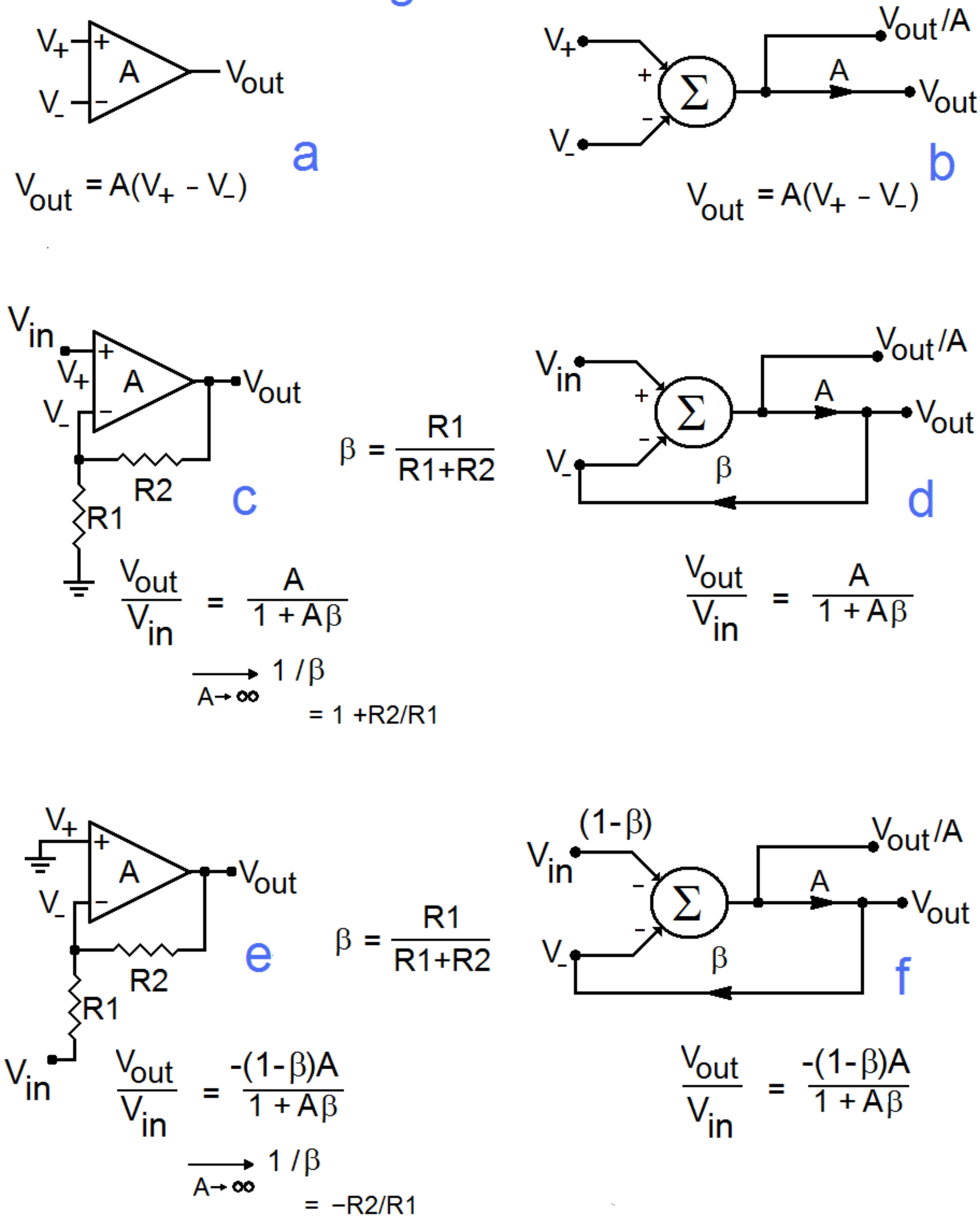


Positive Feedback Increases Gain to 3

It is probably most useful for the readers of our notes to have examples that relate to things we design easily and have constructed, so here we will look at some examples using ordinary op-amp circuits. We will begin with a case where we have an amplifier of unity gain and want to increases the gain to 3 using positive feedback. This is not intended as a practical circuit but as a very simple illustration (Fig. 6).

In Fig. 6 first note that if the red resistor is removed, we just have two unity-gain inverters in series. Changing the gain to 3 would be a simple matter of changing either of the two R resistors in the feedback paths of the op-amps to 3R. Here in using our notion of an overall positive feedback, equation (5) is equal to 3 if g = 2/3. Adding the red resistor of (3/2)R from the second stage provides this positive feedback. Does this give us a gain of 3? The three currents at the left summing node sum as:

Fig. 7



V+
V-
A
Vout
a
Vout = A(V+ - V-)
Σ
Vout /A
b
Vin
R2
R1
c
β = R1 / R1+R2
β
d
Vout / Vin = A / 1 + Aβ
A→∞ 1 /β
= 1 +R2/R1
e
(1-β)
f
Vout / Vin = -(1-β)A / 1 + Aβ
= −R2/R1

$$\frac{V_{in}}{R} + \left(\frac{2}{3}\right)\frac{V_{out}}{R} = \frac{V_{out}}{R} \quad (6)$$

Which solves for V_{out}/V_{in} = 3 as we had anticipated would be the case.

In Fig. 6 we have assumed ideal op-amps with standard "summing node" analysis and have more-or-less just realized a block diagram type system. By far the more traditional presentation of feedback involves feedback around an existing amplifier stage, and a finite gain differential amplifier is a good choice to examine now.

In Fig. 7a we show the standard op-amp noting that here it is non-ideal by giving it a finite (but – we usually assume – very large) gain of A. [We have noted many times that this finite gain model is not too useful, except as it shows that very high gains (say A = 1,000,000 as compared to A=100) make little practical difference to the gain analysis. And, we can and often do make use of a realistic frequency dependent model A = A(s) = G/s where G is the gain-bandwidth product by just doing a substitution.]

Fig. 7b shows a block diagram of the corresponding open-loop amplifier. We know that op-amp applications are usually closed-loop with negative feedback, comparators and Schmitt-triggers (see Fig.16 below) being exceptions. Fig. 7c shows the familiar non-inverting amplifier as our typical example. Here the finite gain differential amplifier is defined by:

$$V_{out} = A(V_+ - V_-) \quad (7)$$

and we note that the feedback loop voltage-divider establishes:

$$V_- = \left(\frac{R1}{R1+R2}\right)V_{out} = \beta V_{out} \quad (8)$$

And thus the closed-loop gain is:

$$\frac{V_{out}}{V_{in}} = \frac{A}{1+A\beta} \quad (9)$$

This, in the limit A→∞ is $1/\beta$ = 1 + R2/R1 which is the well-known ideal op-amp result. The fact that A is part of the "loop gain" is clearly seen in the Fig. 7d, block diagram equivalent of Fig. 7c.

