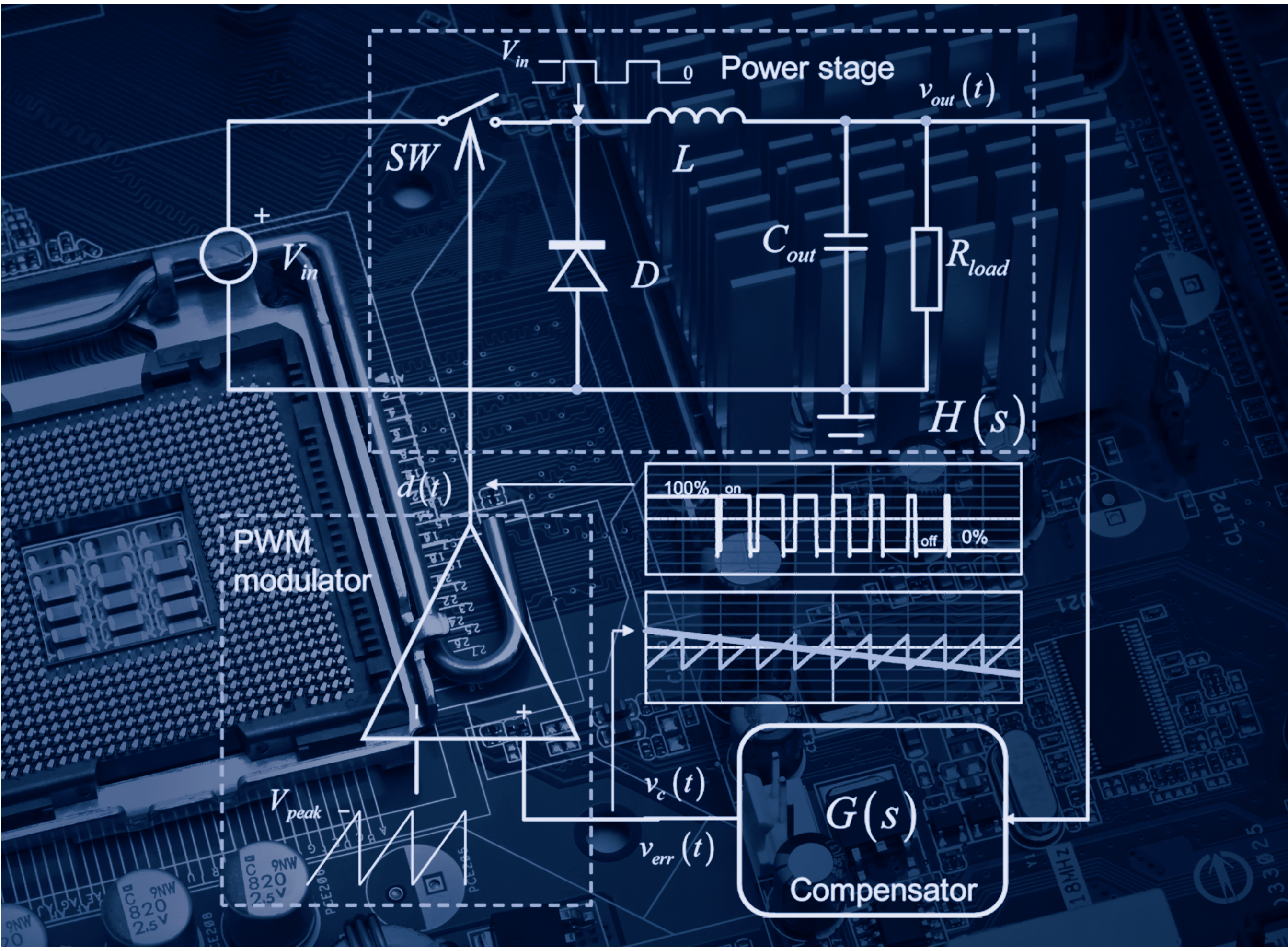
# DESIGNING CONTROL LOOPS for LINEAR and SWITCHING POWER SUPPLIES

## A TUTORIAL GUIDE




CHRISTOPHE BASSO

# Designing Control Loops for Linear and Switching Power Supplies

## A Tutorial Guide

Related Artech House Titles in Power Engineering

*The Advanced Smart Grid: Edge Power Driving Sustainability,*
Andres Carvallo, and John Cooper

*Battery Management Systems for Large Lithium Ion Battery Packs,*
Davide Andrea

*Designing Control Loops for Linear and Switching Power Supplies: A Tutorial Guide,*
Christophe Basso

*Electric Systems Operations: Evolving to the Modern Grid,*
Subramanian Vadari

*Energy Harvesting for Autonomous Systems,*
Stephen Beeby, and Neil White

*Power Line Communications in Practice,*
Xavier Carcelle

# Designing Control Loops for Linear and Switching Power Supplies

## A Tutorial Guide

Christophe Basso

ARTECH HOUSE
BOSTON | LONDON
artechhouse.com

**Library of Congress Cataloging-in-Publication Data**
A catalog record for this book is available from the U.S. Library of Congress.

**British Library Cataloguing in Publication Data**
A catalog record for this book is available from the British Library.

ISBN-13: 978-1-60807-557-7

**Cover design by Vicki Kane**



10 9 8 7 6 5 4 3 2 1

# Contents

## CHAPTER 4

## CHAPTER 5

## Operational Amplifiers-Based Compensators 253

# Foreword

In his previous book, *Switch-Mode Power Supplies: Spice Simulations and Practical Designs*, Christophe Basso gave valuable, numerous, and detailed explanations of converter topologies and PFC circuits by way of practical circuit designs. Although modeling and feedback control were well explained and utilized in that book, the emphasis was on SPICE simulations.

In this book, Christophe has delved deep into understanding and analyzing specific control circuits used with power converters. He has examined in detail the performance of converters ranging from dynamic load response and stability to line rejection for various controllers, including some unconventional ones for educational purposes to emphasize some of the subtle aspects of feedback control methods. He has derived numerous analytical results to design a compensator that achieves a particular type of performance goal or stability criterion.

In addition to the rigors of control theoretic aspects of this book, I was pleased to see the analysis of converter circuits being carried out using fast analytic circuit techniques with answers in low-entropy form very much in the spirit of Dr. R. D. Middlebrook. If design is the reverse of analysis, Middlebrook would say, then the only kind of analysis worth doing is design-oriented analysis, which yields low-entropy expressions. Christophe has taken this to heart and done a very fine job with it.

Such a book is necessary because of the continued isolation and demise of the field of analog electronics of which power converter design is part. Control theory books abound, whereas analog electronic textbooks are increasingly becoming less adequate and cookbook like. For the new graduate in electrical engineering who wants to work in the field of power electronics, the necessary knowledge to get started is diffused and rather difficult to consolidate. Good sources exist to teach one-week courses to practicing engineers, but these assume some working knowledge or experience with power converters. This book in my opinion fills the gap between control theory and converter design *rigorously*. Also, that such a book is necessary in my opinion is based on my experience, albeit limited, with a few control problems I have had to work on in addition to designing power converters full time. I have seen control systems or rather control schemes that are far too elaborate and unnecessary for stabilizing a platform or regulating the temperature of a laser. Their performance, often marginal, could have been tremendously improved by designing a far simpler feedback control circuit if only the designer understood the equivalent circuit model of the transducer in the first place. Christophe has done an excellent job in exposing the concept of modeling and rigorously compensating converter circuits. It is a book that I recommend to all power electronics engineers who, every once in a while, may have to tackle a control problem for a device that is not a power converter.

*Vatché Vorpérian*
*Jet Propulsion Laboratory*

# Preface

When I started this book in January 2009, I had the intention to write a quick booklet exclusively covering compensator structures. The idea germinated as I realized that most of the available documents covered compensator examples implementing operational amplifiers. This is the way I learned how to stabilize a loop at university in the eighties. Later, as an engineer, I wanted to put my knowledge at work with a TL431 or a transconductance amplifier to which an optocoupler was hooked. As you can imagine, the connections were missing between my school books and the practical circuit I was working on.

Literature does not abound on the subject of compensator structures, so I had the choice to either plunge into analytical analysis or start tweaking the circuit through trial and error. Obviously, the second approach was wrong, but when time pressure becomes an unbearable situation, I understand that engineers have no other alternative for their ongoing design. I felt there was a gap to fill in the technical literature to show how compensation theory could apply to electronic circuits different than op amps. On the run, I wrote Chapters 5, 6, 7, and 8. Then, I decided to write a little about loop control theory, something engineers could use to refresh their memory on the topic. As I wrote Chapter 1, I discovered that most of the theory I knew on the subject was not really what I had been taught at school. In fact, when I graduated from Montpellier University (in France, not Maine!) and tried to apply that fresh knowledge to a project, I was stuck: I could not bridge the stuff I knew to what I was asked to do. My teachers talked about PID coefficients, and I had to place poles and zeros.

I want this book to be the companion you look at when you need to stabilize a power converter. For that purpose, I have tried to balance the useful theory—you do not need to know everything in the field of loop control to be a good engineer—and the necessity to make it work on real projects. In the nine chapters I wrote, Chapter 1 starts with generalities on the subject. If you are a beginner, you must read it. Chapter 2 introduces transfer functions and the formalism to write them correctly. Fast analytical techniques are used throughout this chapter, and I encourage you to dig further into the subject. Chapter 3 is an important part of the book, as it details the stability criteria to build rugged control systems. Back to my university time, I had been told to maintain the phase margin to at least 45° and that was it, with no further explanations or origins of this number. There is nothing new here, but I have derived the equations so you can link phase margin numbers and the expected closed-loop transient performance. The same applies to the crossover frequency that you will no longer arbitrarily select. Chapter 4 explains compensation basics, starting with the PID blocks and expanding to what you will deal with: poles and zeros placement. Then I introduce several compensation methods, including output impedance shaping for high-speed dc-dc converters.

The next chapters, 5, 6, 7, and 8, teach you how to compensate your converter with an op amp, a TL431, a transimpedance amplifier, or a shunt regulator. This is the strength of this book: I strived to exhaustively cover all the possible configurations, with and without optocoupler, regardless of the active compensation element. You will even find TL431 internals secrets that are not often disclosed in data-sheets or application notes. Finally, Chapter 9 closes the subject with measurement methods and design examples.

In the text, I will bring you to the important matter being examined but will suddenly digress to a mathematical tool needed for understanding. This is my writing style. A lot of books simply assume you know the concerned technique and continue the explanation, leaving the reader with a fragmented knowledge. I have tried to avoid this situation and it is the raison d'être of the appendixes at the end of some of the chapters.

In this book, I have derived more than 1,550 equations in three years. Despite all the reviewers' care, it is impossible to have trapped all the typos, missing signs, or wrong numerical results that could have escaped my attention. I sincerely apologize in advance: I know the frustration, as a reader, when you discover errors in a book you want to trust. To help improve the content, I would like you to kindly report the mistake you spot while reading and I will maintain an errata list in my webpage with credit given to the discoverer. It worked very well for the previous book and helped to maintain analytical integrity. Thank you in advance for your kind help. Please send comments to cbasso@wanadoo.fr.

As a conclusion, I have spent three great years writing this book. I learned a lot when tackling some of the subjects. A few of them were tough, and I confess there was highs and lows. But the final content confirms that I was on the right path. I hope your comments will acknowledge that point. Above all, I hope you often come back to this companion book while fulfilling your engineering tasks. Happy reading to you all!

# Acknowledgments

There is no way I could have written a book like this without the help and involvement of many people. My warmest thanks and love go to my dear family: Anne, my wife, who let me write this book despite the numerous long nights during which I struggled with equations or lack of inspiration! My two children, Lucile and Paul, whom I scared by the amount of equations I derived: no, don't worry, kids—this does not happen in the real life!

I was also fortunate to exchange and debate ideas at work with my ON Semiconductor colleagues and friends, Thierry Sutto, Stéphanie Cannenterre, Yann Vaquette, and Dr. José Capilla. Special thanks go to my friend Joël Turchi, who spent long hours reviewing my work throughout the years, kindly pointing out mistakes or flaws in some of my unorthodox approaches! A lot of email exchanges as well as trainings with experts in this field like Dr. Ray Ridley (Ridley Engineering), Larry Meares (Intusoft), Dr. Richard Redl (ELFI), and Dr. Vatché Vorpérian (JPL) helped me to understand the necessity of describing circuits through analytical analysis. This is the best approach to unmask hidden parasitic elements and find the neutralizing parry.

I had the privilege to form a reviewer team made of worldwide experts who spent time reading—they corrected the work but also polished my English. I warmly thank them for the time they kindly allocated for this book review: Joël Turchi (ON Semi), Dr. José Capilla (ON Semi), Jeff Hall (ON Semi), Yann Vaquette (ON Semi), Nicolas Cyr (ON Semi), Jim Young (ON Semi), Patrick Wang (ON Semi), Dr. Vatché Vorpérian (JPL), Dr. Richard Redl (ELFI), Dhaval Dalal (Innovatech), Analogspiceman, Roland Saint-Pierre (Power Integrations), Steve Sandler (Picotest), Georges Gautier (ESRF), Christopher Merren (International-Rectifier), Arnaud Obin (E-Swin), Dr. Chung-Chuieh Fang (Advanced Analog Technology), Dennis Feucht (Innovatia), Dr. Mike Schutten (General Electric), Dr. Germain Garcia (LAAS-CNRS), and Dr. Didier Balocco (AEGPS).

Last, but not least, I would like to thank Deirdre Byrne and the all the team at Artech House for giving me the opportunity to publish my work with them.

CHAPTER 1

# Basics of Loop Control

Often without knowing it, our everyday life utilizes loop control techniques: stretching our muscles to reach a pitcher and pour water into a glass, keeping the bicycle speed constant despite a sudden uphill climb, or maintaining the right pressure on the gas pedal to stay slightly below the maximum speed limit on a long straight road. In all these cases, we have implemented a *feedback control system*: the brain defines a *setpoint* and uses muscles or mechanical power to execute the order. The brain receives information on how the order is executed as the nerves (spinal, optical) permanently feed the information back to it. In the described chains, we have associated a power-amplified control (of which the brain and the muscles represent the *direct path*) with a *return path* (the nervous system). As the information leaves the brain and returns to it via the nervous systems to correct or modify the muscular motion, we say the system operates under *closed-loop* conditions. On the contrary, when the return path is broken, the information confirming that the order has been well or poorly executed is missing. The system is said to run *open loop*. Yes, if you try to ride your bicycle while wearing a blindfold, the biofeedback loop is lost and your body operates in open-loop conditions with all the associated risks!

## 1.1 Open-Loop Systems

As outlined in the previous section, a system can either run in open-loop or closed-loop conditions. An open-loop system transforms a control signal, the *input*, into an action, the *output*, following a specific relationship that links the output to the input. In this open-loop system, the control input $u$ is independent from the output $y$. Figure 1.1 shows a simple representation in the time domain of a system where the output of the system relates to the input by a gain factor $k$. This system is often referred to as *the plant* in the literature and is noted $H$.

In this diagram, the rectangle represents the transmission chain, whereas the arrows portray the physical input and output variables. Please note the usage of letters $u$ and $y$ to respectively designate the input and the output signals as commonly employed in textbooks. In this drawing, the relationship linking the output $y$ to the input $u$ is simply:

$$y(t) = ku(t) \tag{1.1}$$

As it is assumed that coefficient $k$ does not change with time, the system is said to be *linear time invariant* (LTI).

An application example fitting this model is a person turning a car steering wheel by an angle of $\theta$ degrees (the input) to force the wheels turning by an amount of $k\theta$ degrees (the output) on the road. This is what Figure 1.2 depicts. In the early

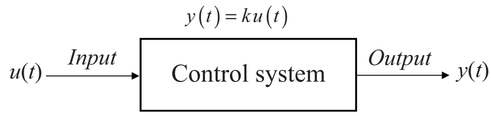



**Figure 1.1** A simple representation of a system where the output depends on the input by a factor *k*.

days of cars, the coupling between the steering wheel and the wheels involved mechanical linkages and hydraulic actuators. Turning the wheels at no or low speed, for instance during parking maneuvers, could quickly turn into a physical exercise for the driver, depending on the size of the car. The power or the force available at the output of the system was almost entirely delivered by the control input—the driver's biceps, in this case!

In modern cars, the control technique has evolved into a power-assisted steering system known as electric power steering (EPS). Sensors detect the motion and the torque applied to the steering column and feed a calculator with their signals. The output of the calculator drives a motor via a power module that provides an assistive torque to the steering gear. The driver can then easily rotate the steering wheel and smoothly induce an angle change on the wheels via the amplification chain. Unlike the previous example, the power or the force available at the output no longer derives from the control input but finds its origins from another source of energy. In a car, it can be an electrical battery, for instance. A simplified schematic of the system appears in Figure 1.3: the steering wheel angle is transformed into a voltage (volts, V) by a potentiometer (or a digital encoder). This signal enters the amplifier that delivers the power (watts, W) to efficiently drive the motor. The motor is coupled to the steering gear and delivers the torque (newton-meters, N.m) to change the wheels position. The succession of the various blocks from the input $u$ to the output $y$ is called the *direct path*.

In Figure 1.2, where no amplification chain exists, it is extremely difficult to convey the force over a long distance from the control source to the output without losses or distortion. The exercise becomes even more difficult when geometric shapes are needed to accommodate a complex environment: curvatures, noncollinear shafts, and so on. Fortunately, this is no longer an obstacle when a power-

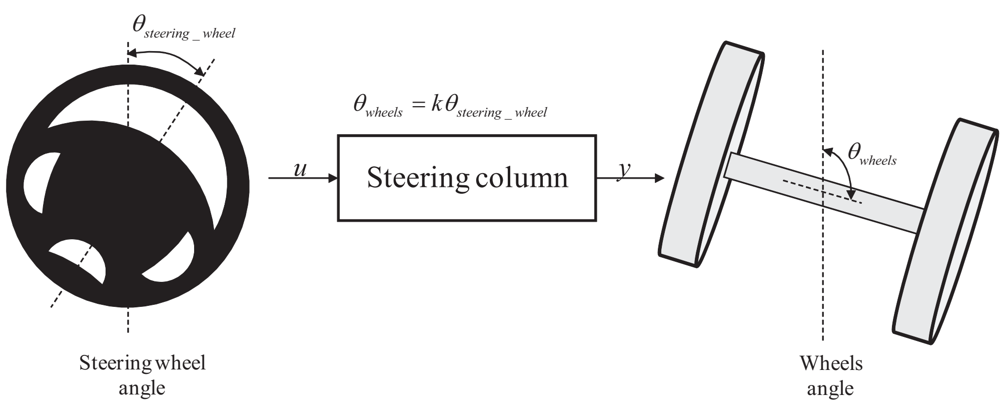



**Figure 1.2** The described system directly converts the angle given by a steering wheel into an angle on the car direction.

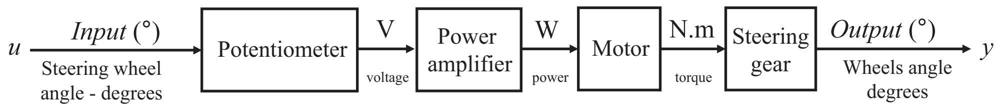



**Figure 1.3** A control system featuring an amplification chain. The input variable does not directly control the output but is conveyed through a series of power systems before it is transformed into the final variable.

amplified chain is implemented, as in Figure 1.3. Despite a long distance between the control input and the delivered output, it is easy to transport an electrical signal through a pair of wires and reach a motor or an actuator placed in a remote location. Named *x-by-wire*, this technique nowadays replaces the mechanical and hydraulic links by an electronic control system using electromechanical actuators located close to the point of action: we have seen the steering control for a car (steer-by-wire), but it can be the controls in an aircraft that actuate the flaps or adjust the engine rotation speed (fly-by-wire).

### 1.1.1 Perturbations

In the described example, we have only considered a single input and a single output. In textbooks, these systems are referred as single-input-single-output (SISO) systems. In reality, any system is affected by several input variables. Therefore, if several input variables are considered, we can envisage as many output variables. In the literature, such a configuration is referred as a multi-input-multi-output (MIMO) system.

Regarding the outputs, we usually select the main output the one that is the most interesting to the designer. In our steer-by-wire system of Figure 1.3, the main output is the angle of the wheels on the road. However, we know that the built-in differential system strives to ensure an evenly distributed torque to each wheel, while allowing them to rotate at different speed. Output variables such as the individual speed of the wheels could then be monitored to deliver information when the car drives along a curve. Another output to consider is the car trajectory, as this is the ultimate goal: forcing the vehicle to follow a curve by acting on the steering wheel without losing car control. Considering all these output variables, the illustration can be updated as Figure 1.4 portrays.

The main input to a control system is usually the one that the output must follow. In our steer-by-wire system, the steering wheel signal obviously represents the main input. However, there could be other inputs that affect the transmission chain. As these inputs are usually undesirable, they are considered as *perturbations*. In our example, assume you try to negotiate a curve in presence of a strong wind. Despite the angle imposed to the car by the steering wheel, you will experience a trajectory deviation. The wind is a perturbation that affects an output, the trajectory. As the

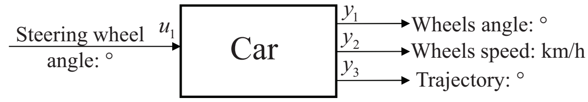



**Figure 1.4** A system can have several outputs depending on the designer interest and the goal to fulfill.

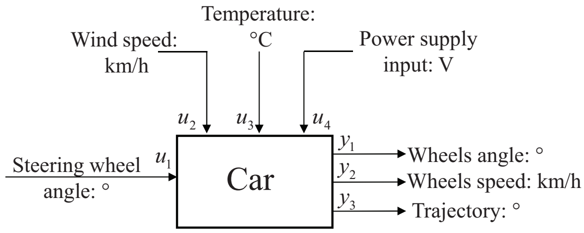



**Figure 1.5** Perturbations affect the transmission chain and can be considered as inputs.

assistive torque is derived from an electric motor, the power source supplying the amplifier is getting weak: the high internal temperature lowers the storage inductor value, and the lack of power affects the actuator responsiveness: the imposed angle is not properly replicated on the wheels. The power supply, when its voltage fluctuates, represents a perturbation affecting the transmission chain. All these perturbations can be treated as multiple inputs that have to be accounted for during the design phase. The ultimate goal is to build a rugged device, insensitive to these perturbations. Figure 1.5 represents the system where these individual contributions appear.

## 1.2 The Necessity of Control—Closed-Loop Systems

In the previous systems, a relationship exists between the input and the output of the system. This is our coefficient $k$, linking the steering wheel angle and the actual angle applied to the direction system. In reality, because of the numerous perturbations or deficiencies in the transmission chain itself, the final output value may never reach the setpoint imposed by the input. For instance, the coefficient $k$ in Figure 1.1 may have great variability itself, linked to the ambient temperature or process variations. How can we counteract the deficiencies in the chain? In our car example, we could imagine a sensor actually measuring the real angle applied to the wheels based on the steering wheel position. This sensor could be a simple potentiometer or a digital rotary encoder. A calculator could then evaluate the difference, or the *error*, between the imposed setpoint and the obtained angle read by the sensor. In the literature, this error is noted $\varepsilon$ (epsilon) and represents the difference between the input setpoint and the final output:

$$\varepsilon = u - y \tag{1.2}$$

From the obtained error, a signal could be generated and injected into the transmission chain to apply a corrective action: if a few degrees are missing or are in excess, the setpoint must be proportionally increased or decreased to compensate the difference. A closed-loop system would therefore be a system where a signal representative of the output is fed back to the control input and acts upon it to ensure the error is kept to a minimum. In such a model, the control signal is no longer the setpoint alone, but the difference signal expressed by (1.2). To upgrade our model representation, we will add a difference box, a circle, illustrating the error signal generation. Figure 1.6 shows the updated sketch. As we deal with electric variables,

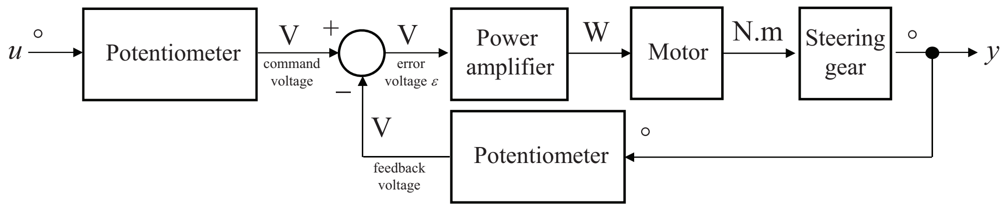



**Figure 1.6** The original setpoint is now affected by a return signal to create an error variable. This error variable controls the whole chain.

a second sensor is added in the *return path* (also called the *return chain*) to convert the actual feedback output angle into a voltage and allow the subtraction with the commanded voltage-image of the input variable. The difference with these provides the error signal $\varepsilon$.

If the output $y$ is too large compared to the target imposed by $u$, the error signal $\varepsilon$ will decrease, commanding the system to reduce the output. On the opposite, if the output is too small, $\varepsilon$ will grow, commanding an increase of the output. As a first approximation, we can say that the system operates adequately and reaches equilibrium as long as the error signal *opposes* the output variation. If for any reason this relationship is lost (i.e., the negative sign in (1.2) becomes positive), the system can run away and will quickly hit its upper or lower stops. We will come back on this important point in a while.

In the Figure 1.6 example, the control input can be variable or fixed. Imagine the driver is on a road and maintaining the wheels straight for a long period of time. In that case, the system simply maintains the output constant and keeps the wheels in the defined axis, fighting against perturbations (i.e., wind) to maintain the trajectory. The system is told to be a *regulator* or a *regulating system*. A regulator is a control system operating with a constant input or setpoint that maintains a fixed relationship between the setpoint and the output, regardless of what the perturbations are. A voltage regulator is another example: it maintains a constant output voltage despite its operating environment, such as input voltage (ac or dc) and output current changes. In this book, we will mostly deal with power converters delivering a fixed output voltage or current, naturally falling into the regulator area.

If the input permanently changes with time, the control system must ensure that the output precisely tracks the input. The French language uses the term "asservissement" (enslavement) to designate such a system. It literally means that the output must be slave to the input. The goal of such a system is indeed to maintain a relationship between the output and the input regardless of the speed and the amplitude at which the control input changes. Such a system is called a *feedback control system*. Audio amplifiers, the autopilot in aircrafts, or marine navigation systems are good examples of feedback control systems, also denominated *servomechanisms* for the latter, as the controlled variable is a mechanical position. They all use complex feedback architectures to ensure the output perfectly follows the changing setpoint regardless of what the perturbations are (wind speed, stream strength, and so on).

In all the cited examples, an extreme precision is needed in the delivered variable: you cannot afford to have a mismatch of several degrees in the control of an aircraft flap for instance. Such a system must be extremely sensitive to the smallest deviation detected between the setpoint and what the sensor returns. To be that sensitive, it is necessary to *amplify* the error signal $\varepsilon$. When amplified, a small output deviation becomes a variation of higher amplitude that the system can properly treat. The *gain* of a system therefore directly relates to the amount of feedback and hence to its precision: a highly precise system exhibits a high *static* (also called dc) gain in the control path. No gain, no feedback—that is to say you must have gain in the loop to realize a control system.