The feedback gain β is, as in Fig. 7d from the output back to the input. Here is an important point: In Fig. 1c, it appeared as though the gain g was from the output of the summer back to the input. We should always consider the output of the summer to feed into an amplifier, even when the amplifier is of gain A=1 (a "wire"). This may be a source of confusion. As mentioned, a frequency dependent model such as A = G/s can also be used.

Fig. 7e shows the corresponding inverting amplifier. The analysis is familiar, using equation (7):

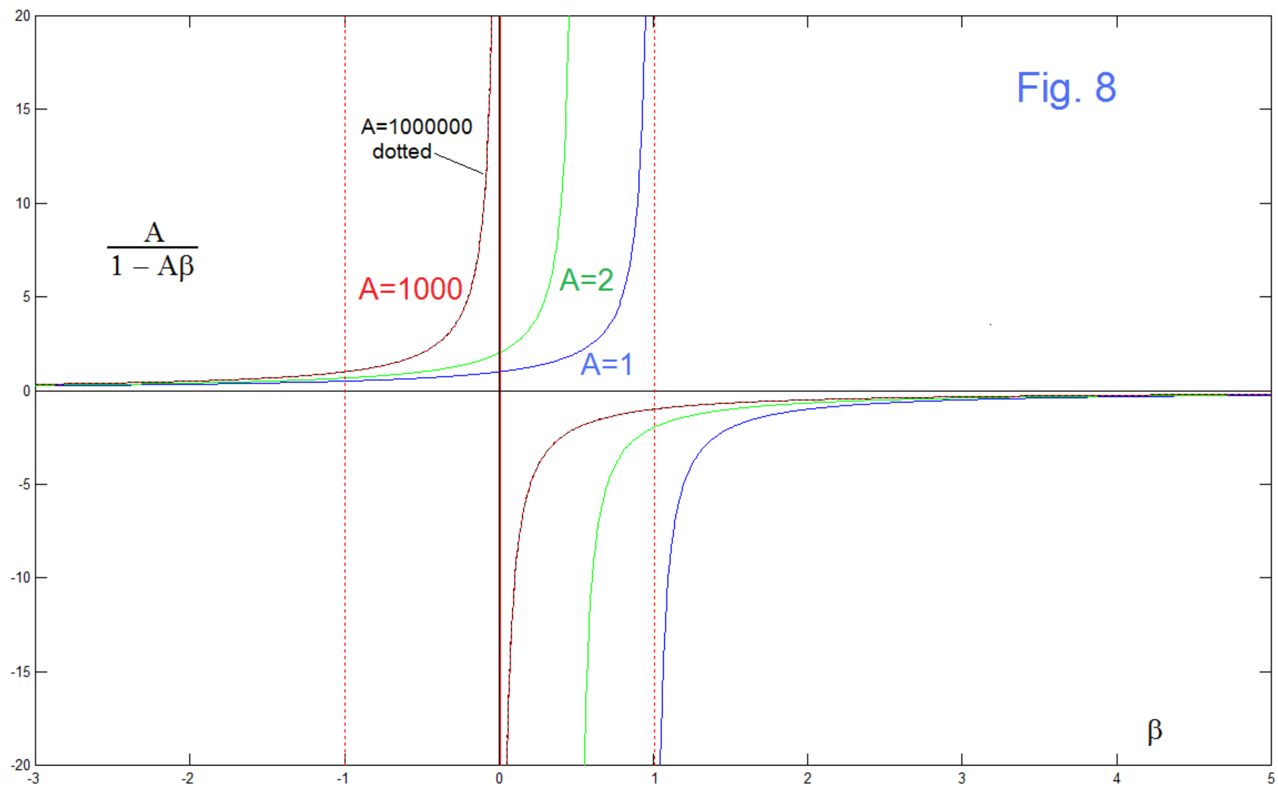
$$V_{out} = -A\left(\frac{V_{in}R2}{R1+R2} + \frac{V_{out}R1}{R1+R2}\right) \tag{10}$$

which can be solved as:

$$\frac{V_{out}}{V_{in}} = \frac{-(1-\beta)A}{1+A\beta} \tag{12}$$

and in the limit A→∞ we have the familiar inverting gain -R2/R1.

We now have three cases, Fig. 6 where we somewhat artificially examined an amplifier with a gain of 1 (and have used positive feedback), and the two more-useful op-amp circuits (Fig. 7c and Fig. 7e), both of which use negative feedback. Do not conflate positive/negative feedback with inverting/non-inverting. Further, positive or negative feedback can change with the sign of β equally as well as the selection of the input polarity of a summer.



In the development above we have replaced the feedback factor g with the "loop gain" which can be seen to be βA. In the examples of Fig. 7c and 7e, β is a voltage divider and is thus limited to values from 0 to +1. We need to see how the gain curves which begin with Fig. 4 change as A and β change.

The curves shown in Fig. 8 show that when A changes from 1 (as in Fig. 4) to 2, there is a drastic shift (blue to green). This continues although much less drastically, and A=1000 (red) is virtually the same as A=1,000,000 (dotted black). Clearly for large A the gain function is essentially -1/β. However, note the interesting limit of A/Aβ as A gets large, but β can still go to zero (at β=0 in Fig. 8).