## 1.3 Notions of Time Constants

A control system, regardless of its implementation, usually reacts with delay when adjusted to a new setpoint or in response to a perturbation. This delay is inherent to the control chain, as the signal is conveyed through mechanical, physical, or electronic paths. For instance, in our electric power steering system, when you turn the steering wheel, it takes a certain amount of time for the order to propagate as an angle change on the wheels. Another classical example is the heating system in your house: you program a particular temperature setpoint in the presence of other room parameters (e.g., volume, air flow), but you will need to wait tens of minutes if not an hour before the system, via the sensor, considers the read temperature to be adequate. In these examples, the time needed by the system to change from one state to another is called the *time constant*, noted $\tau$, (Greek letter "tau") and relates the system response to a time (s). The time constants can vary from a few milliseconds in the first case to seconds or hours in the given example.

There are many possible ways to represent the time constant of a first-order LTI system; however, they all obey a common differential equation. For electrical engineers, a simple representation is an $RC$ filter as shown in Figure 1.7.

The output voltage $y(t)$ of such a network can be obtained after a few lines of algebra:

$$y(t) = u(t) - Ri(t) \tag{1.3}$$

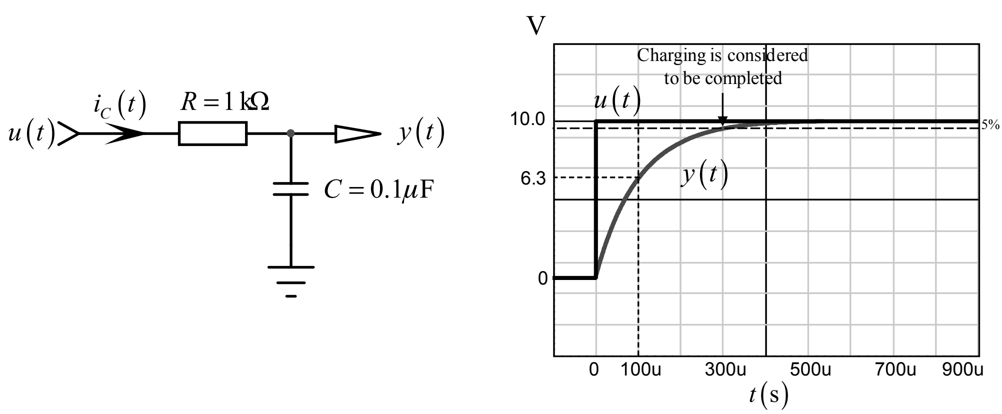



**Figure 1.7** An RC low-pass filter response depends on the time constant in response to an input step.

The current in the capacitor depends on the voltage variation across its terminals:

$$i_c(t) = C\frac{dy(t)}{dt} \tag{1.4}$$

Substituting (1.4) into (1.3) and rearranging, we have an equation describing the behavior of a linear time invariant first-order system:

$$y(t) = u(t) - \tau\frac{dy(t)}{dt} \tag{1.5}$$

With $\tau = RC$, the time constant of the system. If $R$ = 1 kΩ and $C$ = 0.1 $\mu$F, we have a 100-$\mu$s time constant.

The solution to such a differential equation can be found by different means such as the inverse Laplace transform, as we will later see. It can be shown that the solution follows this form:

$$y(t) = A + Be^{-\frac{t}{\tau}} \tag{1.6}$$

where A and B are constant numbers found by solving a 2-unknown/2-equation system for $t = \infty \left(e^{-\frac{\infty}{\tau}} = 0\right)$ and the *initial condition* at $t = 0 \left(e^{-\frac{0}{\tau}} = 1\right)$. An initial condition represents the value of the *state* variable at the beginning of the observation. It can be the current in an inductor or the voltage across a capacitor (for instance, when $t$ = 0). In Figure 1.7, if we consider the input signal $u(t)$ to be a voltage step of amplitude $V_{cc}$, then the voltage across the capacitor $v_C(t)$ is simply

$$v_C(t) = V_{cc}\left(1 - e^{-\frac{t}{\tau}}\right) \tag{1.7}$$

This is the familiar exponential curve shown on the right side of Figure 1.7. The time constant can be determined for $t$ equals $\tau$, reached when $e^{-\frac{t}{\tau}} = e^{-1} = \frac{1}{e}$ in (1.7). Solving for the corresponding value of $v_C(t)$, we have:

$$v_C(t) = V_{cc}\left(1 - \frac{1}{e}\right) \approx 63\%\ V_{cc} \tag{1.8}$$

We can apply this definition to our example. As $V_{cc}$ equals 10 V, the capacitor will reach 6.3 V after $\tau$ seconds. If we read the x-axis for $v_C(t)$ = 6.3 V, we can determine the system time constant. In this example, it corresponds to 100 $\mu$s, the value of the $RC$ product used in the circuit. In a first-order system, the output is within 5 percent of the final target after $3\tau$ have elapsed. In our example, if the input voltage corresponds to a sudden setpoint change and our control system exhibits a 100-$\mu$s time constant, the output of the system will be considered within limits after 300 $\mu$s.

### 1.3.1 Working with Time Constants

The time constant represents one of the most troublesome natural parameters in a control system. Why? Because when the loop observes the output while the control input changes, it sees a variable out of range for a certain time: you change the temperature setpoint in the house and no significant temperature variation happens on the sensor before tens of minutes. As a result, the error signal between the setpoint and the returned value is maximum, pushing the system in its upper or lower power limit with all the associated problems (over power, runaway risks, and so on). Is there any wise control strategy, leading to a more reasonable reaction? One widely adopted solution consists of combining the following actions:

- The principle is to link the control voltage amplitude to the difference sensed by the system between the setpoint and the controlled output. Why overamplify a moderate change if a small increase of the control voltage is the right answer? On the other hand, if the change amplitude is large, then it should be compensated with a stronger control voltage. In other terms, it is desirable to have a control voltage amplitude *proportional* to the detected change in the error signal.
- If a drift on the controlled variable or a slow setpoint transition is sensed, why immediately push the control signal to its upper or lower stop? If the perturbation or the change in the operating point is slow, then let's slow down the reaction. On the other hand, if the control chain senses a fast-moving perturbation, let's make it react quickly. The control voltage must thus be sensitive to the slope of the error signal. How do we compute the slope of an error signal? By taking its time *derivative*.
- Finally, we want to obtain a very precise controlled output, exactly matching the setpoint. One way to fulfill this goal is to increase or decrease the control voltage until the detected error signal is null, meaning the output is right on target. In a control system featuring a permanent error between the setpoint and the output, the error voltage is flat: there is nothing the system can do to correct the situation. How do we transform this flat voltage into a growing or decreasing control signal to force the error reduction to zero? By *integrating* the error voltage. This way, we will obtain a ramp, up or down, automatically driving the control voltage until the error becomes zero.

These functions are usually implemented in a *compensating* block taking place after the error signal. The output of this block becomes the new control voltage, $v_c$. As designers combine a little bit of these functions via the tweaking of their associated coefficients, we call this block a *proportional-integral-derivative (PID)*

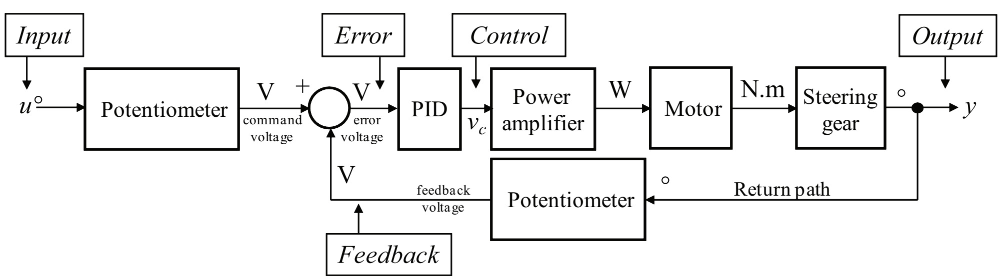



**Figure 1.8** A PID block inserted after the error signal offers a way to shape the behavior of the control system.

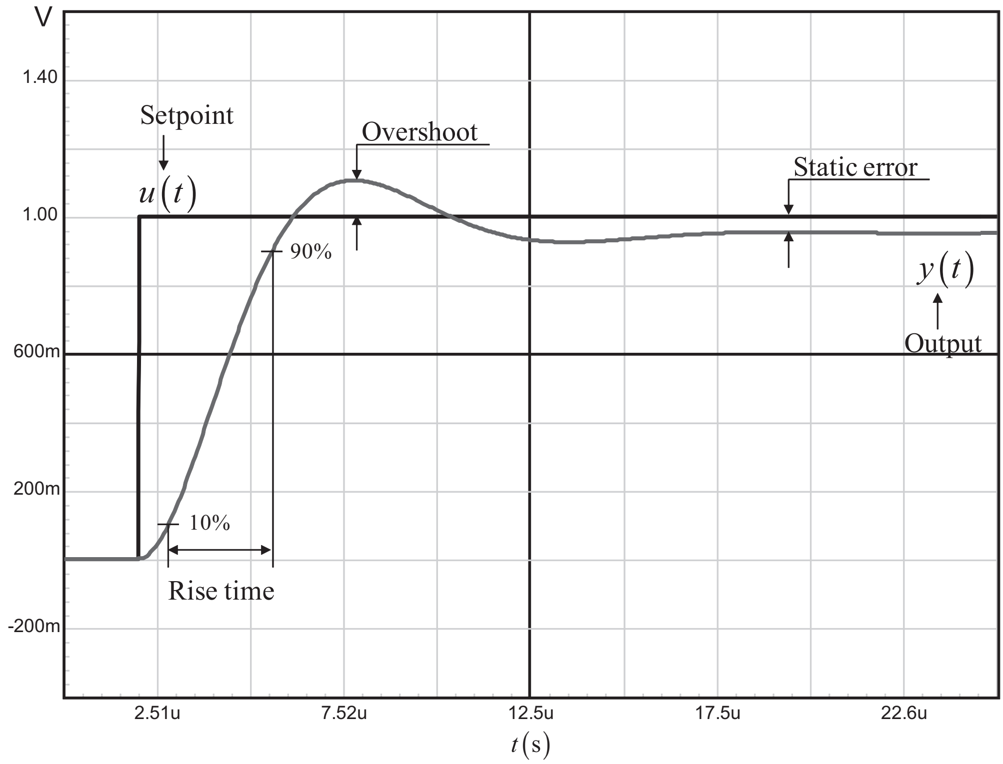



**Figure 1.9** The typical response of a control system illustrating some of the aforementioned parameters.

*compensator*. The term *compensator* means that we want to compensate for system imperfections by purposely tailoring the return chain. Figure 1.8 shows the corresponding update on our control system, including the signal names that we will adopt.

Figure 1.9 shows the typical response of a second-order closed-loop system to a change in the setpoint and illustrates the imperfections we talked about. You can see the output $y(t)$ takes off toward the target and needs a certain time before reaching it. At some point, it even exceeds this target to stabilize to a lower value, giving birth to a permanent mismatch. Implementing PID compensators or correctors will help to minimize these imperfections, leading to control systems exhibiting precision and speed without overshoot. We will study this response in Chapter 2.

Please note that the rise time is measured from 10 to 90 percent of the rising waveform, but the definition can vary. In Chapter 2, it is considered from 0 to 100 percent.

### 1.3.2 The Proportional Term

The idea is simple: if the deviation or the mismatch with the target is big, the control signal $v_c$ is increased. As the opposite, if the distance from the output variable to the expected target is small, a control signal of low amplitude will be generated. With this approach, a proportional relationship exists between the error signal and the deviation amplitude. This link is implemented with a *proportional gain* introduced

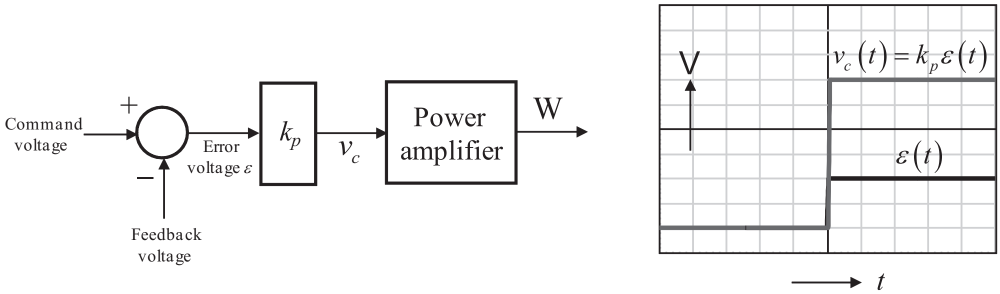



**Figure 1.10** The proportional term amplifies the error signal before reaching the power chain.

in the control chain. It is usually noted $k_p$, as shown in the left side of Figure 1.10. How much gain is needed? In a heating system, if the generated power is high (high $k_p$ value) the output (the heat in our case), will make the temperature rise at a fast pace. On the contrary, if $k_p$ is small, the heat increase will be slower. In case the temperature in the room changes too rapidly, because the heater is pushed to the maximum ($k_p$ is big), there are chances that the target is reached and then exceeded before the loop detects it: an overshoot is created. On the contrary, if you accept a slower but steady reaction ($k_p$ is small), you will limit the overshoot amplitude when reaching the adequate temperature: a proportional gain affects the reaction speed but also the overshoot amplitude.

According to Figure 1.10 drawing, the error signal defined by (1.2) now undergoes a transformation before controlling the amplifier. It becomes

$$v_c(t) = k_p\varepsilon(t) \tag{1.9}$$

If the error voltage $\varepsilon(t)$ is a step, its transformation through (1.9) becomes the scaled-up signal shown on the right side of Figure 1.10.

### 1.3.3 The Derivative Term

If a gain factor $k_p$ is needed for reaction speed, you could also pay attention to the error signal slew-rate. For a slowly moving error signal, why rush the loop and risk the output overshoot? On the contrary, if the error signal is moving quickly, you must make sure the control signal is sufficiently large to impose a fast-paced change on the output. How do we know if the system requires a small driving signal or a larger one? By looking at the error signal slope. This slope can be assessed by the introduction of a derivative coefficient, noted $k_d$ in Figure 1.11.

According to the drawing, the control voltage $v_c$ becomes

$$v_c(t) = k_d \frac{d\varepsilon(t)}{dt} \tag{1.10}$$

In presence of a fast setpoint change, the control voltage will quickly react thanks to the derivative term in (1.10). This is what the right side of Figure 1.11 shows you. For the opposite, when the change is slow, the control voltage will be of lower amplitude. We will later see that the presence of the derivative term contributes to slowing down the system as it also opposes any output variation, including

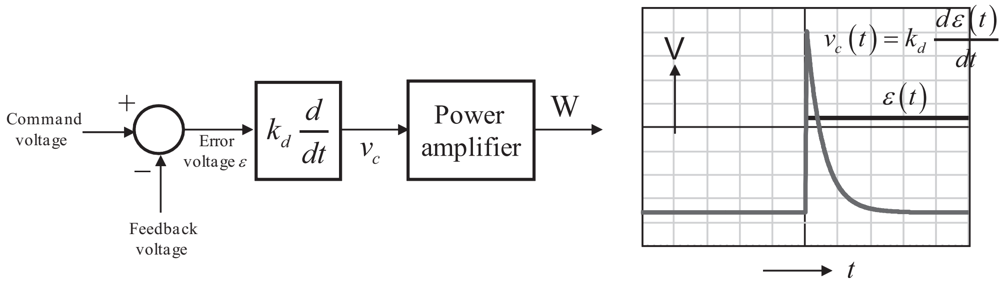



**Figure 1.11** The derivative term produces a control voltage sensitive to the slope of the error voltage.

that driven by the loop in response to a perturbation. As a result, the recovery time is affected by the derivative term. Its presence, however, naturally limits the output overshoot.

### 1.3.4 The Integral Term

What you want from a control system is precision—in other words, the least error between the setpoint and the controlled variable. If the weighted combination of coefficients like $k_p$ and $k_d$ help to reach the target on time while minimizing the overshoot, you need the system to keep its precision over time. In other words, if $k_p$ and $k_d$ clearly affect the transient response, you need another function block that accumulates or *integrates* over time all the long-term errors, eventually canceling the static/dc error. This coefficient is an in *integral* term, noted $k_i$. It appears in the left side of Figure 1.12.

With the presence of the block, the control voltage in the time domain becomes

$$v_c(t) = k_i \int \varepsilon(t)\,dt \tag{1.11}$$

If you integrate a constant signal of $k_i\varepsilon$ amplitude, you obtain a ramp following $v_c = k_i\varepsilon \cdot t$, where $t$ is the elapsed time. As the resulting control signal increases permanently for a constant error, we assume that the target will be exactly matched after a certain amount of time: systems including an integral term are called null-error systems.

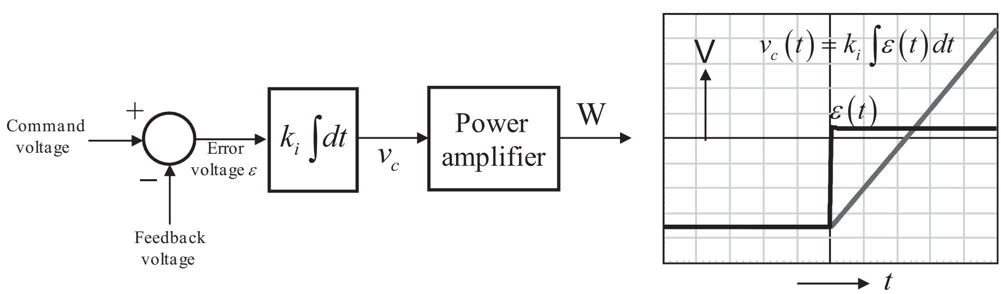



**Figure 1.12** The integral term fights all the long-term errors drifts.

### 1.3.5 Combining the Factors

A well-designed control system reacts quickly to perturbations without excess overshoot or undershoot and exhibits a small static error. However, it is important to understand the necessity of tradeoff between precision (small or zero error) and stability (overshoot amplitude). Addressing this dilemma is partially obtained by combining the above functions through a PID block, where the specific coefficients $k_d$, $k_p$ and $k_i$ are individually tweaked to reach the desired performance.

This combination appears in Figure 1.13. We can clearly see that the control voltage $v_c$ is actually the sum of all blocks outputs:

$$v_c(t) = k_p\varepsilon(t) + k_i\int \varepsilon(t)\,dt + k_d\frac{d\varepsilon(t)}{dt} \tag{1.12}$$

The fine tuning of each individual coefficient is beyond the scope of this book. However, for the vast majority of power converters (linear or switching), this individual tuning is not directly performed by the designer. Rather, the designer will indirectly control the derivative and integral terms by respectively positioning *zeros* and *poles* in the compensator transfer function to match certain design criteria such as *crossover frequency* and *phase margin*.

It is worth noting that a designer can also favor a term in particular, or two (PI or PD). For instance, it is very possible to stabilize a converter with a proportional term or an integral term alone. Power factor correctors are often designed with a simple integrator in the return chain. We will come back in greater details on this PID type of compensator in Chapter 4.

## 1.4 Performance of a Feedback Control System

As control inputs and perturbations can be arbitrary by nature, it is extremely difficult to assess the performance of a control system if you ignore the input signal or the perturbation shapes. Furthermore, a control system can cross various operating modes within which its output must stay within known boundaries. These modes can be transient or steady state (i.e., permanent) and must be separately studied to predict the output variations in a variety of situations. In practice, designers judge

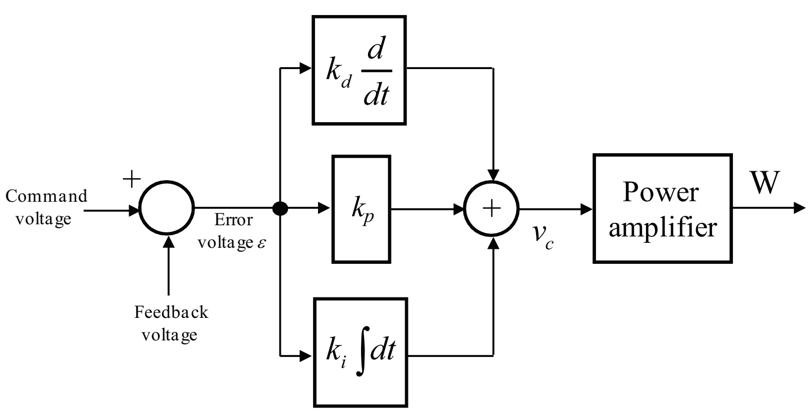



**Figure 1.13** A PID system combines all three coefficients to form the control voltage $v_c$.

the performance of a feedback control system based on its response to a set of test waveforms. There are several of them: the step, the ramp, the Dirac impulse, and the sinusoidal stimulus. As they are among the most commonly used during the analysis of a power converter prototype, we will only look at the step and the sinusoidal input. However, before exploring these signals, it is important to understand the differences between the terms transient and steady state.

### 1.4.1 Transient or Steady State?

In electrical or electronic systems, the term transient designates a period of time during which the system under study is not at its equilibrium state. For instance, it can be the time necessary to let the system start up and reach its nominal output. When you first power up a 5 V converter, all capacitors are discharged, and its output starts from zero and then rises toward the expected nominal value. Further to a certain time designated as a *transient mode*, the output stabilizes to the regulated value: the converter is in *steady-state operation*. Figure 1.14 illustrates this event for a switching converter.

A transient event can also be seen as a sudden change that makes the system deviate from its equilibrium state: (e.g., the load is steady at 100 mA and you suddenly increase it to 1 A). Observing the output of the converter while this sudden change is applied will give you information on its *transient response*.

During a transient period, the system crosses highly nonlinear states and can no longer be approximated to a linear system for its analysis. The transient response of a converter tells you a lot on the way it has been compensated and will be explored

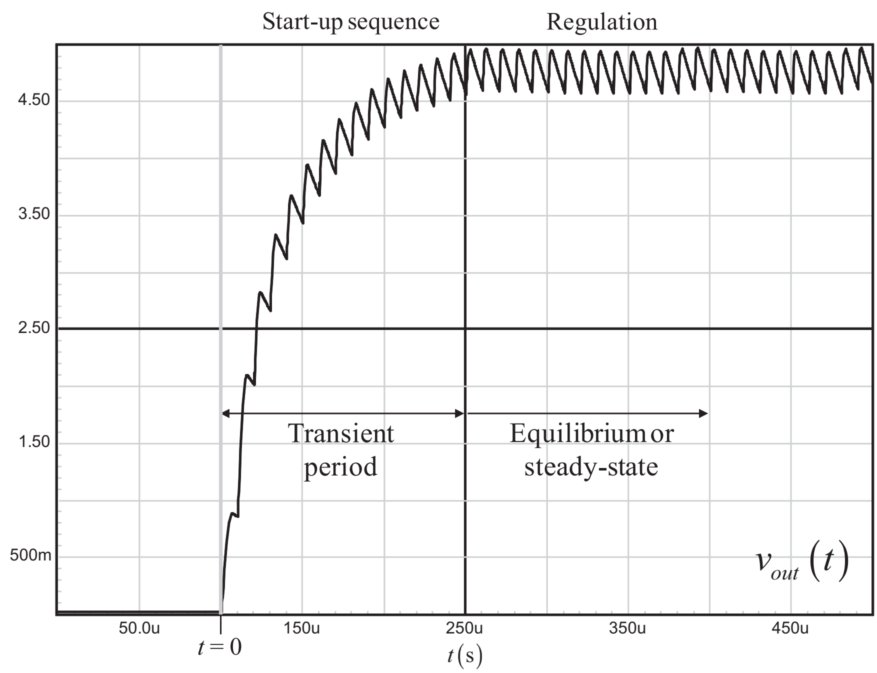



**Figure 1.14** The time needed by the converter to stabilize to 5 V is designated as a transient period.

in Chapter 2. Figure 1.15 depicts the typical response of a converter subjected to a transient load step. As you can observe, the current step has been applied after the converter reached equilibrium. The response of Figure 1.15 is typical of a good design: the output slightly deviates from the target and comes back quickly without overshoot or oscillations.

*Steady state* is the time that the system under study is considered to be at equilibrium: the output has reached a stable operating condition from which it does not deviate. As we will later see, a harmonic analysis can be carried out once the converter reaches steady state. The injected signal perturbs the converter around its *operating point*. Also called *bias point*, it is a working point at which you study the converter. For instance, when you specify a Bode plot or a small-signal analysis, you must indicate the operating conditions at which these data were captured (e.g., $V_{in}$ = 20 V with an output voltage of 5 V delivering 2 A). Further details of the bias point include the duty ratio, the error amplifier output level, and so on. It is important to check these data in a simulation as they indicate a correct dc analysis prior to running the ac sweep. Then, as the deviation brought by the injected harmonic signal around this bias point is small, the converter keeps operating in a linear region, and its response can be analyzed. This is called *small-signal analysis*.

When working with switching converters, observing the current flowing in a capacitor instructs you whether the converter is in steady-state or still in a transient mode. After the converter has stabilized and provided that no external excitations are applied, the time-averaged current in any of the capacitors used in the system must be exactly zero. The same observation can be applied to any inductor where the time-averaged voltage across its terminals must be zero at steady state. Deviations from these values indicate that the converter is not in steady state, suffers instability, or undergoes an ac sweep.

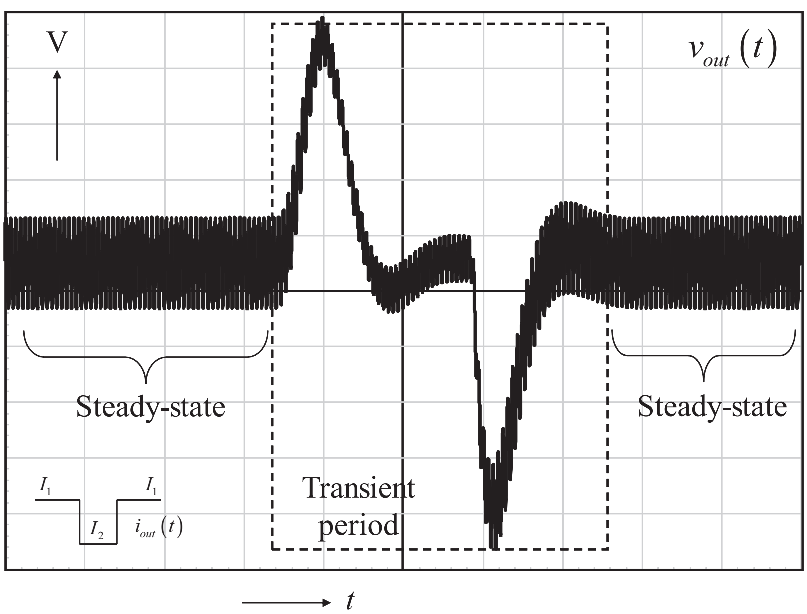



**Figure 1.15** This curve combines a steady-state operation before and after the transient event.

### 1.4.2 The Step

A step function, also named a heaviside function, is a mathematical function whose amplitude is zero for $t < 0$ and equals a constant value for $t \geq 0$. In a closed-loop system, the control input is rarely zero at rest. It can rapidly change from one constant value to another one (for instance, to correct a sudden perturbation). When you study a voltage regulator, the control input is a fixed scaled-up reference voltage $V_{ref}$ you want to replicate on the output (e.g., a 2.5-V reference voltage used to build a 12-V regulator). The perturbations, in this case, are the input variables that can change the operating conditions of the system: the input voltage or the output current. Any of them can thus be stepped to test the system response to a perturbation such as an output current change. Figure 1.16 shows an example where the output voltage of a converter, $v_{out}$, has been subjected to a steep output current increase.

As Figure 1.16 illustrates, stepping the output of two converters A and B via a current source (or using a resistive load with a switch) displays various things on the performance of these converters and also on the internal loop implementation of their respective control section. Both converters deliver an output voltage of $V_0$ when loaded by a current $I_0$. When the current is increased to $I_1$, converter A output severely dips by a voltage $\Delta V_A$. We say the output *undershoots*, meaning that it passes momentarily below the regulated output level. Then, it recovers by going up quickly, exceeding the output—the system now slightly overshoots—before it stabilizes to $V_0$, missed by a very small deviation of $\Delta V_{AA}$ amplitude: this is the

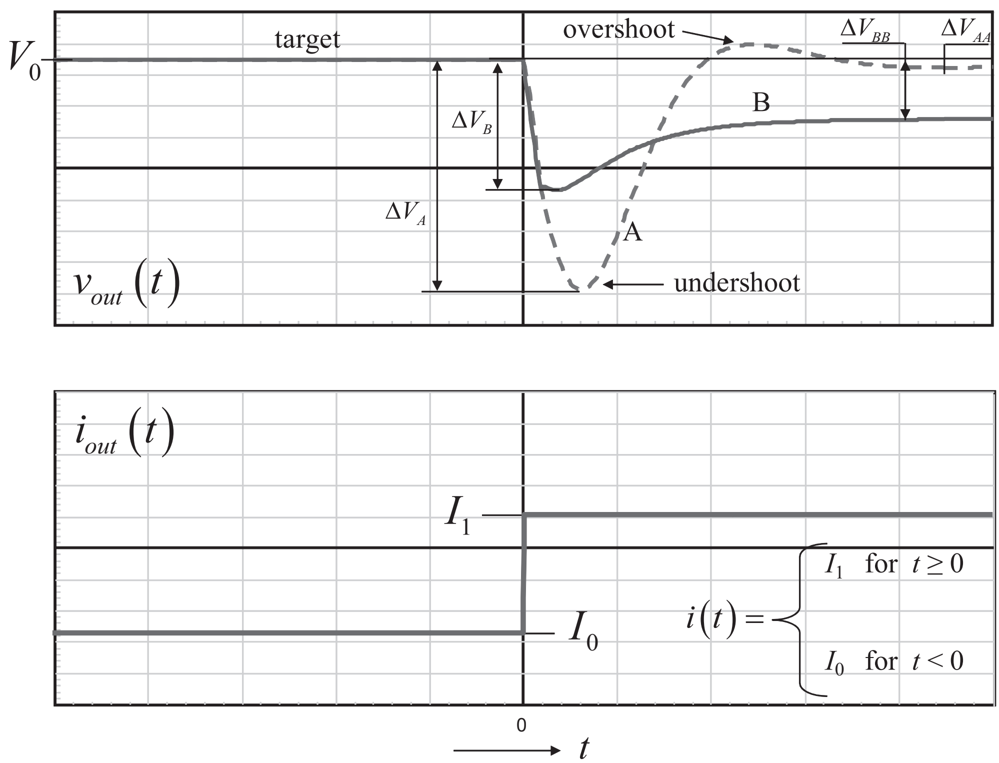



**Figure 1.16** The step function can be seen as a switch that suddenly closes (or opens) to apply a sharp discontinuity to the system under study. Here, the output of a converter is suddenly loaded, and the control system tries to correct the perturbation as much as it can.

*static error* that the system cannot correct. We will later see that a system affected by a large open-loop gain exhibits a very small static error. This theoretical error approaches zero when an integral term (i.e., a pole placed at the origin) is inserted in the control loop. On the second converter, the undershoot $\Delta V_B$ is smaller than that of converter A but the static error $\Delta V_{BB}$ is larger, almost 10 times the previous one. This teaches us that converter B has a lower gain than converter A, hence a larger static error. As observed, the lower gain system does not generate an overshoot.

### 1.4.3 The Sinusoidal Sweep

The sinusoidal stimulus offers an alternative to the input step for studying control systems. A sinusoidal signal is used to reveal the *transfer function* of a given system by ac-sweeping one of its inputs while observing one of its outputs. However, as the transfer function study concerns the output response to one particular perturbed input, the other inputs must be biased at a steady-state level during the sweep. For instance, in a converter, the inputs can be the supply voltage, the output current $i_{out}$, or the control pin $v_c$. If you study $v_{out}$ to $v_c$, then the output current and the input voltage are frozen during the ac sweep. A signal of constant amplitude is injected into the selected input and its frequency varied from a starting value (e.g., 10 Hz) to a stop value (e.g., 100 kHz). At each frequency step, the output amplitude is recorded, as well as its phase difference relative to the input. The amplitude of the injected signal must stay within certain limits, a small-signal analysis, to guarantee the system will not be overdriven and remains in a linear zone throughout the sweep. A possible text fixture appears in Figure 1.17, where an oscilloscope can either be used to capture the points of interest or to control the linearity of the output signal. At the end of the sweep, we end up with a series of data points containing the input amplitude $V_{in}$ (kept constant during the sweep), the output amplitude ($v_{out}$), the phase difference between both signals $\varphi$, and, finally, the frequency $f$ at which these points were stored. This series of data points, amplitude phase couples, are representative of the control system transfer function: when a perturbation or a control setpoint change occurs, how does it propagate in terms of amplitude and phase through the system to finally affect the output? This is the answer the study of the transfer function must give us.

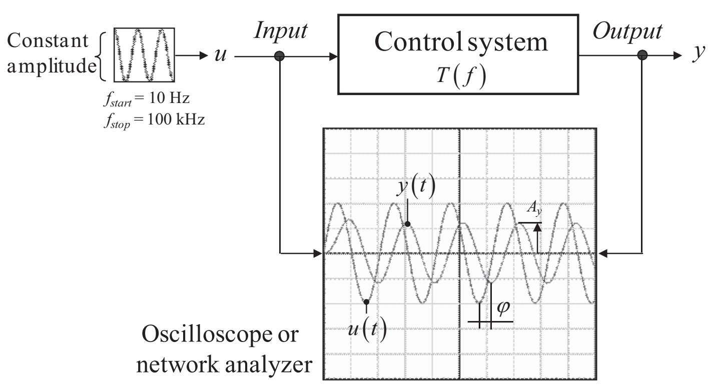



**Figure 1.17** The input is ac-swept while the output signal characteristics (amplitude and phase) is recorded at each frequency step.

### 1.4.4 The Bode Plot

The most common way for plotting the transfer function is to display the magnitude of the ratio $V_{out}/V_{in}$ versus frequency in one graph, while the phase versus frequency appears in a second graph. However, as both the ratio and frequency variations can be quite large, it is usual to logarithmically compress the $x$ and $y$ axis. The final representation becomes a so-called *Bode plot*, after H. Bode, an American engineer working at Bell Labs in the late 1940s. Such a plot is made of two graphs, magnitude and phase, sharing a common horizontal axis graduated in hertz (the log-compressed swept frequency). The upper graph (the magnitude curve) has a vertical scale graduated in decibels (dB), whereas the lower graph simply displays the phase difference in degrees.

The decibel, one tenth of a bell, is a logarithmic unit of measurement commonly used to express the magnitude of a physical quantity (power or current intensity, for instance) relative to a reference level. For instance, when two power levels, $P_1$ and $P_0$, need to be compared, the following formula can be used:

$$G_p = 10\log_{10}\left(\frac{P_1}{P_0}\right) \tag{1.13}$$

If a power source $P_0$ of 10 W is chosen as the reference and you measure a second source $P_1$ of 30 W, you would say that $P_1$ is larger than $P_0$ by 4.8 dB or $P_0$ is smaller than $P_1$ by –4.8 dB.

In our case, as we want to compare input and output voltages, the output of our control system to its input stimulus, the formula needs revision. Going back to (1.13), if we consider that power levels $P_1$ and $P_0$ are obtained by two rms voltages $V_1$ and $V_0$ applied across a common resistor $R$, then the ratio of powers could be reformulated as follows:

$$G_v = 10\log_{10}\left(\frac{V_1^{2}/R}{V_0^{2}/R}\right) = 10\log_{10}\left(\frac{V_1}{V_0}\right)^2 = 20\log_{10}\left(\frac{V_1}{V_0}\right) \tag{1.14}$$

To draw the magnitude curve of our Bode plot, we will simply apply this formula to the collected data points:

$$G_v(f) = 20\log_{10}\left(\frac{V_{out}(f)}{V_{in}(f)}\right) \tag{1.15}$$

For every frequency step, you will thus record and compute the magnitude in decibels plus the phase difference between both input/output signals. Once this information is graphed, you obtain the Bode plot as shown in Figure 1.18 for a first-order system. What kind of frequency step must we select? Usually, to avoid ending up with too many data points, it is recommended to take around 100 points per decade (e.g., 100 points between 10 Hz and 100 Hz and so on). However, if sharp peaks must be observed, there are chances that the 100 points are scattered and the resonance can be masked. In that case, it is recommended to increase the amount of points to, let's say, 1000, to the detriment of the sweep speed of course. By the way, if we take 100 points per decade, what is the resolution of a step then? If we start from $f_{start}$, the next point will be $f_2 = f_{start} \cdot x$ where $x$ is the ratio increase between the second starting point and the starting point. The third point will be at

$f_3 = \left(f_{start} \cdot x\right) x = f_{start} \cdot x^2$. If we select $n$ data points, the equation we need to solve is the following one:

$$f_{start} \cdot x^n = f_{stop} \tag{1.16}$$

Knowing that, over a decade, $f_{start}$ and $f_{stop}$ are linked by a ratio of 10, we have

$$x = 10^{\frac{1}{n}} \tag{1.17}$$

If we start at 10 Hz to end up at 100 Hz, we have $x = 1.02329$ or an increase of 2.33 percent between each point. The second point will therefore be at $f_2 = 10.2329$ Hz, the third at $f_3 = 10.4712$ Hz, and so on.

There are several pieces of information you can extract from the Bode plots represented in Figure 1.18:

- The *cutoff frequency* also called the *corner frequency* is the frequency either below or above which the transfer function magnitude is reduced/increased by 3 dB. In Figure 1.18, we can see that the magnitude is flat in the low frequency area and falls by 3 dB at 1 kHz.
- The cutoff frequency of this first-order system is also the frequency for which the phase lag between the output and the input is 45°. It would be 90° for a second-order system.
- The slope of the magnitude curve is classically given by the vertical displacement divided by the horizontal displacement. In other words,

$$S = \frac{20\log_{10} y_2 - 20\log_{10} y_1}{20\log_{10} x_2 - 20\log_{10} x_1} \tag{1.18}$$

By looking at the graph, for a frequency decade between two points $x_1$ and $x_2$, we read that a 20-dB magnitude difference links $y_2$ and $y_1$. In other words, using (1.15), $y_2 = 0.1y_1$. Given the decade between $x_1$ and $x_2$, we have $x_2 = 10x_1$. Updating (1.18) with these definitions, we have

$$S = \frac{20\log_{10} 0.1y_1 - 20\log_{10} y_1}{20\log_{10} 10x_1 - 20\log_{10} x_1} = \frac{20\log_{10}\left(\dfrac{0.1y_1}{y_1}\right)}{20\log_{10}\left(\dfrac{10x_1}{x_1}\right)} = \frac{-20}{20} = -1 \tag{1.19}$$

When using linear-logarithmic (lin-log) scales for the vertical and horizontal axis, it is possible to draw the ac response through asymptotic curves. These curves for both the magnitude and the phase represent a template made of straight lines. In Figure 1.18, we can see the magnitude and phase response of the following first-order transfer function:

$$\frac{V_{out}(s)}{V_{in}(s)} = \frac{1}{1 + s/\omega_p} \tag{1.20}$$

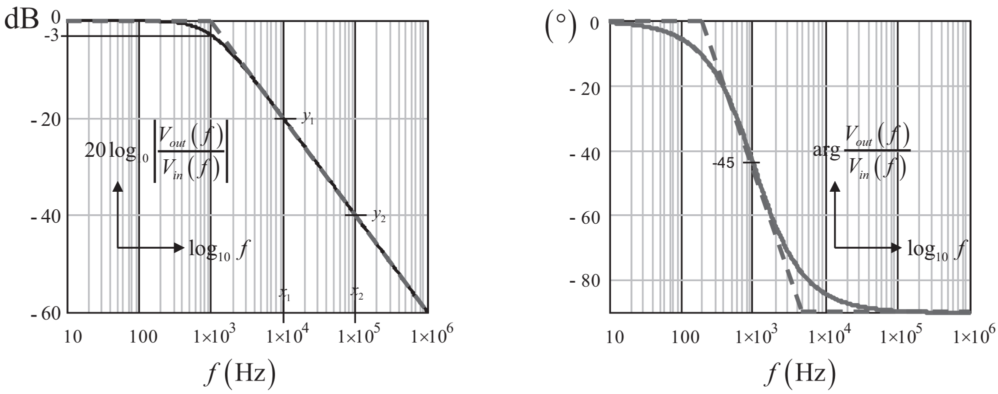



**Figure 1.18** The Bode plot is a first-order system exhibiting a –1 slope.

For frequencies well below the pole position (1 kHz), the magnitude curve is almost flat and can be represented by a straight line up to that point. Beyond, the magnitude decreases following a negative slope of –20 dB per decade, also called a –1-*slope*, as given by (1.19). The phase follows almost the same scenario: 0° well below the cutoff frequency and then lags down to 90° when the frequency is far beyond this value. It is usual to draw a 0° flat line up to one fifth of the cutoff point (200 Hz in our example). A falling line then joins a –90° point placed at five times the cutoff value (5 kHz). They are represented as dashed lines in Figure 1.18. As you can see, the magnitude and phase curves deviate from the asymptotes at the corner points. The deviation values at various observation points can be computed as described in the various references given at the end of this chapter.

A first-order system exhibits a down or up slope of 1 or –1, respectively, implying an increase (single zero) or a decrease (single pole) of 20 dB per decade of frequencies. For a second-order system, this slope becomes 2 or –2 (–40 dB per decade) depending if the magnitude increases (double zero) or decreases (double pole) as the frequency is swept.

## 1.5 Transfer Functions

A control system is characterized by the relationship relating its output, $y$, to its input $u$. As our input signals are arbitrary by nature, the *time-domain* analysis offers a known means to study a control system: if, over time, the input $u(t)$ changes by a certain amount, how does it affect the output $y(t)$? Performing this study requires the usage of *differential equations*. A differential equation uses the notion of derivative, a mathematical tool that measures the rate of change of a function when its input varies by a certain quantity. For instance, in the time-domain equation (1.4), we state that the current inside the capacitor depends on the time-derivative of its terminals voltage (actually the *slope* of the voltage applied on the capacitor) and the capacitor value itself. This equation is then substituted into (1.3) to form the final first-order differential equation of the system under study, (1.5). We say first order because we only differentiate once in relationship to a time interval, *dt*. Should we need to differentiate twice, implying a variable affected by the slope of the slope,

$dt^2$, we would have a second-order system. Generally speaking, the order of the equation depends on the number of distinct energy storage elements present in the circuit. Should you have one inductor and two independent capacitors (not in parallel or series), this is a third-order equation or a third-order system.

### 1.5.1 The Laplace Transform

Looking at the output $y(t)$ delivered by these types of equations requires mathematical skills that power electronics engineers, including myself, are often lacking or have forgotten. Rather than solving differential equations, engineers prefer the *Laplace transform*. For our usage in electronics, we can say that the Laplace transform, noted $\mathcal{L}$, is a mathematical tool that converts complex linear differential equations of any order into a simpler set of algebraic expressions. Once the solution of these algebraic equations is found, the expression can be transformed back to the time domain using the *inverse Laplace transform*, denoted $\mathcal{L}^{-1}$.

When used in electrical circuits, the Laplace transform can also be seen as a tool processing periodic or nonperiodic time-domain functions (e.g., $u(t)$ and $y(t)$) to map them into a two-dimensional plan, the complex frequency domain. In this domain, the new expressions, now noted $U(s)$ and $Y(s)$, are a function of a complex argument $s = \sigma + j\omega$ (also known as $p$ in some countries). The resulting function of $s$ now features an argument (the phase) and a magnitude (the amplitude).

The Fourier transform also maps the time-domain function but into a one-dimensional frequency domain, $\omega$. When a linear system is stable and the initial conditions are zero, the Fourier and Laplace transforms can be used interchangeably; they give identical results including transient response. When a system is unstable (i.e., it features a pole in the right half plane), the Fourier transform of the response simply does not exist because the Fourier integral does not converge. The Laplace integral, with complex frequency $s = \sigma + j\omega$, can be made to converge because of the presence of $\sigma$, which is at the core of making a stability assessment of a system based on a bounded response. Because part of studying systems is stability assessment, and not just determination of a response, the Laplace transform is the appropriate one to use.

The *unilateral* Laplace transform, meaning that positive time only, is considered (the function is said to be *causal* (e.g., zero before $t = 0$)) looks as follows:

$$U(s) = \mathcal{L}\{u(t)\} = \int_0^\infty u(t)e^{-st}\,dt \tag{1.21}$$

This equation can be used to find the response of a linear circuit to a nonsinusoidal excitation such as a ramp, a step, and so on. By involving the inverse Laplace transform on the output equation that is now a function of $s$, we can reconstruct the output signal in the time domain. If we are solely interested in a harmonic analysis in steady state, such as what Figure 1.17 shows, $s$ becomes a pure imaginary number equal to $j\omega$: $\omega$ is the signal angular frequency in radians per seconds, or $2\pi f$, with $f$ being the waveform frequency, in hertz. In this particular case, the mathematical definition of the Laplace transform becomes that of the unilateral Fourier transform:

$$U(j\omega) = \int_0^\infty u(t)e^{-j\omega t}\,dt \tag{1.22}$$

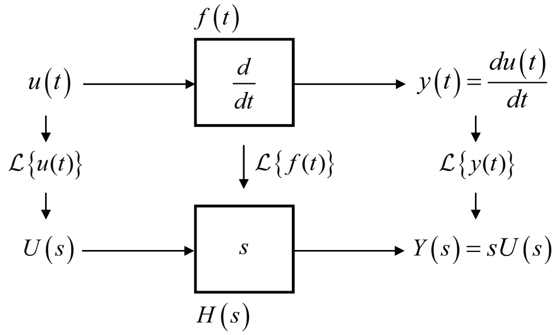



**Figure 1.19** The Laplace transform helps to convert a differential equation into a simple algebraic expression. Here, the initial condition $u_0$ is 0.

In this expression, the term $e^{-j\omega t}$ represents a *phasor*, a reduced mathematical notation including the amplitude and the phase of a sinusoidal signal. The function $U(j\omega)$ will thus carry these characteristics, essential to its description in the complex frequency domain, giving us access to complex parameters such as phase and magnitude. This tool is widely used in the electronic world and in particular in the study of control systems.

There are several interesting properties of the Laplace transform. Among them, the derivative or the integral of a function $u(t)$ are respectively transformed into the Laplace domain by a multiplication or a division by $s$. The Laplace transform of a derivative is

$$\mathcal{L}\left(\frac{du(t)}{dt}\right) = sU(s) - u_0 \tag{1.23}$$

where $u_0$ in the expression is the initial state of $u$ at $t = 0$.

Let's imagine, as in Figure 1.19, a box taking the time derivative of the input signal $u(t)$. If we apply a Laplace transform to this expression, considering null initial conditions for $u$, it becomes a simpler algebraic equation where the multiplication by $s$ now symbolizes the differentiation.

When you write that the voltage across an inductor is defined by

$$V_L(s) = sI_L(s)L \tag{1.24}$$

you simply write that the voltage across the inductor is obtained by differentiating its current $I_L$ further multiplied by the inductor value. In this expression, the term $sL$ is homogenous to the inductor impedance.

The integration follows the same principle and implies a division by $s$:

$$\mathcal{L}\left(\int u(t)\cdot dt\right) = \frac{U(s)}{s} \tag{1.25}$$

Figure 1.20 illustrates the implementation of this principle using our boxes arrangement.

When you write that the voltage across a capacitor is defined by

$$V_C(s) = \frac{I_C(s)}{sC} \tag{1.26}$$

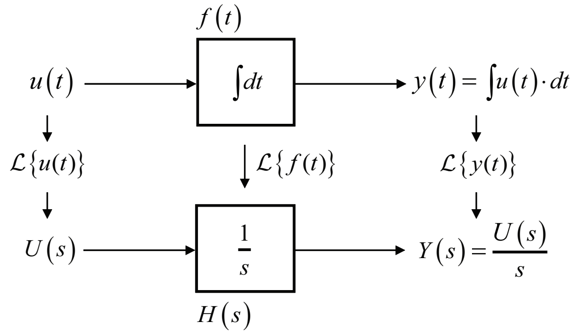



**Figure 1.20** The integration term becomes a simple division by s once going through the Laplace transform.

you simply write that the capacitor voltage is obtained by integrating its current further divided by the capacitor value. In this equation, the term $1/sC$ is homogenous to the capacitor impedance.

### 1.5.2 Excitation and Response Signals

By looking at Figures 1.19 and 1.20, we can see that a relationship now links the output $Y(s)$ to the input $U(s)$ in the frequency domain:

$$Y(s) = H(s)U(s) \tag{1.27}$$

This relationship is called the *transfer function* and is noted $H(s)$ in both figures:

$$\frac{Y(s)}{U(s)} = H(s) \tag{1.28}$$

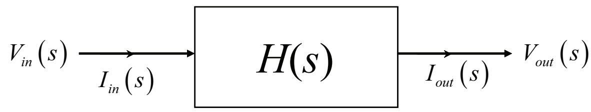
A transfer function is usually expressed by a quotient made of a numerator and a denominator:

$$H(s) = \frac{N(s)}{D(s)} \tag{1.29}$$

For some values of $s$, this transfer function can either be zero, $N(s) = 0$, or can go to infinity for $D(s) = 0$. The *roots* of the numerator $N$ are called the *zeros* of the transfer function. The roots of the denominator $D$ are called the *poles* of the transfer function. We will see later on that these roots can be real, complex, or purely imaginary.

As explained in Dr. Vatché Vorpérian's book (see the recommended books list), a transfer function is characterized by an *excitation* signal and a *response* signal. In our previous expressions, $U(s)$ was the excitation and $Y(s)$ the response. Because excitation and response signals can be a current or a voltage, you can easily combine the variables together, as Figure 1.21 details.

There are six possible types of transfer functions. For the voltage and current gains but also for the transadmittance and the transimpedance definitions, the excitation and response signals are collected at different places in the circuit. How-




$$H(s) = \frac{V_{out}(s)}{V_{in}(s)} \quad \begin{array}{l} \longrightarrow \text{Response} \\ \longrightarrow \text{Excitation} \end{array}$$

$$A_v(s) = \frac{V_{out}(s)}{V_{in}(s)} \text{ voltage gain} \qquad A_i(s) = \frac{I_{out}(s)}{I_{in}(s)} \text{ current gain}$$

$$Y_t(s) = \frac{I_{out}(s)}{V_{in}(s)} \text{ transadmittance} \qquad Z_t(s) = \frac{V_{out}(s)}{I_{in}(s)} \text{ transimpedance}$$

$$\left.\begin{array}{l} Y_{in}(s) = \dfrac{I_{in}(s)}{V_{in}(s)} \\ Y_{out}(s) = \dfrac{I_{out}(s)}{V_{out}(s)} \end{array}\right| \text{ admittance} \qquad \left.\begin{array}{l} Z_{in}(s) = \dfrac{V_{in}(s)}{I_{in}(s)} \\ Z_{out}(s) = \dfrac{V_{out}(s)}{I_{out}(s)} \end{array}\right| \text{ impedance}$$

**Figure 1.21** A transfer function implies an excitation and a response signal.

ever, unlike these four transfer functions, the excitation and response signals for the impedance and admittance are observed at a similar location. If you are measuring an impedance, you usually apply a current source—the *excitation*—at the point where you need the impedance value and read the resulting voltage signal—the *response*—at this very point. For an admittance measurement, you apply a voltage—the excitation—and read the resulting current—the response. When you calculate either the impedance or the admittance, you actually compute a transfer function.

These equations will be affected by the frequency of the excitation signal. A Bode plot tells us how the transfer function evolves in gain and phase when studied in the frequency domain, giving the poles and zeros of the transfer function. The construction of the plot can be undertaken by calculating the magnitude of $H(s)$:

$$|H(s)| = \frac{|N(s)|}{|D(s)|} \tag{1.30}$$

but also by evaluating the phase shift it brings:

$$\arg H(s) = \arg N(s) - \arg D(s) \tag{1.31}$$

### 1.5.3 A Quick Example

The simple $RC$ circuit of Figure 1.7 lends itself very well to a quick application example of the Laplace transform. The time-domain equation of the circuit is given by the following equation:

$$y(t) = u(t) - RC\frac{dy(t)}{dt} \tag{1.32}$$

Considering the capacitor fully discharged at $t = 0$ ($y_0 = 0$), we can apply the Laplace transform to (1.32):

$$Y(s) = U(s) - sRC \cdot Y(s) \tag{1.33}$$

Factoring $Y(s)$ on the left side, we have:

$$Y(s)[1 + sRC] = U(s) \tag{1.34}$$

Rearranging, we have the transfer function we want:

$$H(s) = \frac{Y(s)}{U(s)} = \frac{1}{1 + sRC} \tag{1.35}$$

This is a typical first-order transfer function. The denominator $D(s)$ equals zero for $s_p = -1/RC$. This is a negative root, indicating that the pole is situated in the left-hand portion of the $s$-plane (see the chapter on poles and zeros). With $s$ equal to $j\omega$, the pole definition is obtained by calculating the magnitude of $s_p$:

$$\omega_p = |s_p| = \frac{1}{RC} \tag{1.36}$$

As $\omega = 2\pi f$, we can easily extract the cutoff frequency equal to:

$$f_p = \frac{1}{2\pi RC} \tag{1.37}$$

Substituting (1.36) into (1.35), we obtain a slightly different form, often used throughout this book and the literature:

$$H(s) = \frac{1}{1 + s/\omega_p} \tag{1.38}$$

If we now replace $s$ by $j\omega$ and solve for the denominator magnitude, we have:

$$|D(s)| = \left|1 + j\frac{\omega}{\omega_p}\right| = \sqrt{1 + \left(\frac{\omega}{\omega_p}\right)^2} \tag{1.39}$$

Therefore, applying (1.30):

$$|H(s)| = \frac{1}{\sqrt{1 + \left(\frac{\omega}{\omega_p}\right)^2}} \tag{1.40}$$

When $\omega$ equals $\omega_p$, the magnitude is $1/\sqrt{2}$. In dB, this number becomes

$$20\log_{10}\left(\frac{1}{\sqrt{2}}\right) \approx -3 \text{ dB} \tag{1.41}$$

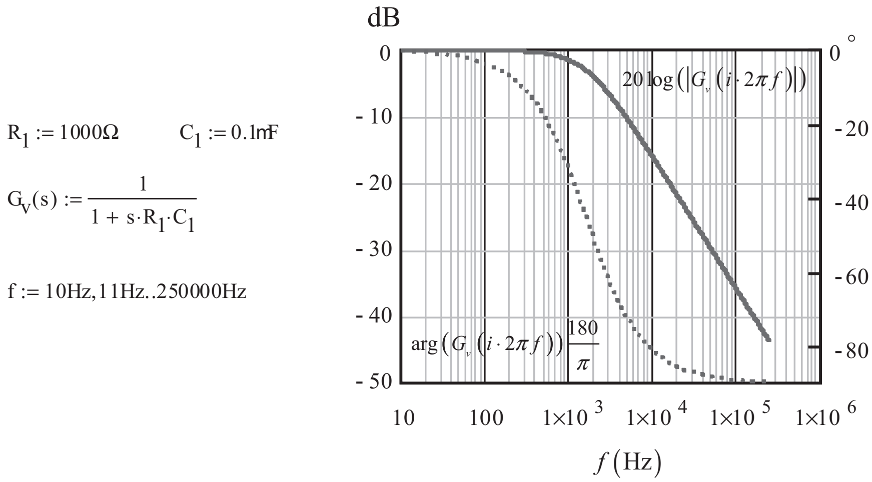



**Figure 1.22** The Bode plot of the first-order network. The component values of Figure 1.7 give a cutoff frequency of 1.6 kHz.