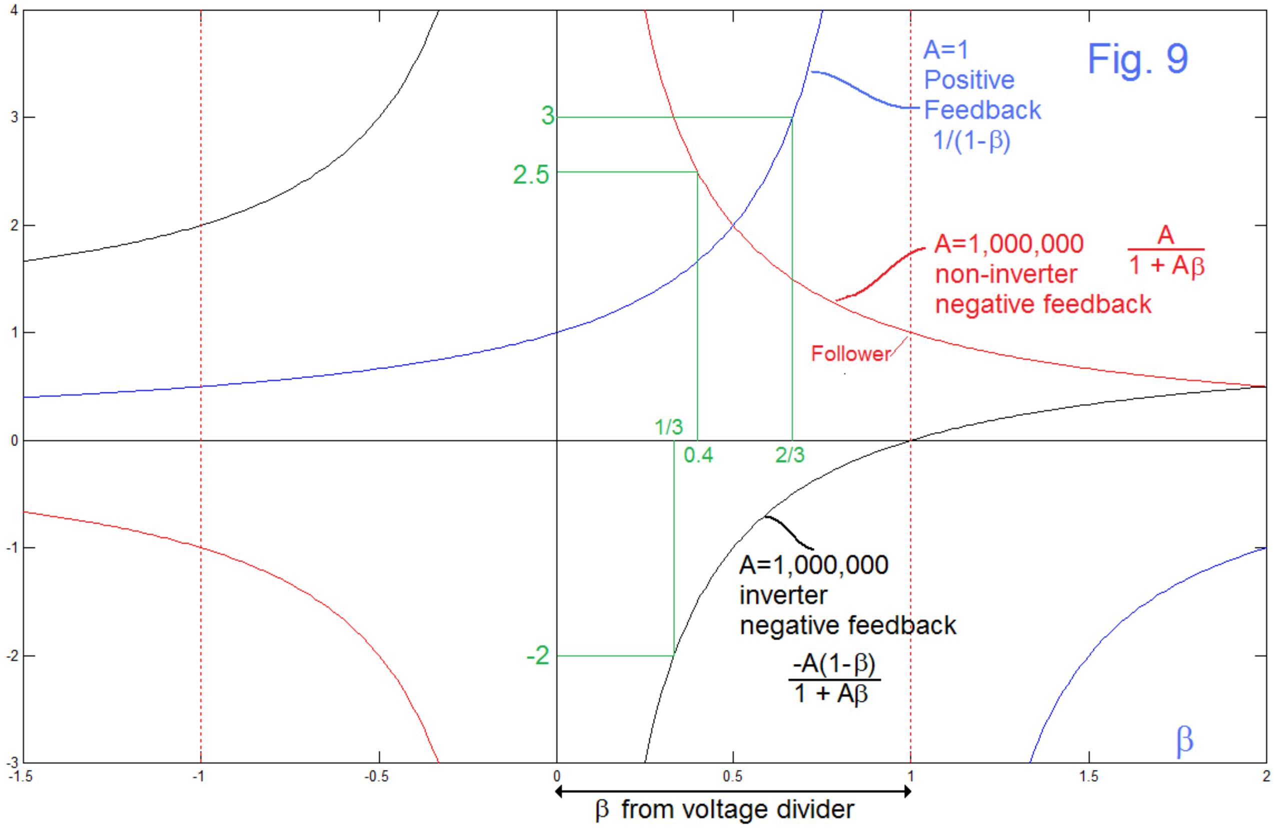


Fig. 9 shows the cases of Fig. 6 and Fig. 7c and Fig. 7e. Here we also have A=1 and A=1,000,000 (very effectively standing in for A = infinity). The red curve here corresponds to equation (9) so is reversed from Fig. 8. We have added as well the inverter of equation (12), the black curve.

THE LIMITS ON β

Because the circuits of Fig. 7c and Fig. 7e have β derived from a voltage divider, β is limited in these two cases to a range from 0 to +1. We can likely always "contrive" a system to test feedback factors in most any range we want to try. As an example, suppose we want to try using negative feedback to decrease the gain of Fig. 6 instead of the positive feedback increasing the gain as we have shown. We could use a third op-amp in an inverter, producing $-V_{out}$ and add a resistor back to the first summing node from that (Fig. 10a). For example, if we used the red (3/2)R resistor from our proposed third stage, we would have β = -2/3 and this 1/(1-β) = 3/5 = 0.6, a gain less than 1 here. (We leave it to the reader to verify the familiar analysis here and in Figures 11, 12, and 13 below). Soon enough we would see that we might as well have just connected the red resistor to the output of the first rather than the third stage, as both are $-V_{out}$. This in turn could be combined with the resistor R already in the (negative) feedback of that stage, for an equivalent resistor of (3/5)R (Fig. 10b). This would give the desired gain without an extra op-amp. We didn't need fancy theory to get this! Note that this point is on the blue curve to the left of β=0. Indeed we can work this negative range of β as

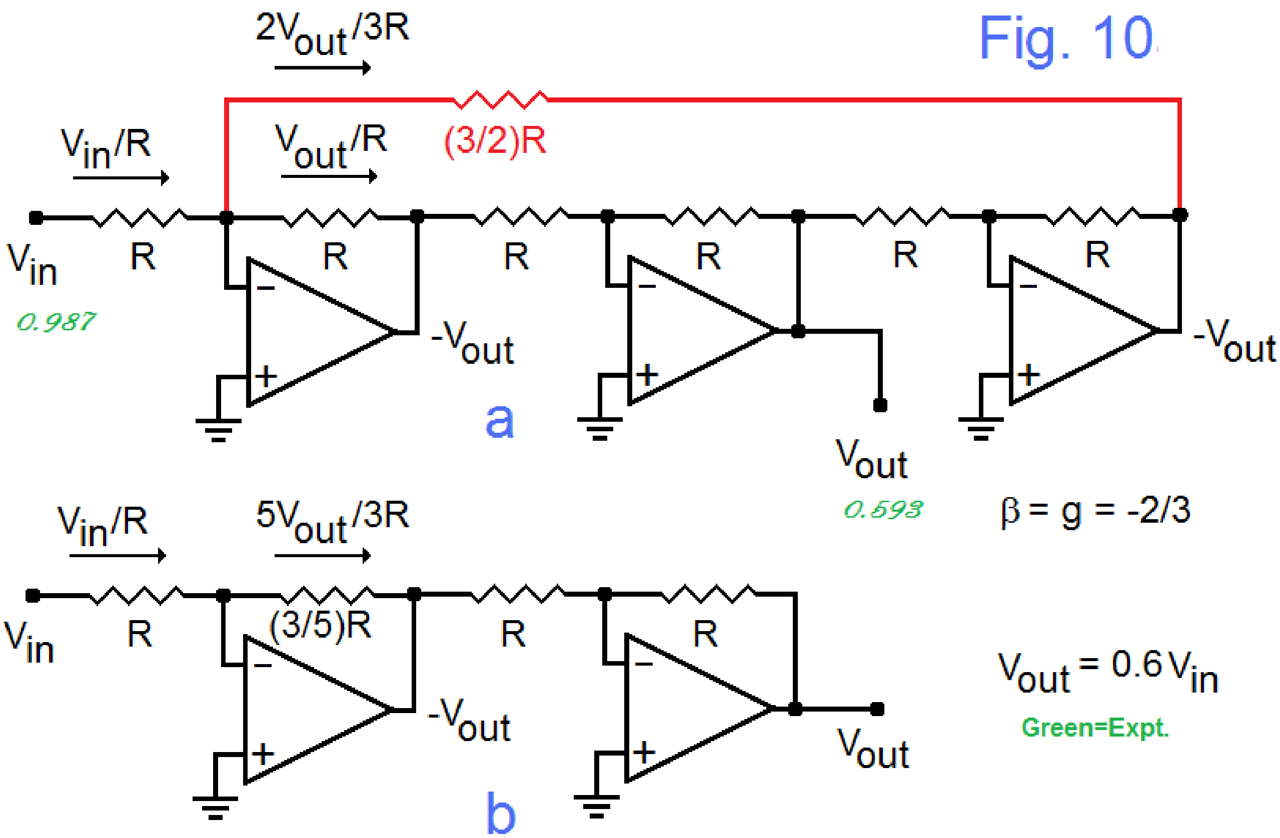

2Vout/3R
Vin/R
Vout/R
(3/2)R
Vin
R
R
R
R
R
R
0.987
-Vout
-Vout
a
Vout
0.593
β = g = -2/3
Vin/R
5Vout/3R
Vin
R
(3/5)R
R
R
-Vout
Vout
Vout = 0.6 Vin
Green=Expt.
b

Fig. 10

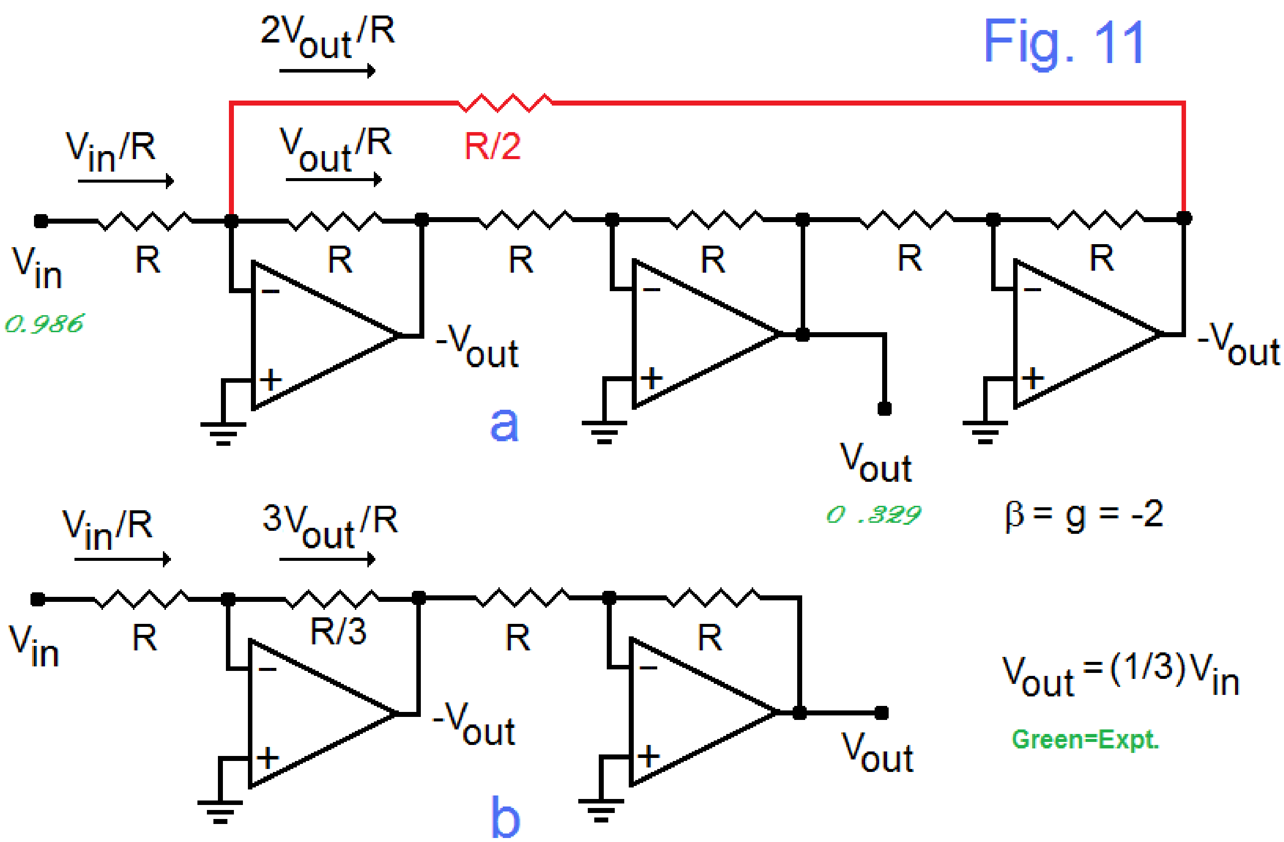

2Vout/R
Vin/R
Vout/R
R/2
Vin
R
R
R
R
R
R
0.986
-Vout
-Vout
a
Vout
0 .329
β = g = -2
Vin/R
3Vout/R
Vin
R
R/3
R
R
-Vout
Vout
Vout = (1/3) Vin
Green=Expt.
b

Fig. 11

far left as we could want, just making the feedback resistor smaller and smaller. Fig. 11 shows the situation similar to Fig. 10 where the feedback is -2 (off scale in Fig. 9) instead of -2/3.

Thus we see that our analysis gives stability for all negative feedbacks (β<0), and an amplification for positive feedbacks if 0 < β < 1. This constitutes the blue curve for positive values of gain (upper half) of Fig. 9. We can easily see that the gain function blows up to ∞ at β=1 and comes back from negative ∞ as β goes more positive then +1 (Fig. 9 and indeed, Fig. 4). Does this happen with actual op-amp circuits? Well – NO.