The argument of this transfer function is now calculated using (1.31):

$$\arg H(s) = \arg(1) - \arg\left(1 + j\frac{\omega}{\omega_p}\right) \tag{1.42}$$

As the argument of 1 is zero, the argument of $H(s)$ is simply

$$\arg H(s) = -\tan^{-1}\left(\frac{\omega}{\omega_p}\right) \tag{1.43}$$

When $\omega$ equals $\omega_p$, the argument reaches –45°.

To plot the transfer function over frequency, we can immediately use (1.40) and (1.43) with mathematical software such as Mathcad®. We then obtain the Bode plot presented in Figure 1.22.

### 1.5.4 Combining Transfer Functions with Bode Plots

Control systems are often made of cascaded blocks, each offering a particular frequency response. The transfer function of the whole chain is then simply the product of each individual transfer function. An illustrating example appears in Figure 1.23.

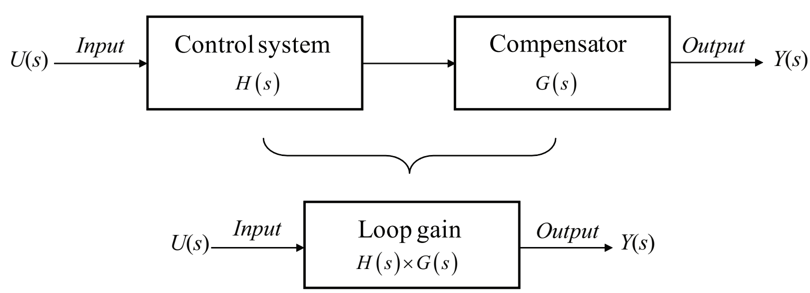



**Figure 1.23** With cascaded blocks, the final transfer function is the multiplication of individual transfer functions.

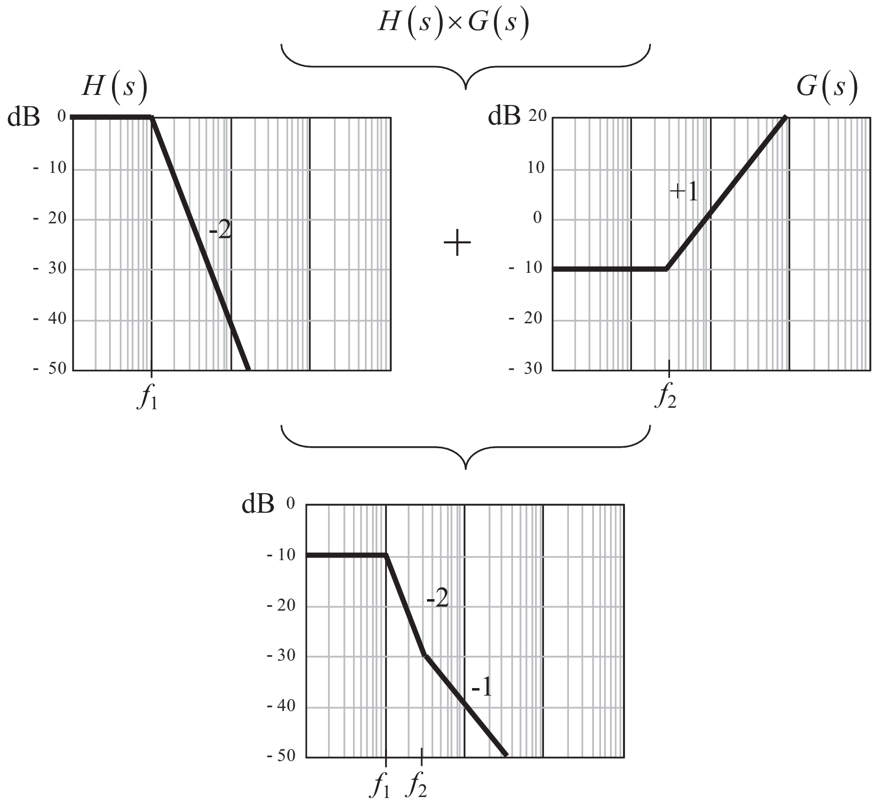



**Figure 1.24** When combining asymptotical responses, the curves and their respective slopes add together to form the final ac answer.

If, rather than assessing each transfer function by its Laplace expression, we have access to individual Bode plots, then we can capitalize on the following property of logarithms, independent from their definition base:

$$\log(A\cdot B) = \log A + \log B \tag{1.44}$$

$$\log\left(\frac{A}{B}\right) = \log A - \log B \tag{1.45}$$

Therefore, if you have individually characterized two cascaded blocks $G(s)$ and $H(s)$ through Bode plots, applying (1.44) means that you just need to sum the graphs points by points to obtain the Bode plot of $T(s) = H(s)G(s)$.

$$20\log_{10}\left[G(s)H(s)\right] = 20\log_{10} G(s) + 20\log_{10} H(s) \tag{1.46}$$

When you consider the slopes, they simply add together as shown in Figure 1.24. For instance, assume you combine a second-order low-pass filter for $H(s)$ with a single-zero response for $G(s)$, the second-order filter offers a flat answer until the cutoff frequency $f_1$ is reached. At this point, its magnitude asymptotically decreases with a –2 slope. The second block magnitude starts with a flat 10-dB attenuation until a frequency $f_2$ is reached. Beyond this breakout point, the magnitude increases with a +1 slope. The combination of both frequency responses will simply be that presented in Figure 1.24 where the +1 slope opposes the –2 slope at the frequency $f_2$ to form a –1 slope. The phase characteristics of each block are also summed together, although not represented here:

$$\arg\left[H(s)G(s)\right] = \arg H(s) + \arg G(s) \tag{1.47}$$

## 1.6 Conclusion

This section ends our quick introduction on the control systems field and its associated terminology. In the coming chapters, we will come back to the topics we've tackled in more detail. Needless to say, the domain is vast and will require effort before mastering it. However, if your interest narrows down to stabilizing simple to moderately complex linear or switching converters, this introduction should get you started.

If you are interested by digging further into the domain of feedback and control systems, the following is a short list of books, articles, and links that will allow you to strengthen your knowledge in that field. Typing in search engines keywords like "modern control theory," "control systems," and so on will lead you to interesting websites and papers.

## Selected Bibliography


[1] Stubberud, A., I. Williams, and J. DiStefano, *Schaum's Outline of Feedback and Control Systems*, New York: McGraw-Hill, 1994.

[2] Vorpérian, V. *Fast Analytical Techniques for Electrical and Electronic Circuits*, Cambridge: Cambridge University Press, 2002.

[3] Saucedo, R. *Introduction to Continuous and Digital Control Systems*, New York: Macmillan, 1968.

[4] Basso, C. *Switchmode Power Supplies: SPICE Simulations and Practical Designs*, New York: McGraw-Hill 2008.

[5] Erickson, B., and D. Maksimovic, *Fundamentals of Power Electronics*, New York: Springer, 2001.

[6] "Course on Modeling and Control of Multidisciplinary Systems," http://virtual.cvut.cz/dynlabcourse, last accessed June 2012.

[7] "Welcome To Exploring Classical Control Systems" http://www.facstaff.bucknell.edu/mastascu/eControlHTML/CourseIndex.html, last accessed June 2012.

[8] Astrom, K., and R. Murray, *Feedback Systems: An Introduction for Scientists and Engineers*, Version 2.10b, February 2009, http://www.cds.caltech.edu/~murray/books/AM08/pdf/am08-complete_22Feb09.pdf.

[9] "Colorado Power Electronics Center Publications," http://ecee.colorado.edu/copec/publications.php.

CHAPTER 2

# Transfer Functions

We learned in Chapter 1 that a transfer function links a response signal to an excitation signal. A transfer function can be written in a lot of different ways whether it has been derived using brute-force algebra or by implementing a smarter approach (e.g., via known tools such as Thevenin/Norton transformations). What actually matters is the insight you can get just by reading the final equation. By insight, we mean your ability to immediately see where poles or zeros are located, if some gain or attenuation exists, just by reading the equation.

## 2.1 Expressing Transfer Functions

Linear network theory teaches us that the transfer function $H$ of a network made of capacitors, resistors, and inductors can be expressed the following way:

$$H(s) = \frac{b_0 + b_1 s + b_2 s^2 + b_3 s^3 + \ldots + b_n s^n}{a_0 + a_1 s + a_2 s^2 + a_3 s^3 + \ldots + a_m s^m} = \frac{N(s)}{D(s)} \tag{2.1}$$

In this equation, it is important that the order of the denominator $m$ is always greater than or equal to that of the numerator $n$. When $m > n$ the magnitude of $H(s)$ goes to zero as $s$ goes to infinity. A transfer function satisfying this property is said to be *strictly proper*. The order of the denominator $D(s)$ reflects the order of the network. The roots that cancel this denominator are called the *poles*. On the other hand, the roots that cancel the numerator $N(s)$ are called the *zeros*. The order of the network can be determined by the number of independent storage elements $C$ and $L$. Should you have two independent capacitors and one inductor, you have three distinct state variables making a third-order circuit: $m$ equals 3 in (2.1) and you have three roots or poles in the denominator.

For instance, if we study a first-order system featuring resistors and a capacitor, its transfer function expression may look like that:

$$H(s) = \frac{b_0 + b_1 s}{a_0 + a_1 s} \tag{2.2}$$

By reading this expression, can you immediately see if the circuit has gain (or attenuation)? Can you identify where the roots of the numerator (the zeros) or the denominator (the poles) are? I cannot. To let you unveil the presence of these elements, you must rearrange that equation into a slightly different format. Factor the terms $a_0$ and $b_0$ to obtain

$$H(s)=\frac{b_0}{a_0}\frac{1+s\dfrac{b_1}{b_0}}{1+s\dfrac{a_1}{a_0}}=G_0\frac{1+s/\omega_{z1}}{1+s/\omega_{p1}} \tag{2.3}$$

Now, from this expression, you can tell that the static gain or the attenuation of the circuit is

$$G_0=\frac{b_0}{a_0} \tag{2.4}$$

and that a pole and a zero cohabit:

$$\omega_{z_1}=\frac{b_0}{b_1} \tag{2.5}$$

$$\omega_{p_1}=\frac{a_0}{a_1} \tag{2.6}$$

If we apply this factorization to (2.1), we obtain

$$H(s)=\frac{b_0}{a_0}\frac{1+\dfrac{b_1}{b_0}s+\dfrac{b_2}{b_0}s^2+\dfrac{b_3}{b_0}s^3+\ldots+\dfrac{b_n}{b_0}s^n}{1+\dfrac{a_1}{a_0}s+\dfrac{a_2}{a_0}s^2+\dfrac{a_3}{a_0}s^3+\ldots+\dfrac{a_m}{a_0}s^m} \tag{2.7}$$

Then, the next step is to factor the polynomials so that the following form appears:

$$H(s)=G_0\frac{(1+s/\omega_{z_1})(1+s/\omega_{z_2})(1+s/\omega_{z_3})\ldots}{(1+s/\omega_{p_1})(1+s/\omega_{p_2})(1+s/\omega_{p_3})\ldots} \tag{2.8}$$

This is the preferred factored pole-zero method as Dr. Middlebrook promoted it in his design-oriented analysis course, [1]. What is important is that a familiar structure appears in the final expression. For instance, it can be difficult to obtain a clean factored form as the previous in a second- or a third-order system. In that case, should you write something like

$$H(s)=\frac{1+s/\omega_{z_1}}{\dfrac{s^2}{\omega_0^2}+\dfrac{s}{\omega_0 Q}+1} \tag{2.9}$$

then the reader will immediately recognize a frequency response combining a zero and the double poles of a second-order system. In that case, we can show that the roots of the denominator are given by

$$s_1,s_2=\frac{\omega_0}{2Q}\left(\pm\sqrt{1-4Q^2}-1\right) \tag{2.10}$$

Depending on the quality factor value $Q$, these roots can be either real or complex conjugates.

### 2.1.1 Writing Transfer Functions the Right Way

As explained in Chapter 1, a transfer function describes the path from an excitation signal, the input, to the delivered response, the output. For instance, let's consider a voltage transfer function $G$ including an origin pole affected by a coefficient $\omega_{po}$, with a zero and a pole. You can write it this way:

$$G(s) = \frac{1 + s/\omega_{z_1}}{\frac{s}{\omega_{po}}\left(1 + s/\omega_{p_1}\right)} \tag{2.11}$$

However, the expression does not give you a lot of insight regarding the overall structure of the transfer function. Ideally, it should fit the format recommended by (2.3) where the first term preceding the $s$-expression is a dc term subscripted with a 0 and sharing a similar dimension as the studied function. Should you write an impedance transfer function, this term should be $R_0$, homogeneous to ohms. For instance, it could look like

$$Z_{in}(s) = \frac{V_{in}(s)}{I_{in}(s)} = R_0 \frac{1}{1 + s/\omega_{p_1}} \tag{2.12}$$

In our case, for a gain, the first term will be called $G_0$ and will be dimensionless. How do we unveil $G_0$ in (2.11)? If we factor $s/\omega_{z_1}$ in the numerator, we obtain

$$G(s) = \frac{\frac{s}{\omega_{z_1}}}{\frac{s}{\omega_{po}}} \frac{\frac{\omega_{z_1}}{s} + 1}{(1 + s/\omega_{p_1})} = \frac{s}{\omega_{z_1}} \frac{\omega_{po}}{s} \frac{1 + \omega_{z_1}/s}{1 + s/\omega_{p_1}} \tag{2.13}$$

If we write

$$G_0 = \frac{\omega_{po}}{\omega_{z_1}} \tag{2.14}$$

Then (2.13) can be rewritten in the following way:

$$G(s) = G_0 \frac{1 + \omega_{z_1}/s}{1 + s/\omega_{p_1}} \tag{2.15}$$

With such a formulation, we can see that when the zero and the pole are fixed, changing the position of $\omega_{po}$ changes the gain of the system. This gain is often called the *mid-band gain*. What is this term, $\omega_{po}$, by the way?

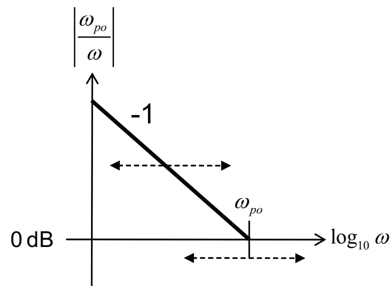



**Figure 2.1** The 0-dB crossover pole is a frequency at which the magnitude of $\omega_{po}/s$ is 1.

### 2.1.2 The 0-db Crossover Pole

I named this factor the *0-dB crossover pole*. The origin pole in an expression such as that described by (2.11) is 0, meaning that when $s = 0$, the quotient goes to infinity. However, when $s$ is combined with a coefficient—for instance when you have $1/sRC(1+\ldots)$—it is advantageous to rewrite it as

$$\frac{1}{sRC(1+\ldots)} = \frac{1}{\dfrac{s}{\omega_{po}}(1+\ldots)} \tag{2.16}$$

in which I call $\omega_{po}$ the 0-dB crossover pole. It equals $1/RC$ in (2.16). It correspond to a cutoff frequency at which the magnitude of $\omega_{po}/s$ simply equals 1 or 0 dB, hence its name. Figure 2.1 graphically represents the magnitude of $\omega_{po}/s$.

Now, when combined with a zero, as in the $G_0$ term from (2.14), it creates a break in the slopes and lets you purposely unveil a gain, changing with the position of $\omega_{po}$. This is what Figure 2.2 shows you for two different values of $\omega_{po}$.

The gain $G_0$ is the mid-band gain defined in (2.14).

## 2.2 Solving for the Roots

The zeros of a transfer function are the frequency points at which the magnitude of the transfer function is zero: the excitation signal, the input, no longer reaches the output. For the opposite, the poles are the frequency points at which the transfer function goes to infinity. If we consider a transfer function as a fraction made of a

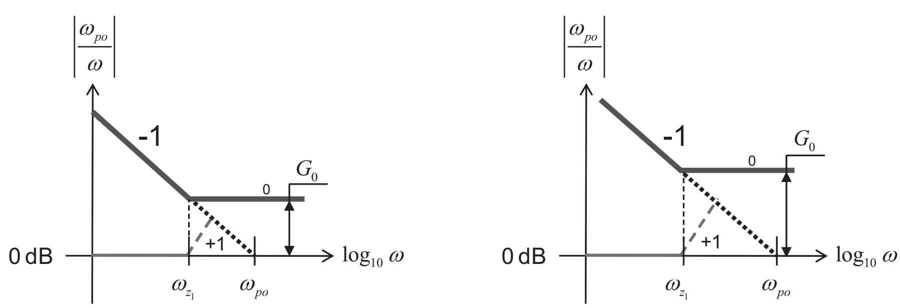



**Figure 2.2** When combined with a zero, the gain $G_0$ changes with the position of $\omega_{po}$.

numerator $N(s)$ over a denominator $D(s)$, then the zeros cancel the numerator while the poles cancel the denominator. In other words, identifying the zeros and the poles of a system is similar to respectively solving for the roots of its numerator and denominator expressions. As all coefficients in (2.7) numerators or denominators are real, the roots can be either purely real or appear in complex conjugate pairs.

Let's go through a few examples where we will use SPICE notations in the upcoming equations: $1k = 1000$, $1Meg = 10^6$, $1m = 0.001$ and $1u = 10^{-6}$. We assume a transfer function $H$ as follows:

$$H(s) = \frac{s+5k}{(s+1k)(s+30k)} \tag{2.17}$$

The zeros of this transfer function are found when $H(s) = 0$ or when $N(s) = 0$. On the contrary, the poles are the roots of the denominator, $D(s)$ and bring $H(s)$ to infinity. Let's rearrange (2.17) to make it look a little friendlier:

$$H(s) = \frac{(s+5k)}{(s+1k)(s+30k)} = \frac{5k(1+s/5k)}{1k(1+s/1k)30k(1+s/30k)} = \frac{1}{6k}\frac{(1+s/5k)}{(1+s/1k)(1+s/30k)} \tag{2.18}$$

$H(s) = 0$ if we have $N(s) = 0$:

$$(1+s/5k) = 0 \tag{2.19}$$

Given the factored form, we can identify the following roots, our real zero:

$$s_{z_1} = -5k \tag{2.20}$$

The poles are obtained when making $H(s) = \infty$ or solving for $D(s) = 0$:

$$(1+s/1k) = 0 \tag{2.21}$$

$$(1+s/30k) = 0 \tag{2.22}$$

Our real poles can be identified at the following position:

$$s_{p_1} = -1k \tag{2.23}$$

$$s_{p_2} = -30k \tag{2.24}$$

When $s$ equals one of these roots, we encounter either a zero or a pole. The poles or zeros frequencies are obtained by computing the roots magnitude:

$$\omega_{z_1} = |s_{z_1}| = 5000 \text{ rad/s or } f_{z1} = \frac{\omega_{z1}}{2\pi} = 796 \text{ Hz} \tag{2.25}$$

$$\omega_{p_1} = |s_{p_1}| = 1000 \text{ rad/s or } f_{p_1} = \frac{\omega_{p_1}}{2\pi} = 159 \text{ Hz} \tag{2.26}$$

$$\omega_{p_2} = |s_{p_2}| = 30000 \text{ rad/s or } f_{p_2} = \frac{\omega_{p_2}}{2\pi} = 4.77 \text{ kHz} \tag{2.27}$$

In these simple examples, the roots are real and you do not need imaginary notation to solve the equations. Let us have a look at a different function:

$$H(s) = \frac{s+4}{(s+0.8)[(s+2.5)^2+4]} \tag{2.28}$$

First, we can massage the expression a little bit to make it easier to work with. Develop the right side of $D(s)$ and factor the terms according to (2.9):

$$H(s) = \frac{4}{0.8\times 10.25}\frac{(1+s/4)}{(1+s/0.8)\left[\dfrac{s^2}{10.25}+\dfrac{s}{\sqrt{10.25}\,\dfrac{\sqrt{10.25}}{5}}+1\right]} = G_0\frac{1+s/\omega_{z_1}}{(1+s/\omega_{p_1})\left(\dfrac{s^2}{\omega_0^2}+\dfrac{s}{\omega_0 Q}+1\right)} \tag{2.29}$$

We can identify a static gain $G_0$:

$$G_0 = \frac{4}{0.8\times 10.25} = 0.488 \tag{2.30}$$

a zero:

$$\omega_{z_1} = 4 \text{ rad/s} \tag{2.31}$$

a pole:

$$\omega_{p_1} = 0.8 \text{rad/s} \tag{2.32}$$

a damped angular frequency $\omega_0$:

$$\omega_0 = \sqrt{10.25} \text{ rad/s} \tag{2.33}$$

and, finally, a quality factor $Q$:

$$Q = \frac{5}{\sqrt{10.25}} \approx 0.64 \tag{2.34}$$

From (2.10), we can see that a $Q$ greater than 0.5 makes the polynomial within the square root a negative number: the roots are complex. Following the definition given in (2.10), we have

$$s_2 = -2.5 + 2j \tag{2.35}$$

$$s_3 = -2.5 - 2j \tag{2.36}$$

These roots are said to be conjugate. The frequency at which these double poles appear is obtained by calculating the magnitude of either $s_2$ or $s_3$:

$$\omega_{p_2} = \omega_{p_3} = |s_{p_2}| = |s_{p_3}| = \sqrt{2.5^2+2^2} = \sqrt{10.25} \text{ rad/s} \tag{2.37}$$

Yes, this is the natural frequency we have evaluated in (2.33). Should we plot (2.29), we would observe the combined action of a pole/zero pair plus a double pole peaking at $\sqrt{10.25}$ radians per second, or 510 mHz.

### 2.2.1 Poles and Zeros Found by Inspection

In the previous paragraphs, we have learned that zeros and poles could be found by solving for the roots of the considered transfer function. To make our life easier, we have seen that rearranging the equation in a factored form could help us identify the poles/zeros positions in a quicker way. Unfortunately, despite the simplicity brought by rearranging the expression, the starting point still remains the transfer function equation that you must derive from node and mesh analysis of the studied network. Rather than deriving the transfer function through classical analysis techniques, is there a way to actually detect where the poles and zeros are hidden and write the transfer function just by looking at the network arrangement—in other words, by *inspection*? After all, we know that the final result should fit the format given by (2.8). Let us see how we could do that.

If we understand that a transfer function links an output signal (the response) to an input signal (the excitation), then a zero at a certain frequency prevents the excitation from reaching the output. Let us try to apply this theory to the passive filter appearing in Figure 2.3. We can see an ac source delivering a signal through a resistor $R_1$ to a network made of a series-parallel combination of two resistors and a capacitor. There is one distinct storage element, the capacitor $C_1$; this is a first-order network. As such, without knowing whether there are poles or zeros, it must fit the format given by (2.3):

$$H(s) = \frac{V_{out}(s)}{V_{in}(s)} = G_0 \frac{1 + s/\omega_{z_1}}{1 + s/\omega_{p_1}} \tag{2.38}$$

The brute-force algebra would be to calculate the expression of impedance $Z_1$ and apply

$$V_{out}(s) = V_{in}(s)\frac{Z_1}{Z_1 + R_1} \tag{2.39}$$

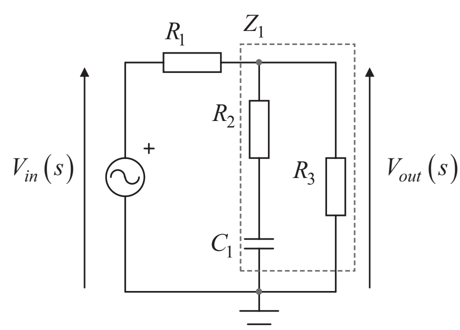



**Figure 2.3** A simple first-order system featuring a zero and a pole.

If you go ahead and develop the equation, you will end up with a moderately complicated expression, but chances to make mistakes during the expansions are real. Furthermore, without additional work on the final result, it is unlikely that poles and zeros pop up at the end.

To start the derivation, let us observe the system at dc, when $s = 0$. This is exactly what SPICE does when it starts the simulation: to calculate the dc operating point of the circuit under study, also called the bias point, SPICE opens all capacitors and shorts all inductors. With the equivalent network, it calculates all dc currents and voltages that the simulator will use for the rest of the simulation. In our network, we can do the same. When $C_1$ is open, we are left with $R_1$ and $R_3$. Therefore, the dc attenuation $G_0$ is simply

$$G_0 = \frac{R_3}{R_1 + R_3} \tag{2.40}$$

Now, if you recall the definition of a zero, it is a frequency point at which the excitation no longer reaches the output. In other words, in Figure 2.3, what element in the input signal path can stop its propagation?

Either an element in series with the signal offers an infinite impedance at a certain frequency or an element linking the signal path to the ground becomes a short circuit, again at a certain frequency point. In our example, the only element that can stop the signal from reaching the output is the series combination of $R_2$ and $C_1$. When its resulting impedance is null (short circuit), we have a zero in the transfer function:

$$R_2 + \frac{1}{sC_1} = \frac{1 + sR_2C_1}{sC_1} = 0 \tag{2.41}$$

You can immediately see the zero position is the root of the numerator:

$$\omega_{z_1} = \frac{1}{R_2C_1} \tag{2.42}$$

We can now write the partial transfer function of our network by combining (2.40) and (2.42):

$$H(s) = G_0\frac{N(s)}{D(s)} = \frac{R_3}{R_1 + R_3}\frac{1 + sR_2C_1}{D(s)} \tag{2.43}$$

We are almost there but we lack the denominator expression $D(s)$. This is where the poles hide.