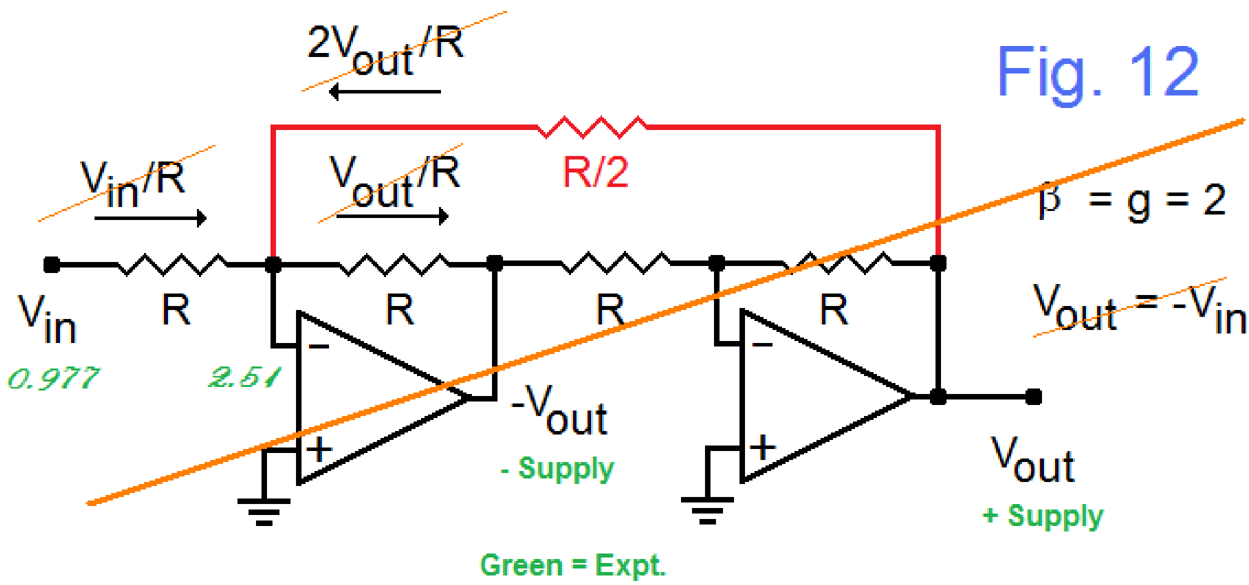
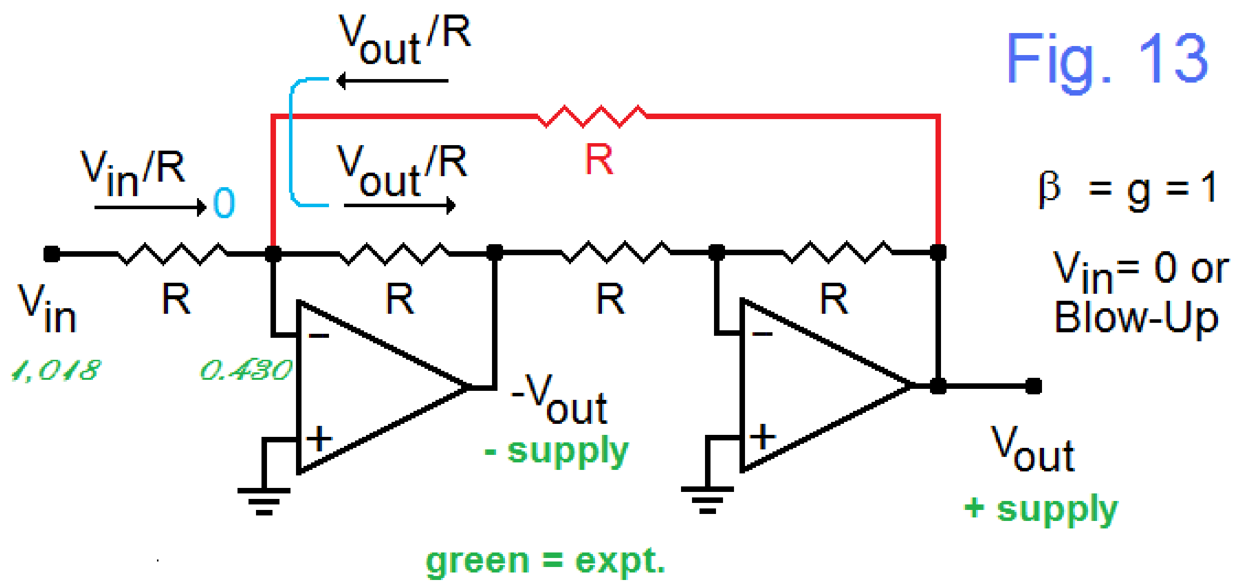


Fig. 12

Fig. 12 shows the case of a positive feedback of β=2. In this case, 1/(1-β) = -1 (inverting). The analysis shows that this really does seem to work out that way, and so it seems that values of β>1 are stable? But the analysis here assumes valid summing nodes. The green experimental data shows clearly that this is not so, as does the fact that the op-amp outputs are pinned at supply rails. So the inversion suggested in the ideal analysis (crossed out) is invalid because the feedback is the opposite of what we first thought) we did get a negative output after all). In consequence, β < 1 is the stable region, and the blue curve in the lower right of Fig. 9 is invalid.

We certainly expect a range of β approaching +1from the low side, in a practical case, to have so much gain that saturation against power supply levels is virtually certain. Fig. 13 shows this case of exactly β=1, and it is curious. Here we have chosen the red feedback resistor as exactly R for β=1. Here summing node analysis is still (just) valid, but impractical. Thus we see the current V_{out}/R comes back through the red R resistor (into the summing node) and the same exact current flows (blue path) out through the R resistor in the feedback of the left op-amp, out of the summing node). The current summing to zero is thus met. Fine, but we can't have this upset by any current V_{in}/R from the input. Thus, V_{in} must be zero, which makes the circuit useless of course, and V_{out} is otherwise arbitrary. In fact, if V_{in} is set to anything other than 0, the summing node of the left op-amp is not 0, but the sum of V_{in}, $-V_{out}$, and V_{out} divided by 3, and thus $V_{in}/3$. This input on the (–) terminal of the left op-amp would saturate the output. In a practical case, because the resistors R are not all exactly equal (and for offset considerations), even with $V_{in}=0$ the outputs V_{out} will be power supply saturation levels.



STABILITY FOR β<1

Here we have looked at feedback (both positive and negative), in terms of gain changes, and in terms of the stability of a system using it. The feedback factor was called g, or Aβ, or just β in the not unusual case where A≡1. We have also looked at the cases where the feedback path includes a delay, shown as a z^{-1}, in the feedback path. We found several things:

(1) In the case of the delay, discrete-time network theory limited stable values of g to the range $-1 < g < +1$ as this corresponded to a network pole inside the unit circle in the z-plane.

(2) When looking at some common op-amp circuits which had an amplification element A (very large) we had useful structures (Fig. 7c and Fig. 7e) where β was limited to the range 0 to +1. This limitation was only due to the fact that it was set by a voltage divider. This was applied to negative feedback, but could be applied to positive feedback (see Schmitt trigger in Fig. 16 below).

(3) Without the delay, we found just above that apparently the values of β that were <u>not</u> allowed was β=1 or greater. Negative β, to include β = -1, was stable.

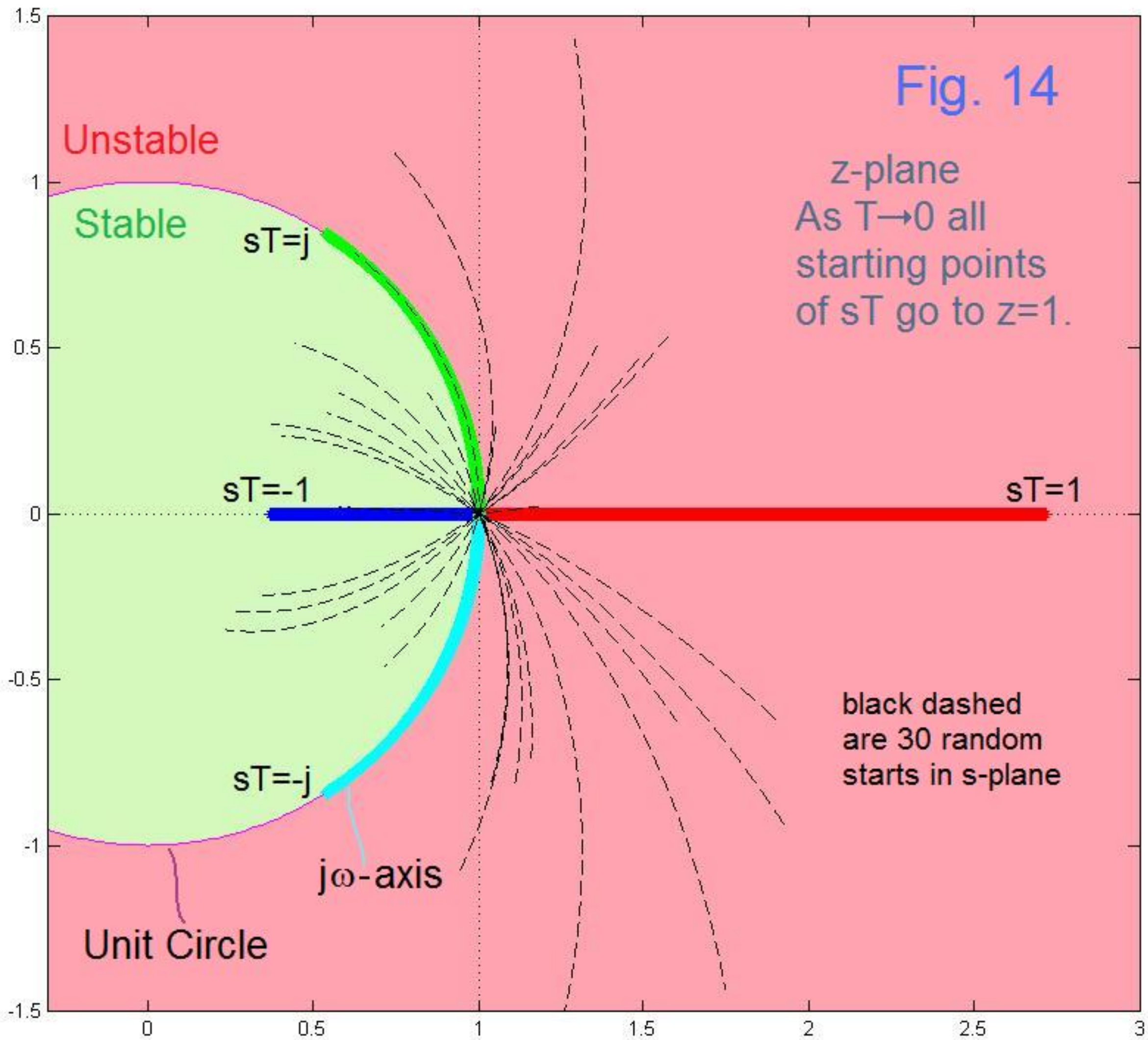
Why the difference between findings (1) and (3)?

If the difference is whether or not a z^{-1} is included in the feedback loop, let's look at what z^{-1} means. We had equation (4) repeated right here:

$$z^{-1} = e^{-sT} = e^{-j\omega T} \qquad (4)$$

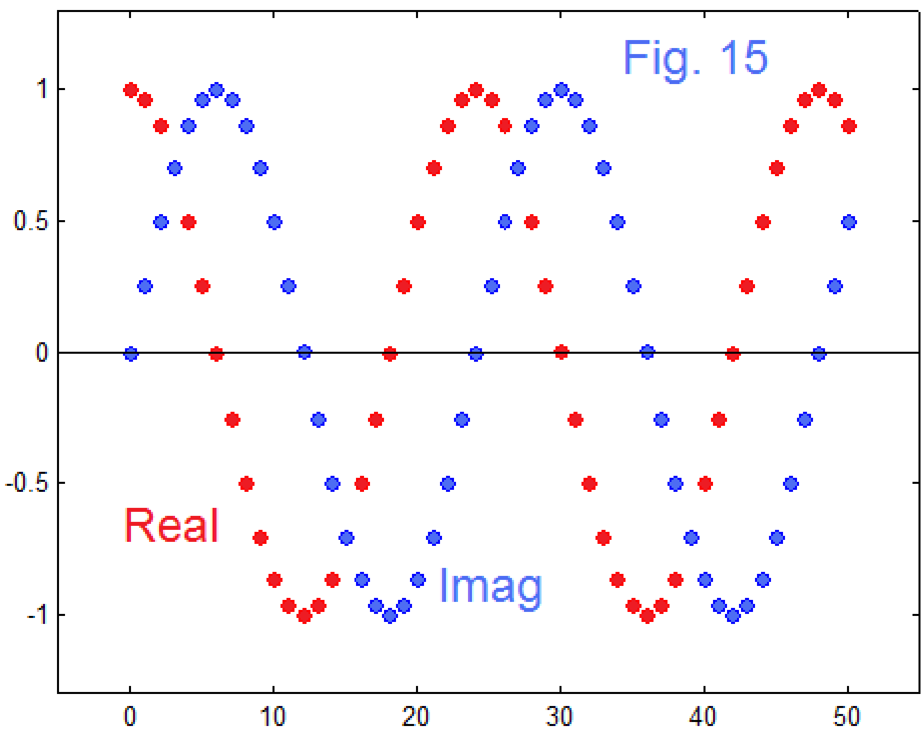
Accordingly, removal of z^{-1} can be accomplished when $z^{-1} = 1$, which is when $e^{-j\omega T} = 1$.

Here we admittedly "hand wave". We can make $e^{-j\omega T} = 1$ with either $\omega \to 0$ or $T \to 0$ (or obviously, both). We note with interest that ω=0 suggests that we look at the DC response of a system. We did this in equation (5) for the frequency response, and got our 1/(1-g) with only a single singularity at g=1. Nothing special anywhere else. Keep in mind that ω=0 is usually thought of as a DC input, and this is a property of an input. The delay T on the other hand is a property of the system, and likely the alternative $T \to 0$ is more in line with exactly what we have in mind in making the delay disappear. This would correspond to a sampling frequency becoming infinite. Thus once around the unit circle in the z-plane is an infinite distance! <u>Compared to ω=∞</u>, all finite frequencies are the same as 0. Frequency-wise, you never get to go anywhere.



On the one hand, we have looked at this from the point of view of moving from a well understood first-order discrete-time (with the delay) network, with its associated unit circle of stability, back to the case where the delay disappears. Accordingly, we ask where the circle went and where the instability at g =-1 "went". Some possible insight into this mystery can be seen in Fig. 14. In one sense, this is the conventional mapping of the s-plane into the z-plane, Here however we have started with 34 points in the s-plane. Four of them are at +1. -1, +j, and –j, and the other 30 are randomly selected points. Here we have chosen not s, but sT. We then consider T to vary from +1 down to 0, thus collapsing the s-plane location from a starting sT to a finish at sT=0, and a corresponding single destination at z=1 in the z-plane. The migration is from the ends of the various lines (away from z=1) to z=1. The four special points have wide, colored traces. Note that the starting points at +j and –j (on the jω-axis) map into the unit circle (green and light blue). The 30 random starts are represented by dashed black lines. Making T go to 0 thus collapses the s-plane to a single point in the z-plane at z=1.