### 2.2.2 Poles, Zeros, and Time Constants

By definition, a gain is a dimensionless expression. When you say the voltage gain of a system is 20 dB, it is another means to say that the gain is 10 V/V or 10. In other words, if we go back to (2.7), and only consider second-order terms (for the sake of simplicity), we can write the transfer equation of second-order network as follows:

$$H(s) = G_0 \frac{1 + \dfrac{b_1}{b_0} s + \dfrac{b_2}{b_0} s^2}{1 + \dfrac{a_1}{a_0} s + \dfrac{a_2}{a_0} s^2} \tag{2.44}$$

A term multiplied by $s$ has the dimension of a frequency, Hz. A term multiplied by $s^2$ has a dimension of a squared frequency, $\mathrm{Hz}^2$. To make sure that all $s$-terms and $s^2$-terms lose their dimension when multiplied by a coefficient, these coefficients must have the inverse dimension. Therefore, the terms $b_1/b_0$ and $a_1/a_0$ must have a dimension of $\mathrm{Hz}^{-1}$, whereas the terms $b_2/b_0$ and $a_2/a_0$ must have a dimension of $\mathrm{Hz}^{-2}$. What offers a dimension of $\mathrm{Hz}^{-1}$ or actually seconds? A time constant does. What has a dimension of $\mathrm{Hz}^{-2}$ or squared seconds? A product of time constants. And this makes sense when you consider the expression of the zero found in (2.42); it has indeed the dimension of a time constant. How to get the poles then? We need to identify the time constants of our system but in a different manner than what we have shown for the zeros. As we stated, the transfer function denominator $D(s)$ of a linear network does not depend on its excitation or response signals. It only depends on the network structure alone. If you look at transfer functions describing a given network, its output impedance, its input admittance, and so on, then you will see that all these equations share a common denominator $D(s)$.

To study the network alone, we are going to bring its excitation signal to zero. How do we do that? Well, if the excitation signal is a voltage source, we set it to zero: replace it by a short circuit. If the excitation signal is a current source, then open circuit it. Let's apply this technique to Figure 2.3 by shorting to ground the left terminal of $R_1$:

The time constant is easily calculated by evaluating the resistance "seen" from the capacitor terminals. The first one is obviously $R_2$, in series with the parallel combination of $R_1$ and $R_3$:

$$R = R_2 + R_1 \,||\, R_3 \tag{2.45}$$

The pole definition, for this simple first-order system, is simply the inverse of the equivalent time constant:

$$\omega_{p_1} = \frac{1}{\tau} = \frac{1}{RC_1} = \frac{1}{(R_2 + R_1 \,||\, R_3)C_1} \tag{2.46}$$

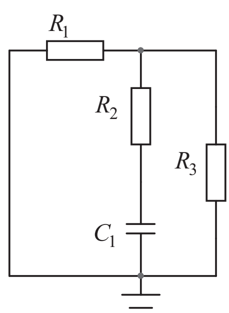



**Figure 2.4** By shorting the input voltage, we can reveal the time constants of the system.

The expression of our denominator $D(s)$ is therefore

$$D(s) = 1 + s/\omega_{p_1} \tag{2.47}$$

The complete transfer function then becomes

$$H(s) = G_0 \frac{N(s)}{D(s)} = \frac{R_3}{R_1 + R_3} \frac{1 + sR_2C_1}{1 + s(R_2 + R_1 \parallel R_3)C_1} = G_0 \frac{1 + s/\omega_{z_1}}{1 + s/\omega_{p_1}} \tag{2.48}$$

This is what is called a *low-entropy* equation by analogy to thermodynamic laws. Simply put, the entropy of a system qualifies its degree of internal disorder: to produce the work the system has been designed for, you need to bring less external energy when its entropy is low (elements are well organized and well ordered) than when it is high (elements are in disorder; this is a chaotic organization). For our equations, a low-entropy expression gives you immediate insight, without further work, on the transfer function it realizes. For the opposite, a *high-entropy* equation does not reveal anything and requires further energy through factorizations or expansions before it tells you where poles and zeros are. The analytical technique we just described lets you write low-entropy equations by *inspection*, just by looking at the network schematic and identifying its time constants. Let us check this again via another simple example that appears in Figure 2.5 with an inductive first-order network. Let us try to derive the transfer function applying what we learned.

First, we start from $s = 0$, the dc transfer function. If capacitors are open-circuited at dc, inductors, for the opposite, are considered as short circuits. When shorting $L_1$, the attenuation $G_0$ is immediately written as

$$G_0 = \frac{R_3 \parallel R_2}{R_3 \parallel R_2 + R_1} \tag{2.49}$$

The zero is found by identifying a network that prevents the excitation from reaching the output of the circuit under study. A series element admittance can become zero at a frequency point or a network connecting the signal path to ground can have an impedance that drops to zero. In our circuit, the series path is $R_1$ and offers a fixed value. On the other hand, a network that can potentially shunt the

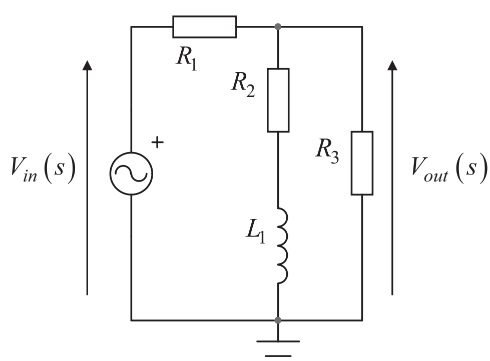



**Figure 2.5** The capacitor has been replaced by an inductor.

excitation signal to the ground is $R_2$ and $L_1$. This is the place where the zero hides. To unveil it, simply solve

$$R_2 + sL_1 = R_2\left(1 + s\frac{L_1}{R_2}\right) = R_2\left(1 + \frac{s}{\omega_{z_1}}\right) = 0 \tag{2.50}$$

We have our zero position:

$$\omega_{z_1} = \frac{R_2}{L_1} \tag{2.51}$$

If the time constant of the resistor $R$ "driving" the capacitor $C$ is $RC$, then the time constant of an inductor $L$ driven by a resistor $R$ is $L/R$. What impedance drives the inductor $L_1$ in Figure 2.5? To discover it, we set the excitation signal to zero. The impedance is the series combination of $R_2$ and $R_1$ in parallel with $R_3$:

$$R = R_2 + R_1 \parallel R_3 \tag{2.52}$$

The pole definition, for this first-order system, again, is the inverse of the equivalent time constant:

$$\omega_{p_1} = \frac{1}{\tau} = \frac{R}{L_1} = \frac{(R_2 + R_1 \parallel R_3)}{L_1} \tag{2.53}$$

The expression of our denominator $D(s)$ is therefore

$$D(s) = 1 + s/\omega_{p_1} \tag{2.54}$$

The complete transfer function then becomes

$$H(s) = G_0\frac{N(s)}{D(s)} = \frac{R_3 \parallel R_2}{R_3 \parallel R_2 + R_1}\frac{1 + s\dfrac{L_1}{R_2}}{1 + s\dfrac{L_1}{(R_2 + R_1 \parallel R_3)}} = G_0\frac{1 + s/\omega_{z_1}}{1 + s/\omega_{p_1}} \tag{2.55}$$

You can see how efficient this fast analytical method can be to express the transfer function of a simple network. The appendix at the end of this chapter shows it at work in a bridge impedance determination. Try to derive it yourself using the classical algebra technique, and you will quickly adopt these fast analytical techniques! Of course, in a small chapter portion, we have just scratched the surface, and as you complicate the network under analysis with more storage elements, you need to apply different techniques such as the extra-element theorem (EET). I encourage you to check [2–6] at the end of this chapter, as they will offer you a means to learn and make this technique efficiently work for you.

## 2.3 Transient Response and Roots

The stability of a closed-loop system can be assessed in different ways. If an ac-sweep teaches us about phase margin at a given crossover frequency, it does not

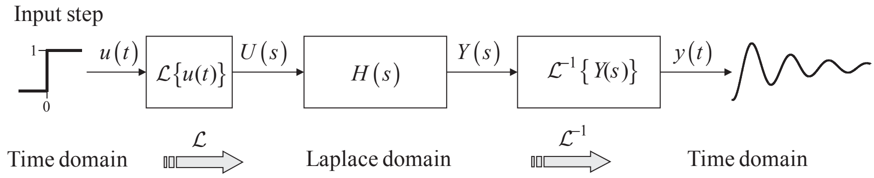



**Figure 2.6** You can assess the time-domain response of a Laplace transfer function by exciting its input with a unity step.

readily tell us how the system will react to an incoming perturbation or a sudden change in the input setpoint. A common test consists of exciting the control system input with a given stimulus. As detailed in Chapter 1, there are numerous types of available stimuli: a step, a Dirac pulse, a linear ramp, and so on. Generally, the response to a step is the most popular choice, in particular for regulators such as linear or switching converters. If stepping an electronic load in the laboratory does not require a particular care, applying the technique to a transfer function implies that some precautions will be observed. First, our transfer function is expressed in the Laplace domain, whereas the step belongs to the time domain. A transformation is needed to transit from one domain to the other. Once the step is converted into the Laplace domain, it becomes the excitation signal $U(s)$ to the transfer function $H(s)$ under study. Then, as its output signal $Y(s)$ is also expressed using Laplace notation, an inverse Laplace-transform is necessary to return to the time-domain and see the resulting waveform. Figure 2.6 shows this process.

What is the Laplace-transform of a unity step? Let's have a look at the waveform that appears in Figure 2.7: it is 0 for all negative time values and equals 1 at $t = 0$ and all values beyond. This is a time-domain waveform and, according to Figure 2.6, we must transpose it into the Laplace domain before driving the transfer function of interest. In Chapter 1, we learned the definition of the Laplace transform:

$$U(s) = \mathcal{L}\{u(t)\} = \int_0^{\infty} u(t)e^{-st}\,dt \tag{2.56}$$

As $u(t)$ equals 1 from 0 to $\infty$, the equation becomes

$$U(s) = \int_0^{\infty} e^{-st}\,dt = \lim_{P\to\infty}\int_0^{P} e^{-st}\,dt = \lim_{P\to\infty}\left[-\frac{e^{-st}}{s}\right]_0^P = \lim_{P\to\infty}\frac{1-e^{-sP}}{s} = \frac{1}{s} \tag{2.57}$$

This is the classical definition of a unity step in the Laplace domain for $s > 0$.

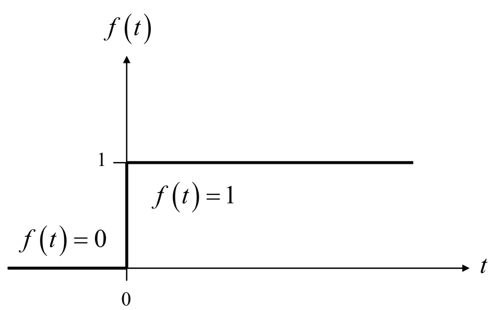



**Figure 2.7** A unity step function is 0 for $t < 0$ and jumps to 1 for $t \geq 0$.

If an input signal 1/*s* enters a transfer function *H*(*s*), the resulting output signal is nothing else than

$$Y(s) = \frac{1}{s}H(s) \tag{2.58}$$

From this Laplace-domain expression, we need to extract its time-domain correspondence through a reverse Laplace-transform:

$$y(t) = \mathcal{L}^{-1}\{Y(s)\} \tag{2.59}$$

Is it as simple as that? Well, it depends if you use a mathematical solver or try to derive the equation yourself. In the second case, you have to realize that the answer is not the product of the inverse Laplace transform of each individual term: you must rewrite the expression as a sum of individual terms, each having its own inverse Laplace-transform equivalent. Since the Laplace-transform is a linear operator, the inverse-Laplace transform of this sum will be the sum of all individual inverse Laplace-transform terms. However, as most of our transfer functions are often in the form of a rational function $N(s)/D(s)$, you need to split them into a sum of ratios of small polynomials. This technique is called *partial fraction expansion*, and articles on the subject can be found on mathematical textbooks or on the Web. The involved algebra looks simple but you need to be careful when expanding complex transfer functions.

Let's try to apply the technique to our first transfer function (2.17) and see its response to a unity-step stimulus:

$$Y(s) = \frac{1}{s}H(s) = \frac{1}{s}\frac{s+5k}{(s+1k)(s+30k)} \tag{2.60}$$

The technique we are going to use is called the Heaviside cover-up method, named after the English electrical engineer Oliver Heaviside. In the case of (2.60), theory tells us that it can be rewritten or expanded into the following terms:

$$Y(s) = \frac{a_1}{s} + \frac{a_2}{s+1k} + \frac{a_3}{s+30k} \tag{2.61}$$

As you can see, each denominator cancels for the roots we already found in (2.23) and (2.24), but it now also includes $s = 0$. To determine the value of coefficient $a_1$, $a_2$, and $a_3$, the idea is to make *s* equal the selected root (0 for $a_1$, $-1k$ for $a_2$, and so on) while multiplying (2.60) by the denominator including that root (by *s* for $a_1$, by $s+1k$ for $a_2$, and so on). Therefore, the concerned denominator on both sides of the fraction naturally disappears, a bit like if you were covering it up with your finger. It then leaves a simple equation in *s*, where *s* takes the value of the selected root. Sounds complicated? Not really as the following details show:

$$a_1 = s\frac{1}{s}\frac{s+5k}{(s+1k)(s+30k)}\bigg|_{s=0} = \frac{0+5k}{(0+1k)(0+30k)} = \frac{1}{6k} = 166.6u \tag{2.62}$$

$$a_2 = (s+1k)\frac{1}{s}\frac{s+5k}{(s+1k)(s+30k)}\bigg|_{s=-1k} = -\frac{1}{1k}\frac{-1k+5k}{-1k+30k} = -\frac{4k}{1k\cdot 29k} = -138u \tag{2.63}$$

$$a_3 = (s+30k)\frac{1}{s}\frac{s+5k}{(s+1k)(s+30k)}\bigg|_{s=-30k} = -\frac{1}{-30k}\frac{-30k+5k}{-30k+1k} = -\frac{25k}{30k \cdot 29k} = -28.7u \tag{2.64}$$

This is it, if we rewrite (2.61) with the right coefficients, we have

$$Y(s) = \frac{166u}{s} - \frac{138u}{s+1k} - \frac{28.7u}{s+30k} \tag{2.65}$$

The inverse Laplace-transform of the previous equation is thus

$$\mathcal{L}^{-1}\{Y(s)\} = \mathcal{L}^{-1}\left\{\frac{166u}{s}\right\} - \mathcal{L}^{-1}\left\{\frac{138u}{s+1k}\right\} - \mathcal{L}^{-1}\left\{\frac{28.7u}{s+30k}\right\} \tag{2.66}$$

We can look at the inverse Laplace-transform tables to get each individual term:

$$\mathcal{L}^{-1}\left\{\frac{166u}{s}\right\} = 166u \cdot \mathcal{L}^{-1}\left\{\frac{1}{s}\right\} = 166u \tag{2.67}$$

$$\mathcal{L}^{-1}\left\{\frac{138u}{s+1k}\right\} = 138u \cdot \mathcal{L}^{-1}\left\{\frac{1}{s+1k}\right\} = 138u \cdot e^{-1k \cdot t} \tag{2.68}$$

$$\mathcal{L}^{-1}\left\{\frac{28.7u}{s+30k}\right\} = 28.7u \cdot \mathcal{L}^{-1}\left\{\frac{1}{s+30k}\right\} = 28.7u \cdot e^{-30k \cdot t} \tag{2.69}$$

Assembling all these terms according to (2.66), we now have our time-domain expression:

$$y(t) = 166u - 138u \cdot e^{-1k \cdot t} - 28.7u \cdot e^{-30k \cdot t} \tag{2.70}$$

The first immediate remark is that the exponents on the exponential terms are the roots of the characteristic equation of $H(s)$. The second remark concerns the zeros. They do influence the time-domain response, but only the poles directly impact the decaying time constants in the response as we just saw. The third remark concerns their signs, negative. It means that when $t$ goes to infinity, since all exponential terms go to zero, the output signal reaches a steady-state value given by the first term: with a 1-V step, the output should reach 166 $\mu$V. By the way, we knew it from the start, after we rearranged (2.17) into (2.18) as the dc gain, $G_0$, is 1/6$k$ or 166u. This is confirmed if we plot (2.70) in Figure 2.8.

From these derivation lines, if we try to obtain the step response from a transfer function that contains positive roots in its characteristic equation, the time-domain response will contain exponential terms featuring positive exponents. As $t$ goes to infinity, these exponential terms do not die out to zero but keep increasing, making the output signal severely diverging: the system response is not bounded; you do not control anything. The question that immediately arises is, if my analysis shows that all roots are negative at the considered operating point, can they move and suddenly become positive?

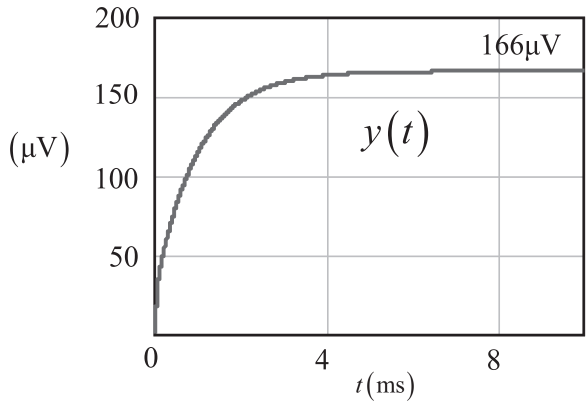



**Figure 2.8** Time-domain response of the transfer function $H(s)$ described by (2.17) when subject to a 1 V input step.

### 2.3.1 When the Roots Are Moving

In the previous examples, we have fixed-values roots (e.g., –1$k$ or –30$k$). They do not depend on other variables, and the transient response to a step won't change as long as the roots remain constant. Now, let us consider a more practical case such as the unity-return regulator appearing in Figure 2.9.

We can see a transfer function $H(s)$ that, for instance, could be the simplified expression of a voltage-mode switching converter operated in the discontinuous conduction mode (DCM), a buck-boost. We can show that the zero $\omega_{z_1}$ is created by the output capacitor $C_{out}$ and its equivalent series resistance (ESR):

$$\omega_{z_1} = \frac{1}{R_{ESR}C_{out}} \tag{2.71}$$

The pole $\omega_{p_1}$ depends on the load $R_{load}$ and $C_{out}$:

$$\omega_{p_1} = \frac{2}{R_{load}C_{out}} \tag{2.72}$$

To stabilize the power supply, we have designed a compensator $G(s)$ that places a pole $\omega_{p_2}$ while offering some gain $G_2$. This is obviously not the best-known compensation, but we keep it simple for the sake of the example. We will learn in Chapter 3 that a unity feedback system closed-loop response $T_{CL}(s)$ obeys the following

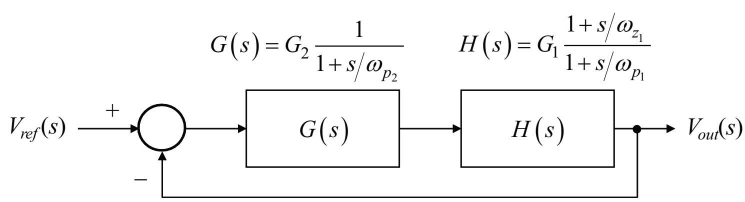



**Figure 2.9** This circuit mimics a first-order converter $H$ stabilized by a 60-dB gain compensator $G$ featuring a single pole response.

expression where the term $1+G(s)H(s)$ represents the characteristic equation of the closed-loop transfer function:

$$T_{CL}(s)=\frac{T_{OL}(s)}{1+T_{OL}(s)}=\frac{G(s)H(s)}{1+G(s)H(s)} \tag{2.73}$$

In this equation, $T_{OL}(s)$ represents the open-loop gain defined as

$$T_{OL}(s)=G(s)H(s)=G_2G_1\frac{1}{1+s/\omega_{p_2}}\frac{1+s/\omega_{z_1}}{1+s/\omega_{p_1}} \tag{2.74}$$

Let us develop (2.73) and see if we can make it fit a familiar format. If we develop and rearrange all the terms, we can show that (2.73) can be rewritten as

$$T_{CL}(s)=\frac{G_2G_1}{1+G_2G_1}\frac{1+s/\omega_{z_1}}{1+s\left(\dfrac{\dfrac{1}{\omega_{p_2}}+\dfrac{1}{\omega_{p_1}}+\dfrac{G_2G_1}{\omega_{z_1}}}{1+G_2G_1}\right)+s^2\left(\dfrac{1}{\omega_{p_1}\omega_{p_2}\left(G_2G_1+1\right)}\right)} \tag{2.75}$$

This format fits a second-order equation:

$$T_{CL}(s)=G_{CL}\frac{1+s/\omega_{z_1}}{1+\dfrac{s}{\omega_0 Q}+\left(\dfrac{s}{\omega_0}\right)^2} \tag{2.76}$$

In which we can identify the following terms:

$$G_{CL}=\frac{G_2G_1}{1+G_2G_1} \tag{2.77}$$

$$Q=\frac{\omega_0}{\omega_{p_1}+\omega_{p_2}+G_2G_1\omega_{p_1}\dfrac{\omega_{p_2}}{\omega_{z_1}}} \tag{2.78}$$

$$\omega_0=\sqrt{\omega_{p_1}\omega_{p_2}(G_2G_1+1)} \tag{2.79}$$

The closed-loop poles of (2.76) are given by the roots of its denominator, $D(s)$. What is worth noting is that the open-loop zero of $H(s)$, $\omega_{z1}$, now appears in the characteristic equation and affects the closed-loop poles. This is always true: the zeros that appear in the plant transfer function or the ones you will place in the compensator $G$ always show up in the characteristic equation. If you place low-frequency zeros to improve the phase margin in a system to be compensated, these zeros will turn into low-frequency poles once the loop is closed. Low-frequency poles imply a slowly reacting system. This is something to keep in mind when evaluating the compensation strategy to follow: is it worth selecting a high crossover frequency requiring the placement of one or two zeros at low frequency, or is it better to adopt a different crossover point and avoid assigning zeros at low frequency?

For stability purposes, what matters now is the expression of (2.76) denominator roots. Since we put our transfer equation under a known second-order form, the roots or the poles of the denominator follow the definition of (2.10):

$$s_{1,2} = \frac{\omega_{p_1} + \omega_{p_2} + G_2 G_1 \omega_{p_1} \dfrac{\omega_{p_2}}{\omega_{z_1}}}{2} \left( \pm\sqrt{1 - 4Q^2} - 1 \right) \tag{2.80}$$

From this expression, what matters is the quality factor value:

1. $Q < 0.5$: the expression under the square root is strictly positive; the roots are separate and real.
2. $Q = 0.5$: the expression under the square root is zero; the roots are coincident and real.
3. $Q > 0.5$: the expression under the square root is negative; the roots are imaginary with a real part.

In the expression of $Q$ given by (2.78), there are parameters that can change, while some are less likely to move. This is part of the design discussion to see whether these parameters can be considered real threats or are unlikely to affect the final result despite small variations. It is your responsibility, as a design engineer, to cover the cases where large variations of parameters are likely to happen. Whether they are due to operating conditions (e.g., temperature, bias points) or production spreads (e.g., tolerance, change of component), you must check their impact on the stability. Here, for the sake of the example, we will consider only the power stage zero brought by the output capacitor equivalent series resistance (ESR) described by (2.71). This parasitic element not only moves in relation to the capacitor temperature (the resistance increases as the temperature drops), it is also affected by a wide spread in production. Figure 2.10 shows the equivalent circuit of a typical capacitor. This model can be supplemented with other parasitics, such as a leakage resistance or an equivalent series inductance (ESL), but this simple approach will do for our example.

Let's assume a 470-$\mu$F output capacitor. Looking into the manufacturer datasheet, we found that its typical ESR is 90 m$\Omega$ at 25°C. However, it can vary

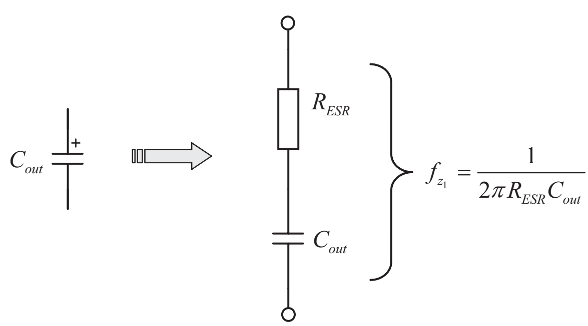



**Figure 2.10** A typical capacitor always exhibits stray elements such as an ESR. This parasitic element introduces a zero in the transfer function.

from 50 mΩ to 200 mΩ if we consider a temperature range from –40°C to +105°C and production spreads. The zero given by (2.71) will thus move between

$$f_{z1,low} = \frac{1}{2\pi \times 200m \times 470u} = 1.7 \text{ kHz} \tag{2.81}$$

and

$$f_{z1,high} = \frac{1}{2\pi \times 50m \times 470u} = 6.8 \text{ kHz} \tag{2.82}$$

For the sake of the example, we will assume that the rest of elements constitutive of $H$ and $G$ have the following values:

$$f_{p_1} = 500 \text{ Hz}$$

$$f_{p_2} = 1000 \text{ Hz}$$

$$G_1 = 0.05$$

$$G_2 = 1000$$

Based on these numbers, the quality factor $Q$ described by (2.78) will change as the zero modifies its position in relations to ESR variations. For the lowest zero position, we have

$$Q\big|_{f_{z1}=1.7 \text{ kHz}} = 0.312 \tag{2.83}$$

when the ESR reduces at higher temperature, the quality factor becomes

$$Q\big|_{f_{z1}=6.8 \text{ kHz}} = 0.97 \tag{2.84}$$

The ESR value for which $Q$ equals 0.5 is found by solving for $\omega_{z_1}$

$$\frac{\sqrt{\omega_{p_1}\omega_{p_2}(G_2G_1+1)}}{\omega_{p_1}+\omega_{p_2}+G_2G_1\omega_{p_1}\dfrac{\omega_{p_2}}{\omega_{z_1}}} = 0.5 \tag{2.85}$$

It happens when the zero reaches

$$f_{z_1} = \frac{G_2G_1\omega_{p_1}\omega_{p_2}}{2\sqrt{\omega_{p_1}\omega_{p_2}(1+G_2G_1)}-\omega_{p_1}-\omega_{p_2}}\frac{1}{2\pi} = 2.9 \text{ kHz} \tag{2.86}$$

With a 470-$\mu$F capacitor, it corresponds to an ESR of 116 mΩ.