Now if we reverse our thinking and consider, as is often done, the delay-less case as the start, we see the appearance of a blow-up at z=-1 not just as occurring at the point, but as the appearance of the entire unit circle. Because we only had a first-order system, we only saw the possibilities of the unit circle at z=1 and z=-1. We can however have a first-order pole (unconjugated) anywhere on the unit circle. For example, Fig. 15 shows the sustained oscillation (real and imaginary parts) when a single pole is place on the unit circle at 15° [g = cos(π/12) + j sin(π/12)].

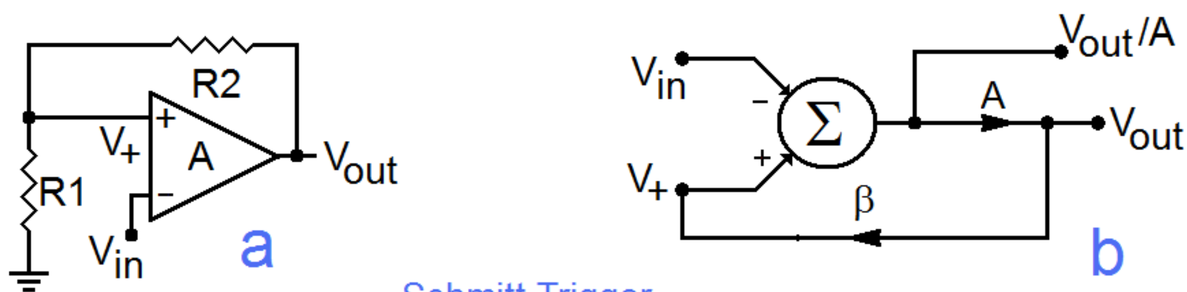


From this perspective, we see that the finding of an unstable point at g=1 in the delay-less cases, and indeed for g > 1 (e.g., Fig. 12), is not so much of a mystery. And, we pretty much understand that there is no corresponding instability for g=-1, or for any negative g. However, the feeling we probably leave here is that something is not totally "tied up" and we invite readers to contribute. A brief graphic summary of this point is in Appendix B.

INTENTIONAL POSITIVE FEEDBACK EXCESS

Fig.16 shows an additional example – a circuit with a value of positive feedback much greater than 1 – indeed a Schmitt trigger. It is the same as Fig. 7c with the op-amp terminals reversed. This circuit is non-linear (it saturates at the power supply levels) and has hysteresis, so we can't use our usual equations. We do however find numerous applications of this, and other non-linear op-amp circuits.

Fig. 16



Schmitt Trigger

SUMMARY

Engineers are likely to use systems with negative feedback in most every project. If not that, we use it when we drive a car and keep it on the road. It's everywhere in our technology and indeed in nature. When an engineer sees long-term stability, he/she thinks negative feedback, not tweaking.

We think less about positive feedbacks. So when someone else uses the term, we need to stop and think about it – understanding it ourselves – and if it can be thereafter explained to a general audience.

Here we have spoken of many generalities. Most importantly perhaps the fact that a positive feedback does not necessarily mean a "blow-up" or run-away" but often just some amplification. Further, many proposed "run-aways" are "fuel limited" or saturated with respect to the amount of "stuff" such as energy that is available. Any mention of a "tipping point" (physical – not just mathematical) should include an inquiry as to the total resources that we would expect could ever come into play, and generally there are multiple feedback mechanisms in play as well.

One last point is that we may expect errors if we try to rely on memory and general considerations. These systems can be complicated. On the other hand, feedback theory is often an alternative path a fuller understand what we do already. The op-amp circuits are a good illustration of this. It gave the answers we already knew.

APPENDIX A - ROGRAM CODE EXAMPLE

No very sophisticated code here. Mostly just making pictures. The Matlab code below made the basis of Fig. 9.

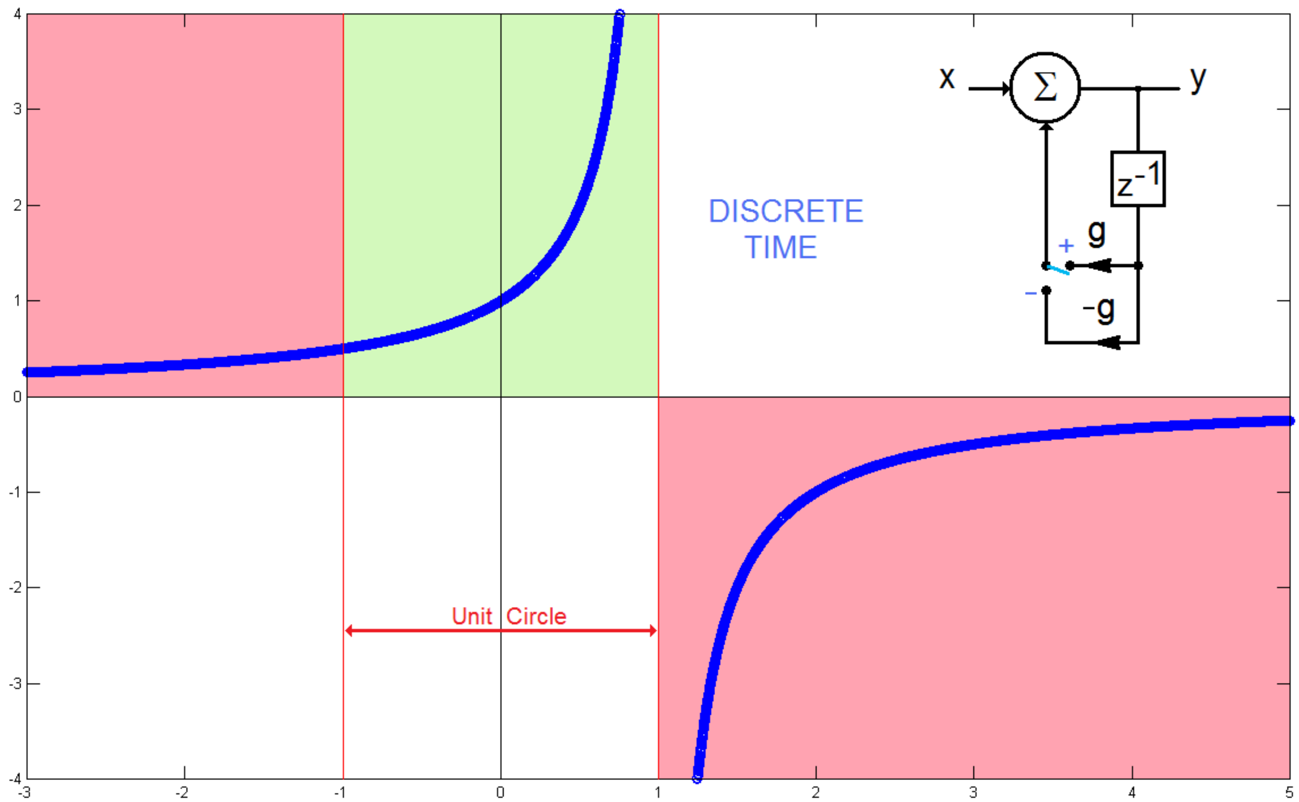
```
 % Gain of 1 amplifier – Blue Curve
k=1
A=1
for beta=-3:.01:5
   gain(k)=A/(1-A*beta);
   k=k+1;
end

plot([-3:.01:5],gain,'b')
hold on
plot([-3 6],[0 0],'k')
plot([0 0],[-20 20],'k')
plot([-1 -1],[-20 20],'r:')
plot([1 1],[-20 20],'r:')

% Non-inverting amplifier – Red Curve
k=1
A=1000000
for beta=-3:.01:5
   gain(k)=A/(1+A*beta);
   k=k+1;
end

figure(11)
plot([-3:.01:5],gain,'r')
hold on
plot([-3 6],[0 0],'k')
plot([0 0],[-20 20],'k')
plot([-1 -1],[-20 20],'r:')
plot([1 1],[-20 20],'r:')

 % Inverting Amplifier – Black Curve
k=1
A=1000000
```