As the ESR varies, it affects the quality factor and changes the nature of the roots. These roots can therefore be purely real; in that case, $Q = 0.312$, and the step response of the closed-loop transfer function is nonringing. According to (2.80), when we have a 1.7-kHz zero, these roots have the following values:

$$s_1 = -90k$$

$$s_2 = -11.1k \tag{2.87}$$

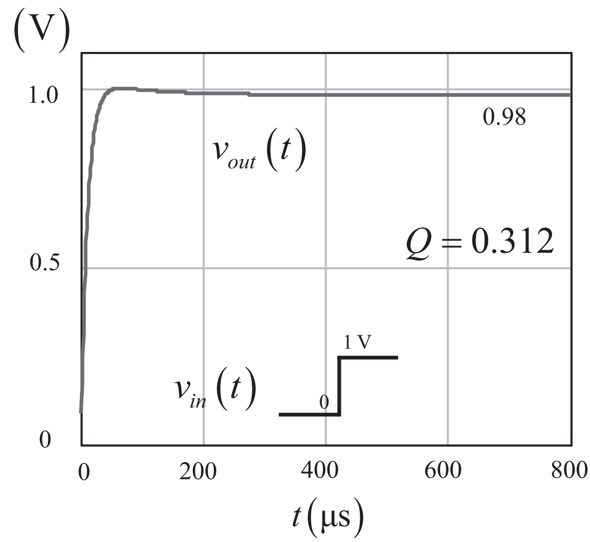



**Figure 2.11** The step response for real roots is a nonringing signal with a very minor overshoot.

If we multiply (2.75) by 1/*s* and extract the time-domain response, we obtain, for these conditions, a signal plotted in Figure 2.11. Please note the presence of a light overshoot though, despite a *Q* less than 0.5. This overshoot is brought by the zero presence that affects the transient response but not the steady-state value. The final value is not 1 V as expected. This is normal; the closed-loop gain defined by (2.77) is less than 1 (0.98 exactly): we have a permanent *static error* of 20 mV.

When the ESR exactly equals 116 mΩ, both roots are coincident and equal:

$$s_1 = s_2 = -32k \tag{2.88}$$

The step response appears in Figure 2.12. Theory tells us that the step response of a second-order system featuring coincident poles should not bring overshoot, but

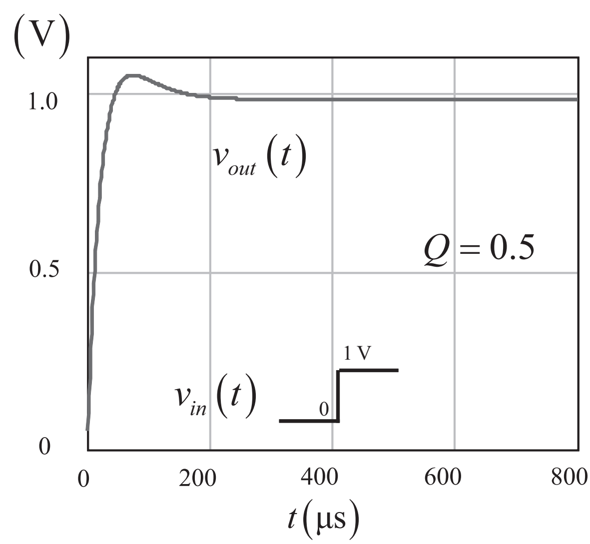



**Figure 2.12** With coincident poles, the response is fast, and the overshoot is still weak.

we can see it in the picture. This is because the study considers only a system having only poles, not an extra zero as we have here. Its presence changes the transient response.

Now, when the ESR is really low, the roots become imaginary conjugates:

$$\begin{aligned} s_1 &= -16.3k - j27.2k \\ s_2 &= -16.3k + j27.2k \end{aligned} \tag{2.89}$$

In that case, a more pronounced overshoot starts to appear, as shown in Figure 2.13.

If for any reason the ESR would become negligible (i.e., you chose a multilayer type of capacitor or parallel a lot of low-ESR capacitors), the roots would simply become

$$\begin{aligned} s_1 &= -5.2k - j31.3k \\ s_2 &= -5.2k + j31.3k \end{aligned} \tag{2.90}$$

The corresponding step response exhibits more overshoot than the previous signal but is still stable. Figure 2.14 details the results:

From (2.90), we can compute the natural angular frequency $\omega_0$ as

$$\omega_0 = |s_1| = |s_2| = \sqrt{5.2k^2 + 31.3k^2} = 31.73 \text{ krd/s or } 5.05 \text{ kHz} \tag{2.91}$$

A similar value is obtained if you use (2.79) or (2.89).

As can be observed, by changing the ESR values, the roots exhibit different real and imaginary coefficients. Actually, we can see that a decreasing ESR makes the quality factor grow while the real part of the roots decreases. This makes sense: the real part of the roots expresses losses that damp the second-order system. The $Q$ increase is simply the result of these damping effects going down as the ESR vanishes.

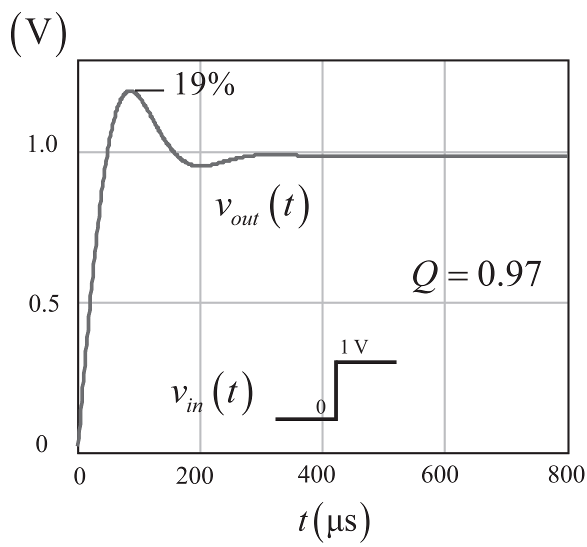



**Figure 2.13** As the imaginary coefficients now appear, the closed-loop ac response starts to peak, inducing some ringing in the time-domain response.

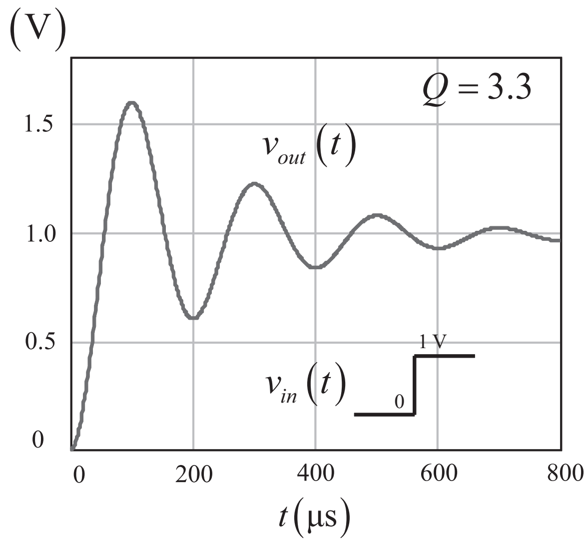



**Figure 2.14** Despite a pronounced overshoot, the response is still stable.

## 2.4 S-Plane and Transient Response

In stability analysis, it is interesting to plot the trajectory of these roots and see how they move in relation to a considered parameter. In the present case, we have chosen the ESR of the output capacitor, but in most of the textbooks the selected parameter is a gain coefficient $k$. A typical example shown in the literature is that of Figure 2.15.

This is a unity-gain return control system, and it is easy to derive the transfer function from input to output:

$$\frac{V_{out}(s)}{V_{in}(s)} = \frac{kG(s)H(s)}{1 + kG(s)H(s)} \tag{2.92}$$

$G$ and $H$ are individual transfer functions, made of a numerator and a denominator:

$$G(s) = \frac{N_G(s)}{D_G(s)} \tag{2.93}$$

$$H(s) = \frac{N_H(s)}{D_H(s)} \tag{2.94}$$

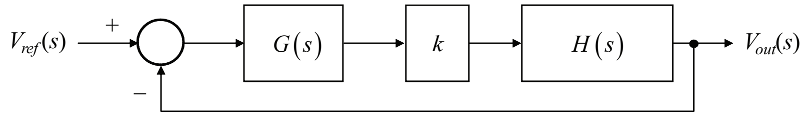



**Figure 2.15** In this typical example, the parameter $k$ is a gain inserted in the control loop and subject to wide variations.

If we now reinject these definitions in (2.92), we obtain:

$$\frac{V_{out}(s)}{V_{in}(s)} = \frac{k\dfrac{N_G(s)}{D_G(s)}\dfrac{N_H(s)}{D_H(s)}}{1+k\dfrac{N_G(s)}{D_G(s)}\dfrac{N_H(s)}{D_H(s)}} = \frac{kN_G(s)N_H(s)}{D_G(s)D_H(s)+kN_G(s)N_H(s)} \tag{2.95}$$

This equation shows that the poles and zeros put in $G(s)$ to shape the open-loop response for adequate crossover frequency and phase margin now appear in the closed-loop transfer equation with the terms $N_G(s)$ and $D_G(s)$. The characteristic equation often noted $\chi(s)$ (Chi, pronounced like "key") can be rewritten as:

$$\chi(s) = D_G(s)D_H(s) + kN_G(s)N_H(s) \tag{2.96}$$

This is an interesting equation because it shows that the zeros you will put in the compensator $G$ (e.g., to boost the phase margin at low frequency) will turn into poles when the loop is closed. A low-frequency open-loop zero turning into a low-frequency closed-loop pole will slow down the response to an incoming perturbation or a setpoint change. If we now look at the parameter $k$, we can see several cases, depending on whether $k$ is small or high:

- $k$ is small, then the characteristic equation can simplify to $\chi(s) = D_G(s)D_H(s)$: the closed-loop poles are those already present in the open-loop gain equation.
- $k$ now increases, and as the roots are a continuous function of $k$ they move away from their open-loop definition to fully satisfy (2.96).
- If $k$ increases further and becomes really high, then the left side of (2.96) can be neglected and the definition becomes: $\chi(s) = kN_G(s)N_H(s)$. The closed-loop poles are now given by the open-loop zeros!

It is important to check the values taken by the roots as $k$ changes. For instance, are there some conditions where the real parts of the roots dangerously diminish, contributing to eliminate the damping? Can these roots suddenly reverse their negative sign and turn positive, making the system output diverging with all consequences behind? This makes sense, since as we discovered with (2.70) the poles directly influence the time-domain response of our system.

Network theory teaches us that the *complete* or *total* response of a system is generally the sum of its *natural* or *free* response $r_n(t)$ and its *forced* response $r_f(t)$. The free response is obtained by setting the input $u(t)$ to zero and considering nonzero initial conditions. On the contrary, the forced response is obtained by solely considering the input and setting all initial conditions to zero. The forced response in (2.70) was 166 $\mu$V, whereas the rest of the terms composed the natural response. The response of a SISO system is thus given by

$$y(t) = r_n(t) + r_f(t) = \sum_{i=1}^{n} C_i e^{p_i t} + r_f(t) \tag{2.97}$$

In this expression, $p_i$ are the poles of the characteristic equation, and $C_i$ are the coefficients of the exponential terms. The number of poles depends on the polynomial degree of the denominator: two poles for a second-order system, three for a third-order system, and so on. The easiest way to study these roots is to place them on a dedicated map called an Argand diagram, commonly denoted as the *s*-plane. The variable *s* is classically defined by

$$s = \sigma + j\omega \tag{2.98}$$

The *s*-plane is thus a two-dimensional graph where the vertical axis represents the imaginary portion of *s*, $j\omega$, and the horizontal axis, its real part, $\sigma$. The area of the plane corresponding to negative roots is called the left half plane (LHP), whereas the area situated to the right of the vertical axis is simply called the right half plane (RHP); the roots placed in this area have positive real parts. The poles are illustrated by a cross, ×, whereas the zeros are designated via a circle, ○. The location of the poles in the *s*-plane will affect the signal delivered by (2.98) as follows:

1. If the pole is real, $p_i = -\sigma$, it is placed in the LHP; it is called a LHP pole. Its contribution to the response is in the form $Ce^{-\sigma t}$, a decaying exponential component of duration $1/\sigma$. Therefore, if the pole location is far from the origin of the *s*-plane, then the response is a quickly decaying signal. For the opposite, as the pole's location moves closer to the 0 point, the response will be a slowly decaying signal. Suppose we have in the time-domain expression a term looking like $5e^{-0.1t}$; the signal starting with an amplitude of 5 V will decay and lasts 10 s before dying out to zero.
2. A pole appearing at the origin, $p_i = 0$, defines a term of constant amplitude, defined by the initial conditions: $5e^{-p_i t} = 5e^0 = 5$ V
3. If the pole is real but this time appears in the RHP, it is called a RHP pole and is of the form $p_i = \sigma$. It is a positive root. Its contribution to the output signal is of the form $Ce^{\sigma t}$, a continuously growing component. If you deal with a system featuring a closed-loop RHP pole, it is unstable.
4. When solving the characteristic equation roots, you can find conjugate pairs of roots, leading to a form $p_i = -\sigma \pm j\omega$. In this case, the contribution to the time-domain signal is a decaying sinusoidal signal of frequency $\omega$ obeying the form $Ae^{-\sigma t}\sin(\omega t + \varphi)$. $A$ and $\varphi$ are imposed by the initial conditions. Again, the real part $\sigma$ represents the losses that damp the response. It is important to note that if the real part is 0, we have imaginary pole pairs of the form $p_i \pm j\omega$. This is an undamped oscillatory component of frequency $\omega$.
5. If the complex pole pair is found to be in the RHP, $p_i = \sigma \pm j\omega$, the response is an exponentially growing sinusoidal signal.

A graphical representation of these various cases appears in Figure 2.16. This is a system featuring two poles, hence a second-order type. These poles could be that of a *LC* filter in which the quality factor is purposely adjusted for instance by an added resistance. In the first part (a), the poles are separate and purely real. The quality factor is very low; the system is overdamped. The response is similar to that of two cascaded *RC* filters. In the second plot (b), the poles are coincident; the quality factor is equal to 0.5. The response is faster but there is no overshoot as the roots do not include imaginary parts. In (c), the poles are split and represented

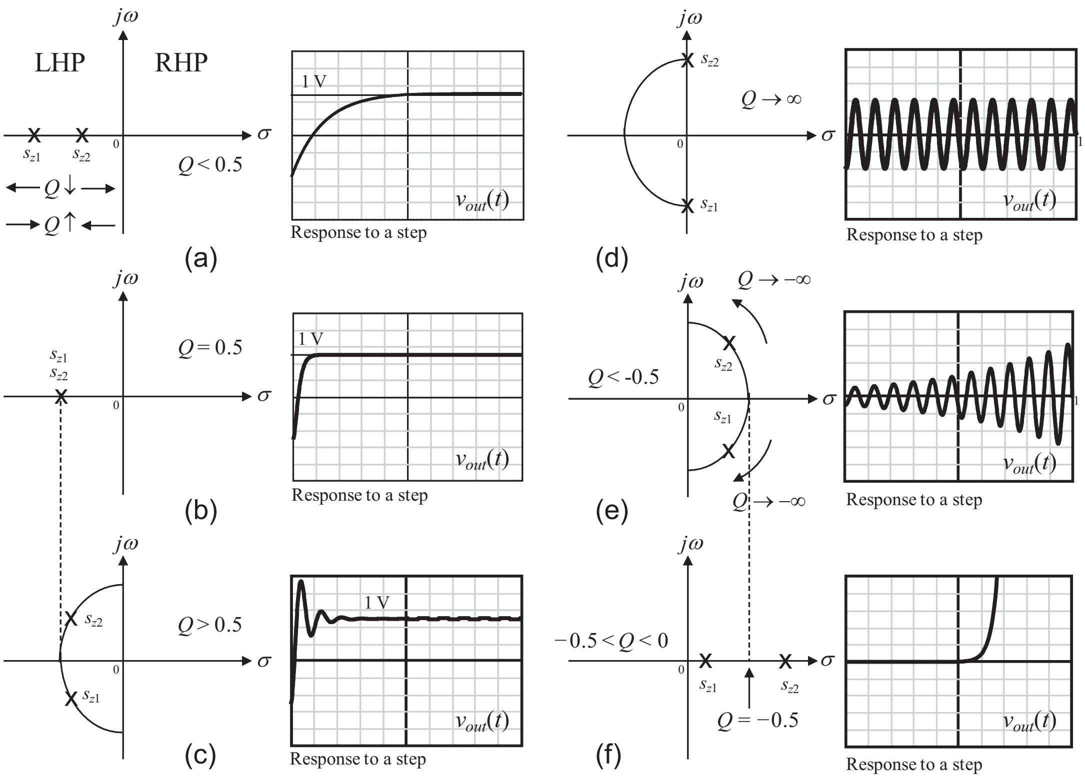



**Figure 2.16** Depending on the closed-loop denominator roots position, the output response of a system converges if there are poles occurring in the LHP. If RHP poles appear, the system is unstable.

by conjugate roots. The quality factor is beyond 0.5. An imaginary portion is now present, and you see oscillations. However, the real parts provide the damping and represent the losses in the network that calm down the oscillations. This is a decaying signal. In (d), all losses have disappeared, and the system response is purely oscillatory. The roots are imaginary, and the real parts (the damping) have gone: we have sustained oscillations. In (e), the quality factor is negative and the poles have jumped in the RHP. The exponential exponent is now positive and oscillations grow, the system diverges. In (f), the imaginary portions have disappeared, the system still diverges but without oscillations.

Figure 2.17 shows another example where several poles and zeros are represented. They correspond to the following equation roots:

$$H(s) = \frac{\left[(s+2)^2 + 4\right]}{\left[(s+1)^2 + 1\right]} \frac{(s-1)}{(s+3)} \tag{2.99}$$

Please note than one of the zeros, $s_{z3}$, lies in the RHP; it is a positive root.

Based on what (2.97) taught us, the time-domain response of such a transfer function to a 1-V step is made of three terms (three poles) plus the forced response. The forced response is found by calculating the dc gain of (2.99):

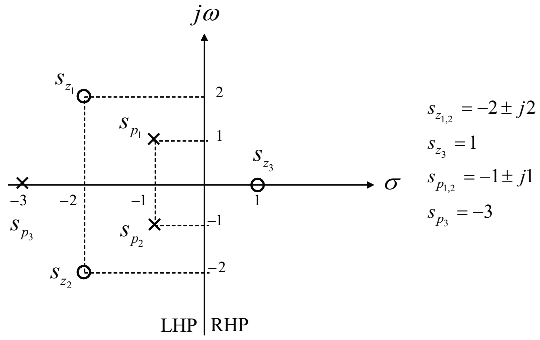



**Figure 2.17** The *s*-plane allows the placement of poles and zeros and helps to check how they move in relationship to a selected parameter.

$$H(0) = \frac{\left[(2)^2 + 4\right]}{\left[(1)^2 + 1\right]} \frac{(-1)}{(3)} = -\frac{4}{3} \tag{2.100}$$

The natural response is made of one pure decaying term (one real pole) and two other terms that are damped sinusoidal signals (conjugate poles pair):

$$\mathcal{L}^{-1}\left\{\frac{1}{s}H(s)\right\} = y_1(t) = \frac{4}{3}e^{-3t} + e^{-t}\cos(t) + 3e^{-t}\sin(t) - \frac{4}{3} \tag{2.101}$$

When all terms have died out to zero, the output is –1.33 V. As you can see, the zeros do not explicitly appear in the various terms, but they play a role in the coefficients—and also in the polarity: the output is negative. Why? We have three zeros and three poles. The two LHP poles and zeros phase lag/lead compensate for each other. The third pole lags by 90°, so what about the RHP zero (RHPZ)? It also lags by 90°, making a total of –180° or a polarity reversal, bringing a negative output for a positive 1-V step! The RHP zero, rather than adding a phase lead as a LHP zero would, further degrades the phase by lagging, just as a pole does. We will see that in more details in a few paragraphs.

Suppose now that $s_{z3}$ becomes a LHP zero, the response would transform into

$$y_2(t) = \frac{2}{3}e^{-3t} + e^{-t}\cos(t) + e^{-t}\sin(t) + \frac{4}{3} \tag{2.102}$$

The sign is now positive since we have an even number of zeros and poles, making the total phase change to 0°: a positive step gives a positive voltage. Please note that shifting of the zero into the LHP has affected the *C* coefficients but not the exponential terms that remain exclusively linked to the locations of the poles.

Now, assume the third pole becomes a RHP pole, $s_{p3} = 3$ in (2.99) and the third zero is still in the LHP as in (2.102). The time-domain response can be expressed as follows:

$$y_3(t) = \frac{116}{51}e^{3t} + \frac{1}{17}e^{-t}\cos(t) - \frac{13}{17}e^{-t}\sin(t) - \frac{4}{3} \tag{2.103}$$

The first term is enough to make the system unstable: the exponent is positive and forces the expression to diverge as $t$ increases. This is the RHP pole effect. If you look at the last term, it is negative again: the two LHP pole/zero pairs compensate each other with a total 0° phase. However, the third LHP zero adds a 90° phase lead that is normally compensated by 90° lag brought by a LHP pole. Here, as this pole is in the RHP, it brings another 90° phase lead, whereas it should be a lag, as with a LHP pole: the total phase lead is now 180°, or another phase reversal. However, as the first positive exponential term immediately dominates the response, you will not see the negative output.

We have graphed all three responses in Figure 2.18. As expected, the RHP pole brings a diverging output. The RHP zero, despite bringing an additional phase lag, does not make the system diverging. As a preliminary conclusion, if a RHP zero forces you to be more careful when compensating the converter, a single RHP pole in the characteristic equation (thus a closed-loop pole) is absolutely insurmountable.

### 2.4.1 Roots Trajectories in the Complex Plane

The analysis of the characteristic equation (the closed-loop transfer function denominator) is usually carried at a certain operating point. However, we have seen that variables affecting the characteristic equation can move. This is often the case for various gain values, as $k$ in the previous example. At the time scientists studied power electronics in the 1950s, valve-based amplifiers could have wide gain dispersions at warm-up or at various ambient temperatures, and stability could be affected. A significant benefit of feedback was to program the gain via an external set of resistors, making the final chain gain insensitive to the amplifier gain varia-

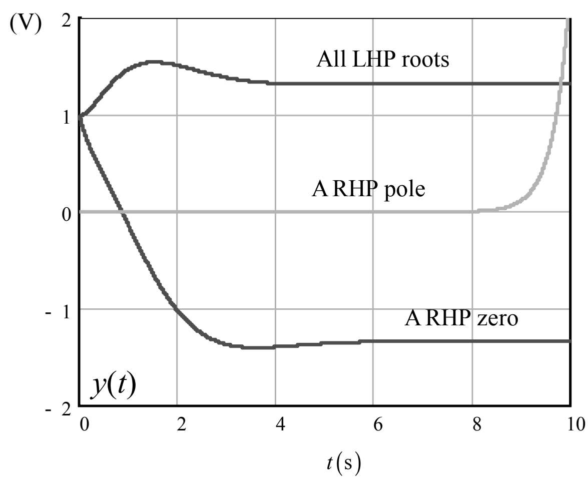



**Figure 2.18** Various time-domain outputs depending where the third poles and zeros are in the $s$-plane.

tions. Besides gain variations, there can be perturbations effects such input voltage or load changes but also open-loop poles or zeros that can slide along the frequency axis when production spreads or loading conditions are involved. In our example from Figure 2.9, sweeping the output capacitor ESR value degraded the transient response. Mathematically, it meant that the roots evaluated at different ESR cases have changed their values. If we would plot these roots in the *s*-plane and link all points together, we would obtain a so-called Evans *root-locus* representation, named after the work of W. R. Evans in the 1950s. By looking at the path taken by the poles, we can check if certain values of the swept parameter bring the roots close to the vertical axis (pure imaginary roots, no damping) or, even worse, make them jump on the right side of the *s*-plane.

Several mathematical programs can compute and plot the poles or zeros on the *s*-plane when a given parameter is swept. Mathcad is one of them with which root locus graphs are very simple to obtain. For instance, Figure 2.19 depicts a root locus plot for (2.80) as the zero $s_{z1}$ is swept from 1.5 kHz to more than 100 kHz. The appendix at the end of this chapter shows how we plotted this picture.

Should the roots approach the imaginary axis or worse, move to the right-half place section, stability would be at stake, as exemplified in Figure 2.16. Fortunately, on the drawing, we see that despite an ESR zero going to almost infinity (the ESR vanishes to 0), the roots always remain in the left plane, guaranteeing some damping. Calculations show that in these conditions, $Q$ does not exceed 3.5 and the step response is stable despite a pronounced overshoot. The trajectory describing the loci of the poles can reveal a lot of information, but the study of the technique is outside the scope of this book. The reader interested in an in-depth analysis of this method will find a lot of information in Chapters 13 and 14 of [6].

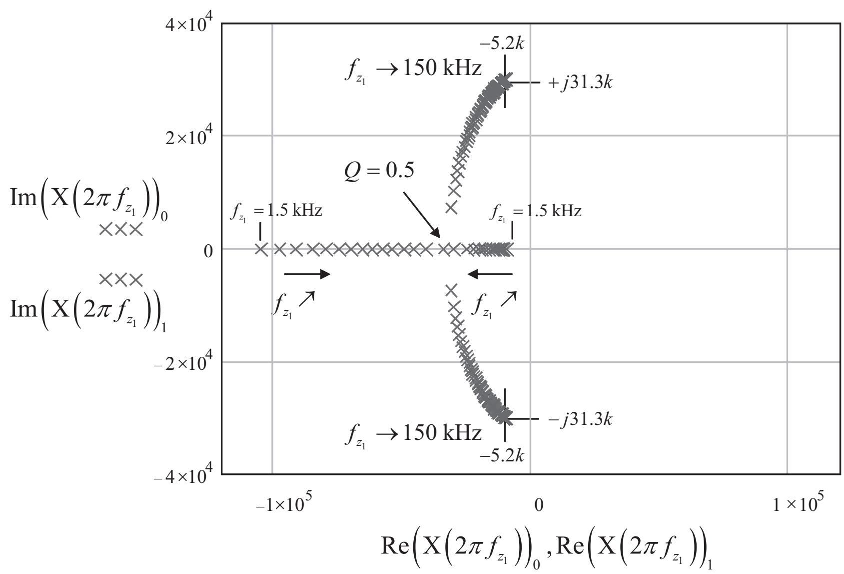



**Figure 2.19** In this picture, we can clearly see how the roots move as the ESR zero is swept from 1.5 kHz up to 150 kHz.

## 2.5 Zeros in the Right Half Plane

Right half plane zeros are usually found in switching converters transferring the energy to the load in two steps: the energy is first stored in the inductor during the on-time and then dumped into the load during the off-time. During the on-time, the load is isolated from the input source while the inductor is energized. During the off-time, the energy stored in the inductor is released to the load: boost, buck-boost, or flyback converters operate this way. These three converters are called indirect energy-transfer converters. They all obey the two-step conversion process we described. For the opposite, the buck converter is a direct energy-transfer converter. You do not need an intermediate step to store energy before transmitting it to the source. This intermediate storing step *actually creates a delay* because the controller must always go through an energy-storing step before answering the increase in the delivered power need. Should it take time to store more energy while the power demand is fast, the converter cannot momentarily keep up its power delivery and the output voltage falls. This event lasts until the stored energy in the inductor has increased.

The output variable momentarily going in the opposite direction compared to what the control expects is the typical signature of a transfer function featuring zero located in the right-half portion of the $s$-plane, also called a RHP zero or RHPZ.

### 2.5.1 A Two-Step Conversion Process

Figure 2.20(a) represents a classical boost converter where two switches appear: a power switch $SW$, usually a MOSFET, and a diode, sometimes called the catch diode. In the continuous conduction mode (CCM) of operation, the inductor current $i_L$ flows in the power switch $SW$ during the on-time or $dT_{sw}$, where $d$ is the instantaneous duty ratio and $T_{sw}$ the switching period. During the off-time, or $(1 - d)T_{sw}$, the power switch is open and the output diode routes the current to an output network made of the capacitor and the load. Regardless of the control method, voltage or current-mode, this configuration assumes that energy is first stored in the inductor during the on-time and then transferred to the output during the off-time.

Figure 2.20(b) shows an equivalent representation of the boost converter, where the switch/diode network has been replaced by a single pole double throw switch

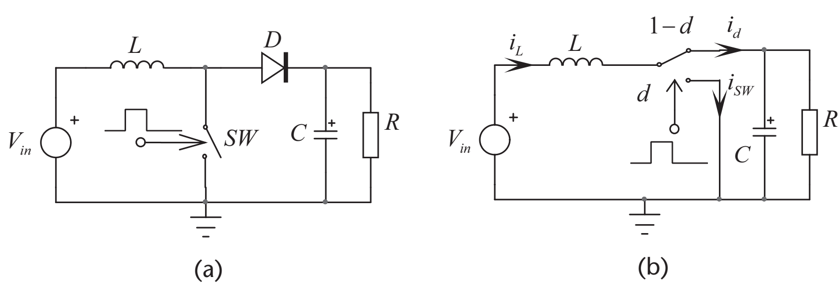



**Figure 2.20** A boost converter features two power switches. They can be replaced by a single-pole double-throw switch that represents the operations of the diode and the transistor.

that alternatively routes the inductor current in the two different branches: the power switch or the output diode. If a designer would observe the currents circulating in the output diode, he or she would see a typical waveform shown in bold in Figure 2.21. Our boost converter is designed to deliver power to a given load. The variable of interest, in our case, is thus the available output current $I_{out}$. This current is actually made of a dc portion on which is superimposed a switching ripple. In theory, the ripple goes into the capacitor and the dc current circulates in the load. The dc current delivered by the boost converter is nothing other than the diode average current $I_d$. Mathematically, this current can be expressed by

$$I_{out} = I_d = I_L(1 - D) \tag{2.104}$$

where $I_d$ is the average diode current also equal to the dc output current $I_{out}$. $D$ represents the averaged duty ratio.

If we graph the circulating currents in the diode at the switch opening, we obtain the drawing presented in Figure 2.21. When the switch opens, the current no longer circulates in the power switch but is routed, via the diode, to the output. The average value—or the dc current—circulating in the load is the area of the surface $A_0$, averaged over a switching cycle $T_{sw}$.

$$I_{out} = \frac{A_0}{T_{sw}} \tag{2.105}$$

Now, if a sudden current demand occurs on the output, the controller senses the transient and immediately increases the duty ratio by a small value— $\hat{d}$—to build more energy in the inductor. This is what Figure 2.22 shows.

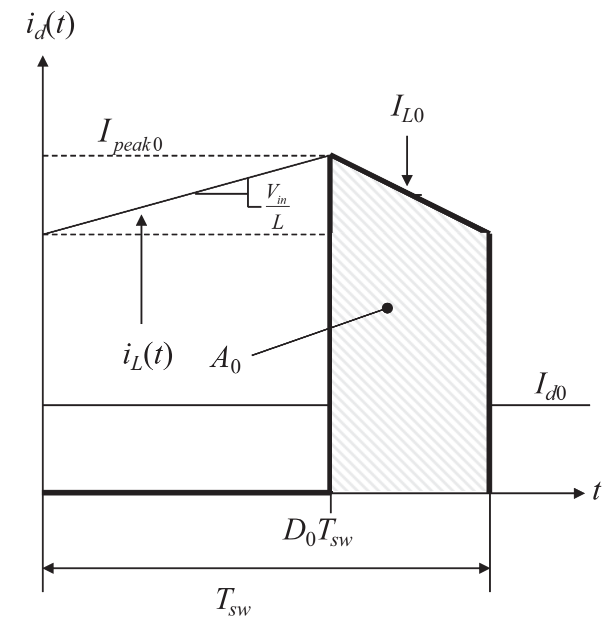



**Figure 2.21** The current observed in the output diode at the beginning of the event.

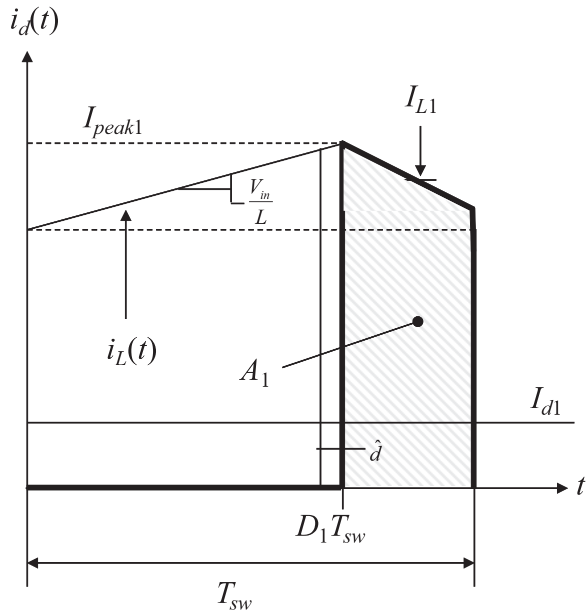



**Figure 2.22** As an answer to the output current demand, the controller asks the inductor to store more energy.

In theory, the new surface, $A_1$, should be larger than $A_0$ to cope with the output power demand increase. However, as the switching period is fixed, the increase in the on-time duration simply shortens the off-time interval. As this is the time during which the current circulates in the diode to satisfy $A_1 > A_0$, the only condition is that the new peak current $I_{peak1}$ is larger than the first one, $I_{peak0}$. This is what is sketched in Figure 2.23.

### 2.5.2 The Inductor Current Slew-Rate Is the Limit

What is the pace at which the average inductor current can change? Lenz's law instructs us that the instantaneous current change rate in an inductor obeys the following formula:

$$\frac{di_L(t)}{dt} = \frac{v_L(t)}{L} \tag{2.106}$$

On average, over a switching cycle, it simply follows

$$\left\langle \frac{di_L(t)}{dt} \right\rangle = \frac{\langle v_L(t) \rangle}{L} = \frac{V_L}{L} \tag{2.107}$$

The exercise now consists of calculating the average value across our inductor. By considering the weighted period of time during which $V_{in}$ or $V_{out} - V_{in}$ are applied across $L$, we have

$$V_L = V_{in}D - (V_{out} - V_{in})(1 - D) = V_{out}(D - 1) + V_{in} \tag{2.108}$$

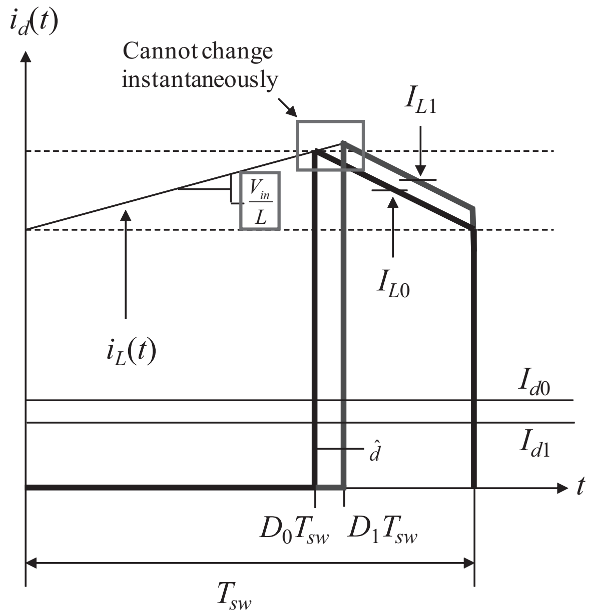



**Figure 2.23** As the on-time reduces the off-time duration, the only way to pass more power is to make sure the new peak current is larger than the first one. Unfortunately, the inductor opposes current changes.

Let's assume the following boost operating parameters:

$V_{in} = 10$ V
$D_0 = 0.583$
$V_{out} = V_{in}/(1 - D_0) = 24$ V
$R_{load} = 240\ \Omega$
$L = 1$ mH
$F_{sw} = 100$ kHz

With a 58.3 percent duty ratio, the converter delivers 24 V. We are in steady state and (2.108) gives 0. Now suppose that the duty ratio jumps to $D_1 = 59$ percent or a difference of 0.7 percent. What is the inductor average current slope in this case? Considering a large output capacitor, the output voltage stays constant during the duty ratio change. Applying (2.108) gives a transient average inductor voltage of

$$V_L = V_{out}(D_1 - 1) + V_{in} = 24 \times (0.59 - 1) + 10 = 160 \text{ mV} \tag{2.109}$$

Back to (2.107), the maximum average current slope authorized by the inductor is therefore

$$\left\langle \frac{di_L(t)}{dt} \right\rangle = \frac{V_L}{L} = \frac{160m}{1m} = 160\ \mu\text{A}/\mu\text{s} \tag{2.110}$$

a rather modest value.

When the duty ratio changes from 58.3 percent to 59 percent, it implies an output voltage change of

$$V_{out} = \frac{V_{in}}{1-D} = \frac{10}{1-0.59} = 24.39 \text{ V} \quad (2.111)$$

With a constant 240-Ω load, the output current will increase to

$$I_{out} = \frac{V_{out}}{R_{load}} = \frac{24.39}{240} = 101.6 \text{ mA} \quad (2.112)$$

Brought back to the inductor change, the output current variation given by (2.112) must be accompanied by an average inductor current variation of

$$\Delta I_L = \frac{V_{in}}{R_{load}}\left[\frac{1}{(1-D_1)} - \frac{1}{(1-D_0)}\right] = \frac{10}{240}\left[\frac{1}{(1-0.59)} - \frac{1}{(1-0.583)}\right] = 1.7 \text{ mA} \quad (2.113)$$

Given an average inductor slope 160 $\mu$A/$\mu$s, this current variation will only be possible within a timeframe of

$$dt = \frac{1.7m}{160u} = 10.6 \ \mu\text{s} \quad (2.114)$$

If the duty ratio is swept from 58.3 percent to 59 percent in less than 10.6 $\mu$s, the inductor current will not build up at a sufficient pace to make the output current rise at the same speed. As an immediate result, the output current drops rather than increases. On the contrary, if the duty ratio sweep is slow enough, the current can increase in the inductor at sufficient speed to compensate the reduction in $(1-d)$: the output voltage goes up. How do we make sure the inductor will always have time to build up enough current in case a fast transient occurs? By simply rolling off the crossover frequency or, in other terms, severely limiting the converter bandwidth so that fast transient demands never translates into a fast duty ratio change. If you fail to limit the bandwidth, in a fast output power transient demand, the output voltage will go down despite an increase of duty ratio. In control theory, you have reversed the control loop and oscillations occur. This situation lasts until the current in the inductor builds up to the right value. To prevent this from happening, the RHP zero effect naturally limits the available bandwidth for a given converter. Intuitively, a converter featuring a large inductor, hence operating in a deep CCM, will have a low-frequency RHP zero, severely hampering its possible response time.

### 2.5.3 An Average Model to Visualize RHP Zero Effects

Average models lend themselves very well to illustrating the effects of a RHP zero. We have built an open-loop boost converter around an auto-toggling model described in [7]. Figure 2.24 portrays the adopted schematic. In this test fixture, a 1-mH inductor driven at a 100-kHz switching period delivers 100 mA to a load ($V_{out}$ = 24 V). The duty ratio is first slowly swept between 58.3 percent and 59 percent. As shown in Figure 2.25, the inductor current nicely follows the demand,

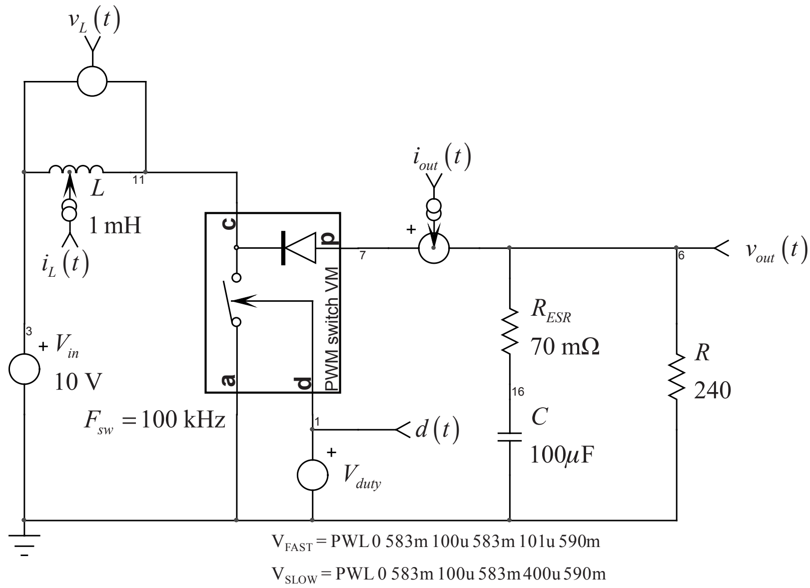



**Figure 2.24** A voltage-mode average model whose control input is swept at two different paces can demonstrate the existence of a RHPZ in the CCM-operated boost converter.

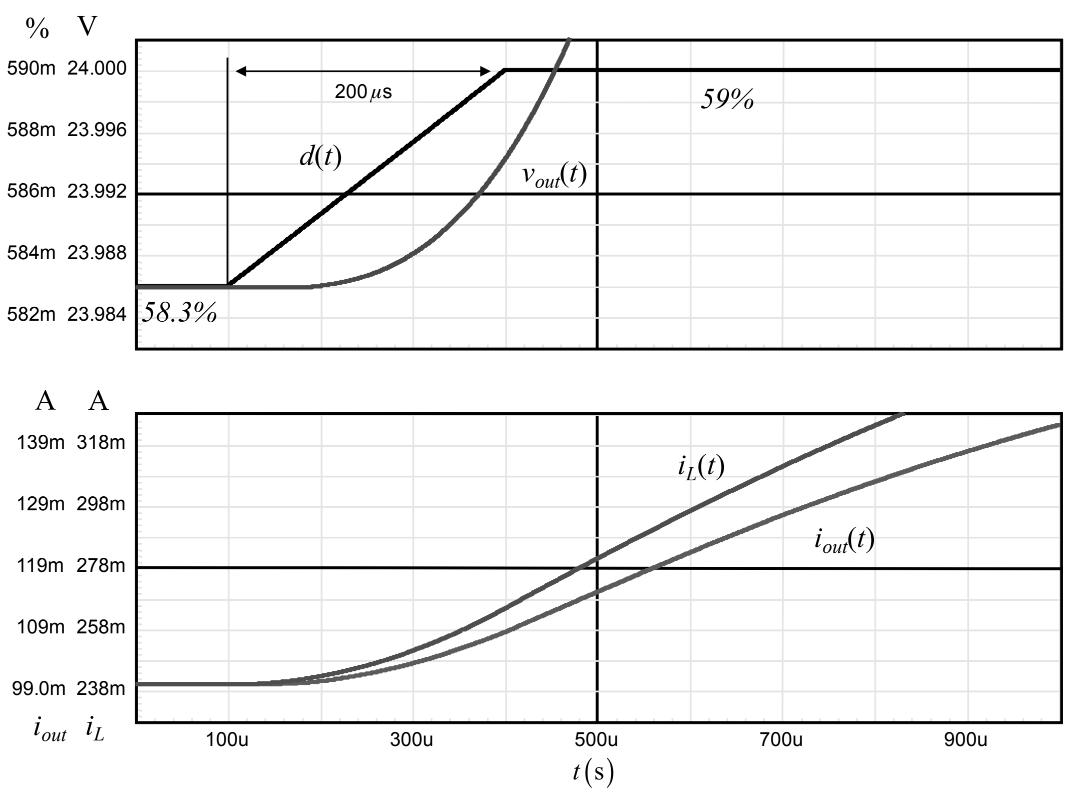



**Figure 2.25** When the duty ratio slowly changes and gives time to the inductor current to build up, the output voltage variation is positive, as it should be.

and the output voltage slope is always positive. If we now sweep the duty ratio at a higher speed, Figure 2.26 indicates that despite a regular slope in the inductor current, it does not build up at a sufficient pace to answer the output current demand. As a result, both $v_{out}(t)$ and $i_{out}(t)$ drop. If the system would operate in closed loop, oscillations would occur because the control law is reversed: the duty ratio increases but the output voltage goes down.

How do we prevent this problem from happening? Well, a solution is to clamp the maximum slew rate on the duty ratio control input. In that way, even if a sudden variation is detected on the output, the error voltage will always rise at a speed where the inductor volt-second limit is never reached, giving sufficient time for the inductor current to build up. How do we limit the slew rate? By rolling off the crossover frequency $f_c$ at a position usually well below the worst-case RHP zero position.

### 2.5.4 The Right Half Plane Zero in the Boost Converter

We have shown the consequences of a RHPZ in a boost converter. It is now interesting to analytically derive its position in the transfer function of the boost converter. To simplify the analysis, we will consider a voltage-mode control. The reader interested in current-mode control and compensation examples can find more informa-

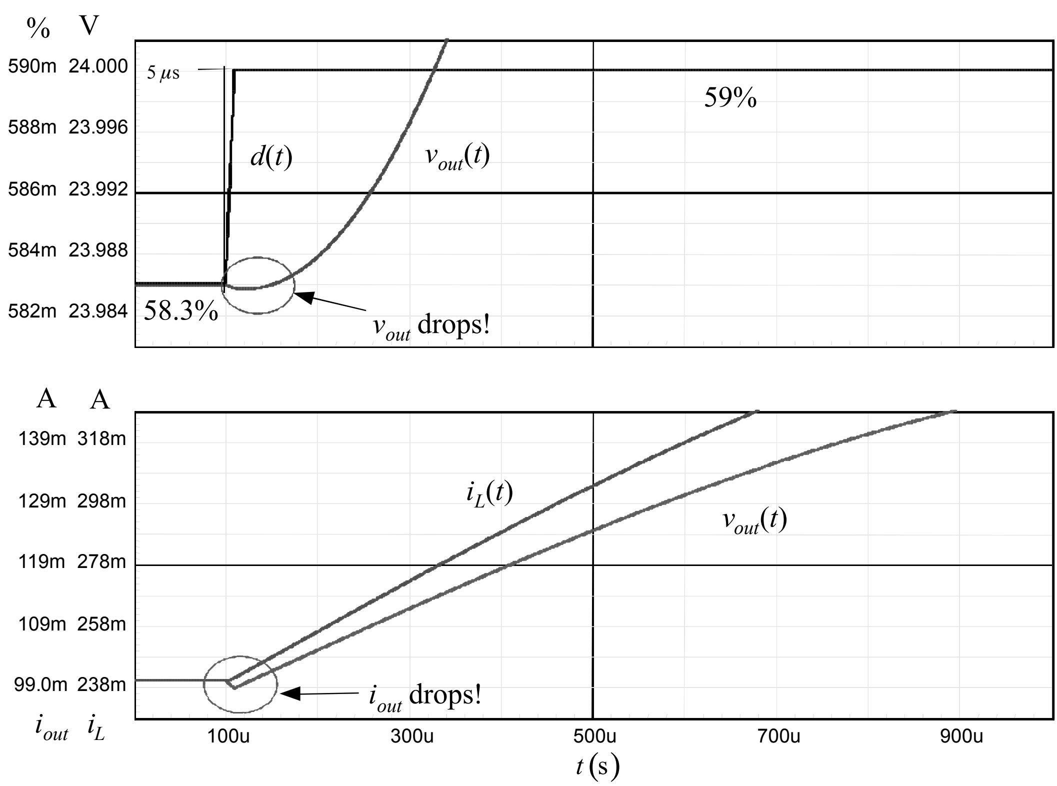



**Figure 2.26** In this example, the inductor current builds up too slowly with regard to the output current demand. As a result, the output current drops until the current in the inductor builds up to the right value.

tion in [8]. To derive the transfer function, we can start from the output current expression given in (2.104). This is a large-signal (nonlinear) equation that we must transform into a small-signal expression. The fastest way to do that is to find partial derivatives coefficients for each of the variables, the duty ratio $D$ and the inductor current $I_L$:

$$\hat{i}_{out} = \left(\frac{\partial I_{out}}{\partial I_L}\hat{i}_L\right)_D + \left(\frac{\partial I_{out}}{\partial D}\hat{d}\right)_{I_L} = \hat{i}_L(1-D) - \hat{d}I_L \tag{2.115}$$

In this equation, the ac inductor current $\hat{i}_L$ appears. What is the expression of an ac inductor current? Simply, it's the ac inductor voltage divided by the inductor impedance. Let us find the expression of the ac inductor voltage by first deriving its average large signal expression, already found in (2.109):

$$V_L = V_{out}(D-1) + V_{in} \tag{2.116}$$

On average, when the converter is at the equilibrium, this equation gives zero. However, under an ac excitation, the average inductor voltage is also ac modulated around zero. By using a partial derivative, we can see that the ac inductor voltage, in this case, is expressed by

$$\hat{v}_L = \hat{v}_{out}(D-1) + \hat{d}V_{out} \tag{2.117}$$

In this equation, the input term $V_{in}$ has disappeared since the input voltage is considered constant during the ac analysis. Furthermore, if we consider a large output capacitor, its impedance at the ac excitation can be considered close to zero. In this case, if we consider $\hat{v}_{out} \approx 0$, we can further simplify the expression:

$$\hat{v}_L \approx \hat{d}V_{out} \tag{2.118}$$

With the ac inductor voltage on hand, it is easy to obtain the ac inductor current we are looking for:

$$\hat{i}_L(s) = \frac{\hat{v}_L(s)}{Z_L} = \frac{\hat{d}(s)V_{out}}{sL} \tag{2.119}$$

Substituting (2.119) in (2.115) gives the final ac output current expression:

$$\hat{i}_{out}(s) = \frac{\hat{d}(s)V_{out}}{sL}(1-D) - \hat{d}(s)I_L \tag{2.120}$$

The average inductor current $I_L$ is the source current $I_{in}$. Considering a 100 percent efficiency power conversion, we can write

$$V_{in}I_{in} = V_{out}I_{out} = \frac{{V_{out}}^2}{R} \tag{2.121}$$

From which we have

$$I_{in} = I_L = \frac{V_{out}^{\ 2}}{V_{in}R_{load}} = \frac{V_{out}}{V_{in}}\frac{V_{out}}{R_{load}} = \frac{V_{out}}{(1-D)R_{load}} \tag{2.122}$$

Substituting (2.122) in (2.120), we obtain

$$\frac{\hat{i}_{out}(s)}{\hat{d}(s)} = \frac{V_{out}D'}{sL} - \frac{V_{out}}{D'R_{load}} \tag{2.123}$$

Now factoring the first term and rearranging, we have

$$\frac{\hat{i}_{out}(s)}{\hat{d}(s)} = \frac{V_{out}D'}{sL}\left(1 - \frac{sL}{D'^2 R_{load}}\right) = \frac{\left(1 - \frac{s}{\omega_{z2}}\right)}{\frac{s}{\omega_0}} \tag{2.124}$$

where

$$\omega_0 = \frac{V_{out}D'}{L} \tag{2.125}$$

$$\omega_{z_2} = \frac{R_{load}D'^2}{L} \tag{2.126}$$

This expression links the output current to the duty ratio input. In this equation, we can see a pole at the origin given by the inductor $L$ and a zero featuring a positive root: this is the RHPZ $\omega_{z_2}$ we are looking for. Please note that both roots depend on the duty ratio and are moving in relation to the input/output conditions.

If we apply the boost converter numerical values from Figure 2.24, we have the following positions:

$$f_0 = 1.6 \text{ kHz} \tag{2.127}$$

$$f_{z_2} = 6.6 \text{ kHz} \tag{2.128}$$

Our interest now lies in the phase lag brought by this transfer function. The argument of a quotient is the numerator argument minus that of the denominator:

$$\arg\left[\frac{\hat{i}_{out}(\omega)}{\hat{d}(\omega)}\right] = \arg[N(\omega)] - \arg[D(\omega)] = \tan^{-1}\left(-\frac{\omega}{\omega_{z2}}\right) - \tan^{-1}(\infty) \tag{2.129}$$

In dc, for $\omega = 0$, this expression becomes

$$\lim_{\omega \to 0} \arg\left[\frac{\hat{i}_{out}(\omega)}{\hat{d}(\omega)}\right] = \tan^{-1}(0) - \tan^{-1}(\infty) = -90° \tag{2.130}$$

The presence of the origin pole justifies this permanent phase lag of 90°. Now, with a normal zero, as the frequency increases, we would expect its contributing phase to reach 90°, canceling the origin pole action. Unfortunately, because this is a RHP zero, the total phase lag at $\omega = \infty$ becomes

$$\lim_{\omega \to \infty} \arg\left[\frac{\hat{i}_{out}(\omega)}{\hat{d}(\omega)}\right] = \tan^{-1}(-\infty) - \tan^{-1}(\infty) = -90° - 90° = -180° \tag{2.131}$$

This is the effect of the RHPZ: its phase lag is 90° compared to the 90° phase lead brought by a LHP zero. To obtain a complete picture, we can obtain the transfer function using the average model of Figure 2.24 but also using (2.124) and a calculation software such as Mathcad. Figure 2.27 shows the plots we have obtained. The superposition of both curves confirms the validity of the analytical equation we have derived. Also, as expected, the total phase lag reaches 180°. This is the effect of the origin pole and the RHP zero, which brings an additional –90° rather than 90° as it should for a normal zero. Please note that this RHPZ also exists, at the same location, in a current-mode boost converter.