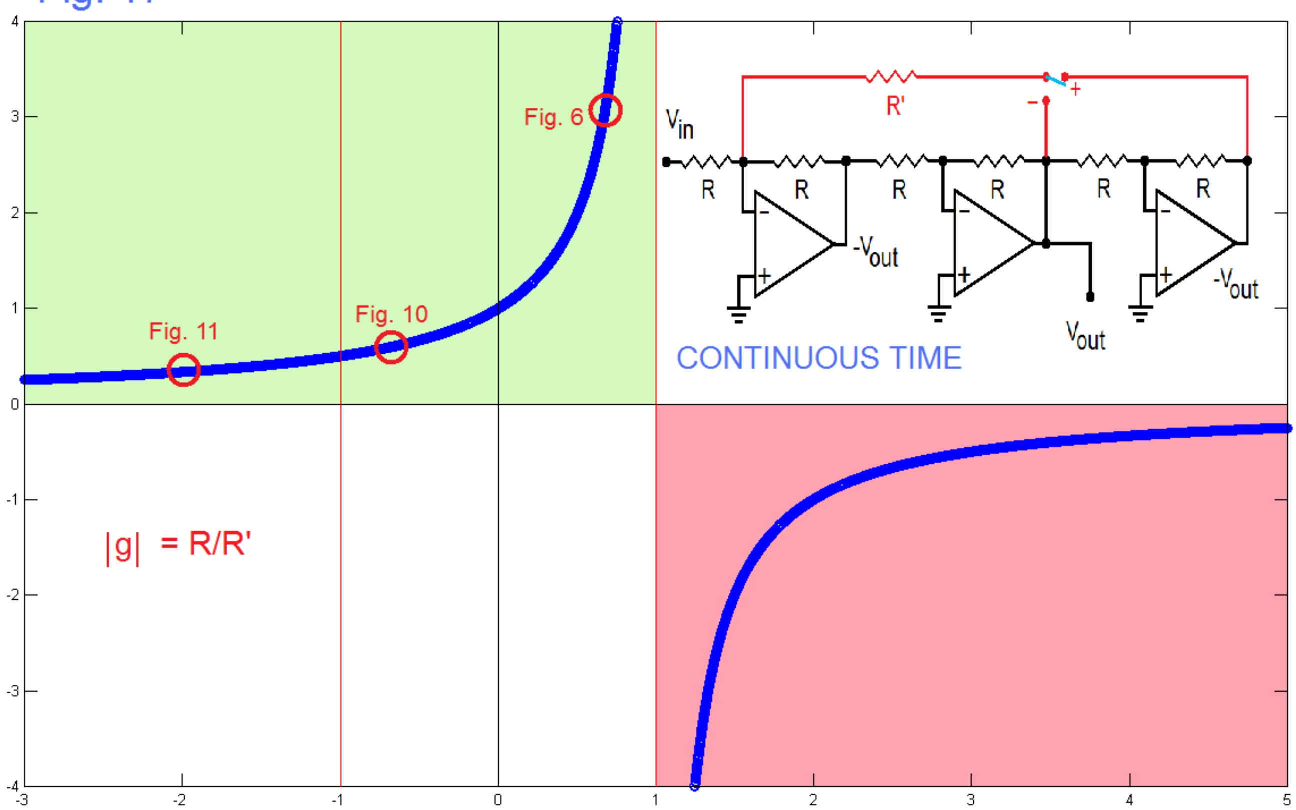
```
for beta=-3:.01:5
   gain(k)=-A*(1-beta)/(1+A*beta);
   k=k+1;
end

plot([-3:.01:5],gain,'k')
hold on
plot([-3 6],[0 0],'k')
plot([0 0],[-20 20],'k')
plot([-1 -1],[-20 20],'r:')
plot([1 1],[-20 20],'r:')

% Example boxes - Green
plot([2/3 2/3],[0 3],'g')
plot([0 2/3],[3 3],'g')

plot([.4 .4],[0 2.5],'g')
plot([.4 0],[2.5 2.5],'g')

plot([1/3 1/3],[0 -2],'g')
plot([1/3 0],[-2 -2],'g')

hold off
axis([-1.5 2 -3 4])
```

APPENDIX B – SUMMARY COMPARISON

Fig. 17 is provided as a "summary" of the findings and remaining puzzles. The top panel shows the discrete-time case, and the bottom one the continuous-time case, both with reference to the same 1/(1-g) curve (blue). The networks show switches to implement either positive or negative feedback. The pink color show unstable regions while the green color shows stable regions. The only difference is in the upper left region of the panels. In both cases, the region of g>1 is unstable, while –1 < g < +1 is stable. Examples in Fig. 2, Fig. 3 and Fig. 5 illustrate the actual responses for discrete-time, while Fig. 6, Fig. 10 and Fig. 11 (as shown) illustrate the continuous-time case with experimental verification in the referenced figures.

While the top panel actually corresponds to the DC gain, we have seen the removal of the delay corresponds to the bottom panel, with the resulting gain or attenuation being little more than a constant valid except at g=1.


DISCRETE
TIME
x
y
z-1
g
-g
Unit Circle
Fig. 17


Vin
R'
R
-Vout
Vout
CONTINUOUS TIME
Fig. 6
Fig. 10
Fig. 11
|g| = R/R'