In this example, should we need to stabilize the converter, the RHPZ position clearly bounds the maximum crossover frequency. To avoid any stability issue, it is recommended to limit the crossover frequency $f_c$ to less than 30 percent of the minimum RHPZ position. In our example, it would imply a crossover frequency of

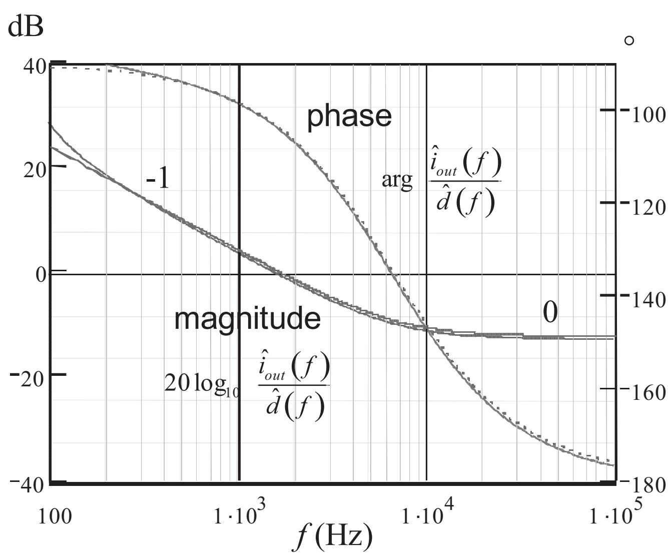



**Figure 2.27** The total phase lag reaches –180° whereas it should be 0° if the zero of the denominator would lie in the left-half portion of the *s*-plane.

$$f_c < 30\%\, f_{z_2} < 0.3 \times 6.6k < 2 \text{ kHz} \tag{2.132}$$

The compensation block $G$ must thus be tailored to force crossover below this value.

One final note on RHPZ: The previous example assumes CCM to illustrate the presence of the RHPZ. It is little known that a RHPZ can also exist in the discontinuous conduction mode (DCM) of operation. However, as it is located in the higher portion of the frequency spectrum, its influence is usually neglected at lower frequencies where crossover takes place.

## 2.6 Conclusion

Understanding transfer functions is key for the design of fast and stable closed-loop systems. Even if you will never do root locus calculations, it is important to realize that poles positions can move in relation to certain operating parameters. Once these parameters are identified (e.g., our output capacitor ESR), you will know how to efficiently compensate these variations over the power supply lifespan, ensuring a robust design. The RHP zero presence can sometimes hamper the available bandwidth in converters like boost or flyback architectures. Again, being able to analytically locate its worst-case position and visualize its effects on a Bode plot is an important point to let you safely pick a crossover frequency value. Finally, we have seen how fast analytical techniques can dramatically improve your analysis speed and unveil poles and zeros in a few minutes. It requires dexterity and practice but once you master the technique, going back to classical brute force algebra will be difficult!

## References


[1] Middlebrook, R. D., "Methods of Design-Oriented Analysis: Low-Entropy Expressions," New Approaches to Undergraduate Education IV, University of California, Santa Barbara, 1992.

[2] Middlebrook, R. D., V. Vorpérian, and J. Lindal, "The *N* Extra Element Theorem," *IEEE Transactions on Circuits and Systems, Fundamental Theory and Applications*, Vol. 45, No. 9, September 1998.

[3] Cochrun, B., and A. Grabel, "A Method for the Determination of the Transfer Function of Electronic Circuits," *IEEE Transactions on Circuit Theory*, Vol. 20, No. 1, January 1973.

[4] Erickson, R. W., "The *n* Extra Element Theorem," http://ecee.colorado.edu/copec/publications.php.

[5] Vorpérian, V., *Fast Analytical Techniques for Electrical and Electronic Circuits*, Cambridge: Cambridge University Press, 2002.

[6] DiStefano, J., A. Stubberud, and I. Williams, *Feedback and Control Systems*, New York: McGraw-Hill, 1990.

[7] Basso, C., *Switch Mode Power Supplies: SPICE Simulations and Practical Designs*, New York: McGraw-Hill, 2008.

[8] Basso, C., "Understanding the RHPZ," Parts I, II, III, and IV, *Power Electronics and Technology*, April, May, June, and July 2009.

## Appendix 2A: Determining a Bridge Input Impedance

We are going to use an example given by Dr. Vatché Vorpérian on page 12 of [1]. The circuit diagram appears in Figure 2.28:

The exercise consists of determining the impedance $Z_{in}(s)$ seen from the left side of Figure 2.28:

$$Z_{in}(s) = \frac{V_{in}(s)}{I_{in}(s)} \tag{2.133}$$

If Dr. Vorpérian used the extra element theorem (EET) to obtain the result, we are going to apply the same technique we used in this chapter to find the input impedance. There is one storage element, the capacitor $C$; this is thus a first-order system. Its general expression can be put under the following form. Analysis will further tell us if a pole or a zero are present:

$$Z_{in}(s) = R_0 \frac{1 + s/\omega_{z_1}}{1 + s/\omega_{p_1}} \tag{2.134}$$

First, let us derive the input resistance seen at dc, $R_0$: open-circuit the capacitors and short-circuit the inductors, if any. Our circuit simplifies to that of Figure 2.29.

The input resistance is simply $R_1$ in series with the series-parallel combination of the remaining elements:

$$R_0 = R_1 + (R_3 + R_4) \,||\, R_2 \tag{2.135}$$

Now, let's see if a zero exists. A zero in this circuit would prevent the excitation signal from reaching the output. As we deal with an impedance expression, the excitation signal is the input current $I_{in}$ and the response is the input voltage $V_{in}$. What in Figure 2.28 would nullify $V_{in}$? A short circuit involving the branch where $C$ lies.

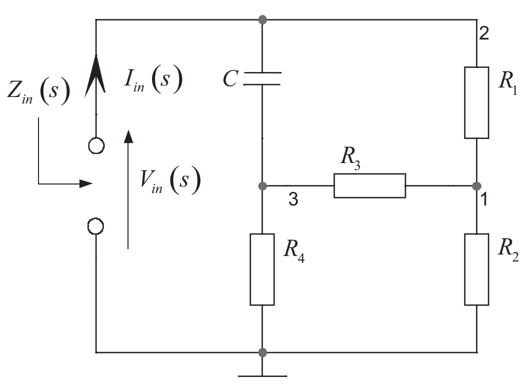



**Figure 2.28** The input impedance of this bridge can be derived in a few steps when using analytical techniques.

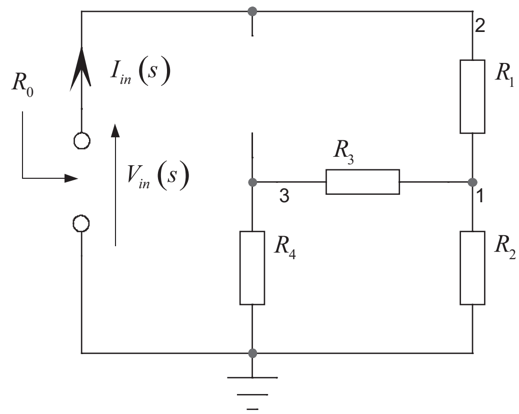



**Figure 2.29** The dc input resistance, when the capacitor is removed, is really easy to find.

If we have a short circuit, then $V_{in} = 0$. If $V_{in} = 0$, then node 2 is grounded and $R_1$ comes in parallel with $R_2$. The circuit appears in Figure 2.30.

The expression seen from the input impedance port is simply

$$Z_{in}(s) = \frac{1}{sC} + R_4 \,||\, (R_3 + R_2 \,||\, R_1) = \frac{1 + sCR_4 \,||\, (R_3 + R_2 \,||\, R_1)}{sC} \tag{2.136}$$

To cancel this expression, we set its numerator to zero and solve for the root. In our case, this is simply

$$1 + sCR_4 \,||\, (R_3 + R_2 \,||\, R_1) = 0 \tag{2.137}$$

Thus,

$$\omega_{z_1} = \frac{1}{sCR_4 \,||\, (R_3 + R_2 \,||\, R_1)} \tag{2.138}$$

Now that the zero has been found, let's give a look to the pole. To derive the denominator expression $D(s)$, we need to set the excitation to zero and find the time

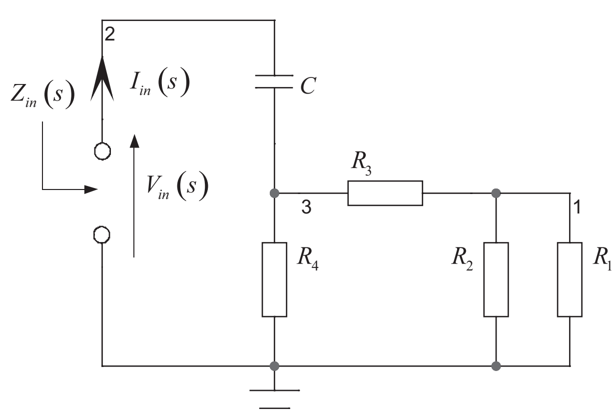



**Figure 2.30** The zero expression is found by nullifying the response signal, which is $V_{in}$.

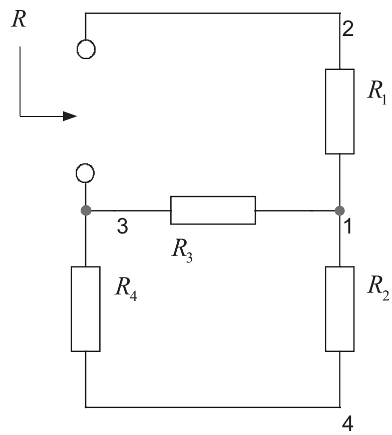



**Figure 2.31** The time constant is found by setting the excitation, $I_{in}$, to zero.

constant of the resulting network. The excitation, in our case, is the input current $I_{in}$. This is a current generator. When set to zero, it transforms into an open circuit, leading to the above updated schematic.

The next step is to find the resistor $R$ driving the capacitor. Again, looking at the open port in Figure 2.31, a simple resistive arrangement is obtained:

$$R = R_1 + R_3 \,||\, (R_4 + R_2) \tag{2.139}$$

The time constant is therefore

$$\tau = \left[R_1 + R_3 \,||\, (R_4 + R_2)\right]C \tag{2.140}$$

Leading to a pole expression of

$$\omega_{p_1} = \frac{1}{\tau} = \frac{1}{[R_1 + R_3 \,||\, (R_4 + R_2)]C} \tag{2.141}$$

Here we are: we have derived the input impedance expression in less than 10 steps! When gathering equations (2.135), (2.138), and (2.141), we have:

$$Z_{in}(s) = R_1 + (R_4 + R_2) \,||\, R_2 \frac{1 + sC[R_4 \,||\, (R_3 + R_2 \,||\, R_1)]}{1 + sC[R_1 + R_3 \,||\, (R_4 + R_2)]}$$

From this expression, it is easy to identify the dc term and the associated pole and zero.

## Reference

[1] Vorpérian, V., *Fast Analytical Techniques for Electrical and Electronic Circuits*, Cambridge: Cambridge University Press, 2002.

## Appendix 2B: Plotting Evans Loci with Mathcad

There are several ways to plot Evans roots loci on a computer. One of them uses the popular mathematical tool Mathcad. Let's assume we want to plot the roots loci of a second-order transfer function defined as

$$H(s) = \frac{1}{\left(\frac{s}{\omega_0}\right)^2 + \frac{s}{\omega_0 Q} + 1} \tag{2.142}$$

The denominator of this expression includes the poles of the transfer function. These poles are the roots of the equation

$$\left(\frac{s}{\omega_0}\right)^2 + \frac{s}{\omega_0 Q} + 1 = 0 \tag{2.143}$$

To use Mathcad for the resolution, let's open a new sheet and enter the following equations:

$$\omega_0 := 20008 \quad Q := 0.1 \quad Y(s,Q) := \frac{1}{\left(\frac{s}{\omega_0}\right)^2 + \frac{s}{\omega_0 \cdot Q} + 1}$$

The $Q$ and the resonant frequency $\omega_0$ are arbitrarily selected for this example. As the expression of the denominator follows the form of a second-order polynomial $f(s) = as^2 + bs + c$, we ask the software to identify the individual coefficients by using the function *denom* where $Q$ is the variable. Should we need to also plot the zeros of the transfer function, we would replace *denom* by the keyword *numer* in the following expression:

$$\mathrm{b(Q){:=}(denom(Y(s,Q)))\ coeffs,s} \rightarrow \begin{pmatrix} 400320064\mathrm{Q} \\ 20008 \\ \mathrm{Q} \end{pmatrix} \begin{matrix} \rightarrow c \\ \rightarrow b \\ \rightarrow a \end{matrix}$$

According to this result, the denominator function to solve is in the form of

$$f(s) = Qs^2 + 20008s + 400320064Q = 0 \tag{2.144}$$

Mathcad® can solve this function via the keyword *polyroots* whose results are passed to a two-dimensional vector, X:

$$X(Q) := polyroots(b(Q))$$

We can extract the real and imaginary parts via two dedicated keywords, *Im* and *Re*:

$$\mathrm{Im(X(Q))} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \qquad \mathrm{Re(X(Q))} = \begin{pmatrix} -1.981 \times 10^5 \\ -2.021 \times 10^3 \end{pmatrix}$$

These two-dimensional vectors include the conjugate roots $s_1$, $s_2$ and will be accessed via subscripted notations. Please note that the subscripted notation for a

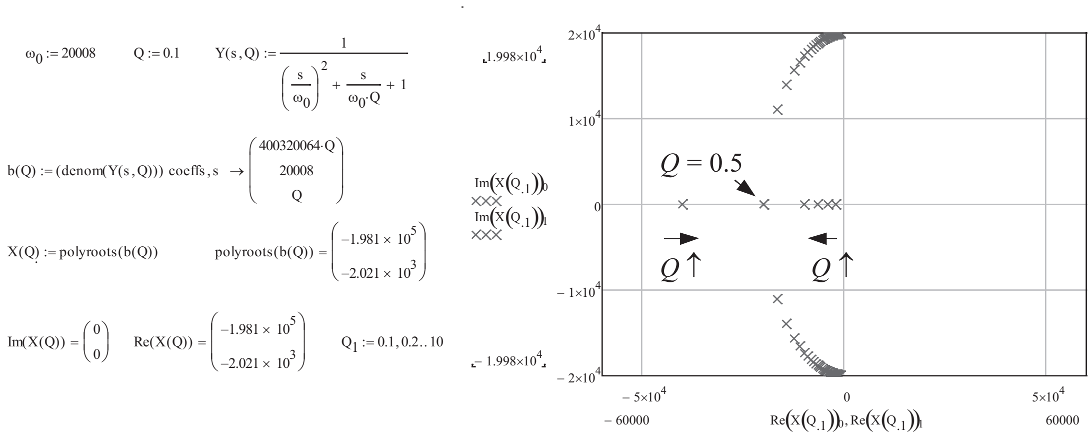



**Figure 2.32** This figure gathers all notations and the plot showing how the roots move as a function of *Q*.

vector manipulation in Mathcad is obtained by pressing "[" right after the vector name:

| | $S_1$ | $S_2$ |
|---|---|---|
| y | $\mathrm{Im}(X(Q))_0$ | $\mathrm{Im}(X(Q))_1$ |
| x | $\mathrm{Re}(X(Q))_0$ | $\mathrm{Re}(X(Q))_1$ |

We now have everything to vary $Q$ from 0.1 to 10 and plot the points individually made of $\mathrm{Re}(X(Q))_{0 \text{ or } 1}$ as the vertical coordinate and $\mathrm{Im}(X(Q))_{0 \text{ or } 1}$ as the horizontal coordinate. The complete calculation sheet including the roots loci graph appears in Figure 2.32.

To display × for the poles, right click on the graph and select traces/symbols. Pick a cross for both variables and you are done. Should it be zeros instead, pick the ○ as the available symbol, affected by the color of your choice.

We can see the poles moving in the same direction as $Q$ increases. When $Q$ equals 0.5, both roots are real and coincident. As $Q$ continues to grow, the poles split, giving birth to an imaginary portion. As $Q$ continues to increase, the real portion (the damping) vanishes and the roots eventually reach the imaginary axis when $Q$ reaches infinity.

There are plenty of solutions to plot roots loci in Mathcad. The presented solution is simple and quick to implement. As a drawback, you will find difficult to associate a given set of roots in the plot with the corresponding $Q$ value. You can always use the available cursor, but it is not really practical. Searching on the Web will give more comprehensive solutions where $Q$ is displayed when a root is identified in the plot but at the expense of an increased sheet complexity.

## Appendix 2C: Heaviside Expansion Formulas

We have seen how to obtain the time-domain response of a system whose transfer function is known when its input is subjected to a step function. You have to decompose the quotient into partial fractions and then sum up the inverse Laplace expression of each individual fraction. Despite its simplicity, the process requires two steps: decompose the quotient into partial fractions, and then find the inverse Laplace function of each partial fraction. Heaviside proposed another method that can give you the time-domain response in one single step. It is not really well known, and I could not find a lot of information on the Web. This is the Heaviside development formula described in [1]. It is expressed the following way:

$$\mathcal{L}^{-1}\left\{\frac{N(s)}{D(s)}\right\} = \sum_{k=1}^{n} \frac{N(s_k)}{D'(s_k)} e^{s_k t} \tag{2.145}$$

The exercise consists of identifying the denominator roots $s_1$, $s_2$ to $s_n$ if you have $n$ roots. Then, you have to take the derivative of the denominator. Once you have these elements on hand, simply apply the following formula and the resulting time-domain equation shows up:

$$\mathcal{L}^{-1}\left\{\frac{N(s)}{D(s)}\right\} = \frac{N(s_1)}{D'(s_1)} e^{s_1 t} + \frac{N(s_2)}{D'(s_2)} e^{s_2 t} + \ldots + \frac{N(s_n)}{D'(s_n)} e^{s_n t} \tag{2.146}$$

Let's take the transfer function already given in (2.60). It represents the system transfer function multiplied by the step function, $1/s$:

$$\frac{N(s)}{D(s)} = \frac{s + 5k}{s(s + 1k)(s + 30k)} \tag{2.147}$$

We have three roots for $D(s)$. These roots are

$$\begin{aligned} s_1 &= 0 \\ s_2 &= -1k \\ s_3 &= -30k \end{aligned} \tag{2.148}$$

We now develop the denominator:

$$D(s) = s^3 + 31000 \cdot s^2 + 30000000s \tag{2.149}$$

and derive it:

$$\frac{dD(s)}{ds} = 3s^2 + 62000s + 30000000 \tag{2.150}$$

We now calculate

$$\begin{aligned} N(0) &= 5k \\ N(-1k) &= 4k \\ N(-30k) &= -25k \end{aligned} \tag{2.151}$$

and

$$\begin{aligned} D'(0) &= 3\times 10^7 \\ D'(-1k) &= -2.9\times 10^7 \\ D'(-30k) &= 8.7\times 10^8 \end{aligned} \tag{2.152}$$

By applying (2.146), we have our time-domain response immediately:

$$\mathcal{L}^{-1}\left\{\frac{N(s)}{D(s)}\right\} = \frac{N(0)}{D'(0)}e^{0t} + \frac{N(-1k)}{D'(-1k)}e^{-1000t} + \frac{N(-30k)}{D'(-30k)}e^{-30000t} \tag{2.153}$$

Developing, we obtain

$$\mathcal{L}^{-1}\left\{\frac{N(s)}{D(s)}\right\} = 166u - 138u\cdot e^{-1000t} - 28.7u\cdot e^{-30000t} \tag{2.154}$$

This is exactly what we found in (2.70) without going through the pain of partial decomposition. Simple, yet elegant, isn't it?

Ok, let's go through another small example. Assume the following transfer function:

$$H(s) = \frac{1+s}{s^2+4} = \frac{N(s)}{D(s)} \tag{2.155}$$

We have two roots for $D(s)$. These roots are

$$\begin{aligned} s_1 &= -2j \\ s_2 &= 2j \end{aligned} \tag{2.156}$$

Let's derive $D(s)$:

$$\frac{dD(s)}{ds} = 2s \tag{2.157}$$

We now calculate

$$\begin{aligned} N(-2j) &= 1-2j \\ N(2j) &= 1+2j \end{aligned} \tag{2.158}$$

and

$$\begin{aligned} D'(-2j) &= -4j \\ D'(2j) &= 4j \end{aligned} \tag{2.159}$$

By applying (2.146), we have our time-domain response almost immediately:

$$\mathcal{L}^{-1}\left\{\frac{1+s}{s^2+4}\right\} = -\frac{1-2j}{4j}e^{-2jt} + \frac{1-2j}{4j}e^{2jt} \tag{2.160}$$

Rearranging and factoring, we have

$$\mathcal{L}^{-1}\left\{\frac{1+s}{s^2+4}\right\} = \frac{1}{2}\left(\frac{e^{2jt}-e^{-2jt}}{2j}\right)+\left(\frac{e^{2jt}+e^{-2jt}}{2}\right) \tag{2.161}$$

We can identify complex sine and cosine functions. The final expression is

$$\mathcal{L}^{-1}\left\{\frac{1+s}{s^2+4}\right\} = \frac{1}{2}\sin(2t)+\cos(2t) \tag{2.162}$$

## Reference

[1] Spiegel, M. R., *Schaum's Outline of Laplace Transforms*, New York: McGraw-Hill, 1965.

## Appendix 2D: Plotting a Right Half Plane Zero with SPICE

We have seen that a RHPZ is a numerator root featuring a positive real part. A RHPZ can be put under the following form:

$$H(s) = 1 - s/\omega_{z_1} \tag{2.163}$$

The negative sign indicates the presence of a positive root, located in the right-half portion in the *s*-plane. To show the effect of a RHPZ in a Bode plot, we can try to artificially create one with SPICE. A possible circuit appears in Figure 2.33. The transfer function of such a configuration is quickly obtained:

$$V_{out}(s) = V_{in}(s) - V_{in}(s)\frac{R_1}{\frac{1}{sC_1}} = V_{in}(s)(1 - s/\omega_{z_1}) \tag{2.164}$$

where $\omega_{z_1} = \dfrac{1}{R_1C_1}$.

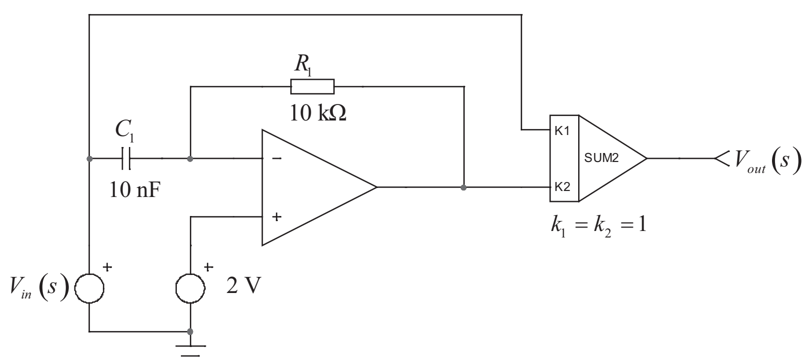



**Figure 2.33** A RHPZ is artificially created with an op amp-based differentiator and an adder.

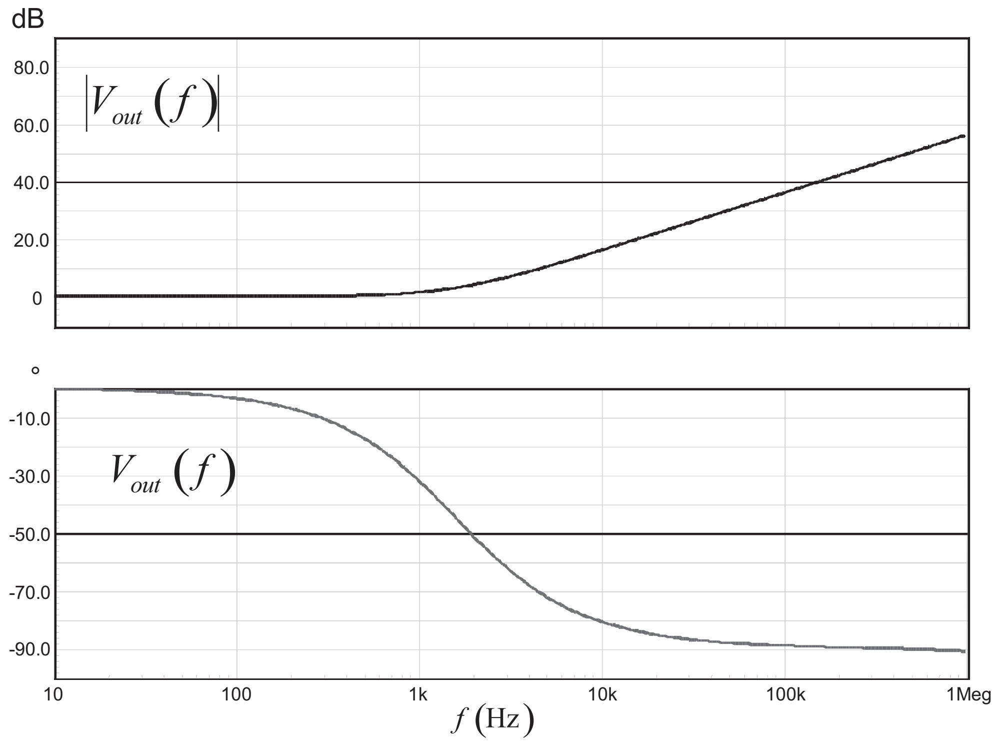



**Figure 2.34** As expected, the simulation results reveal a magnitude similar to that of classical zero, but the phase, rather than going up, is actually going down to –90°.

The ac plot of such a transfer function is given in Figure 2.34. As expected, the magnitude is actually that of normal zero: it increases as the frequency increases. But the phase, rather than also heading towards +90°, it actually lags by 90°, as a pole would do.

CHAPTER 3

# Stability Criteria of a Control System

In previous chapters, we have learned that a closed-loop system works by permanently comparing its output variable with the setpoint imposed by the control input. The difference between both variables gives birth to an error signal, $\varepsilon$, further processed by the compensator block. The resulting control signal $V_c$ drives the transmission chain to force both input and output variables to match. To operate properly, the control signal must *oppose* the output variations: for instance, if the output signal increases beyond the target, as it subtracts from the control input, it naturally instructs the system to reduce its output, bringing it back within the accepted range. If for any reason the control signal no longer opposes the output but actually amplifies it, the system becomes unstable and goes out of control, with all possible consequences. Understanding stability is key to designing robust and reliable control systems. This third chapter focuses on this subject, introducing parameters such as *phase margin and crossover frequency* but also less known stability criteria such as *modulus* and *delay* margins.

## 3.1 Building An Oscillator

In the electronic field, an oscillator is a circuit capable of producing a self-sustained sinusoidal signal. In a lot of configurations, cranking up the oscillator involves the noise level inherent to the adopted electronic circuit. As the noise level grows at power-up, oscillations are started and self-sustained. This kind of circuit can be formed by assembling blocks such as those appearing in Figure 3.1. As you can see, the configuration looks very similar to that of our control system arrangement.

In our example, the excitation input is not the noise but a voltage level, $V_{in}$, injected as the input variable to crank the oscillator. The direct path is made of the transfer function $H(s)$ while the return path consists of the block $G(s)$. To analyze the system, let us write its transfer function by expressing the output voltage versus the input variable:

$$V_{out}(s) = \varepsilon(s)H(s) = [V_{in}(s) - G(s)V_{out}(s)]H(s) \tag{3.1}$$

If we expand this formula and factor $V_{out}(s)$, we have

$$V_{out}(s)[1 + G(s)H(s)] = V_{in}(s)H(s) \tag{3.2}$$

The transfer function of such a system is therefore

$$\frac{V_{out}(s)}{V_{in}(s)} = \frac{H(s)}{1 + G(s)H(s)} \tag{3.3}$$

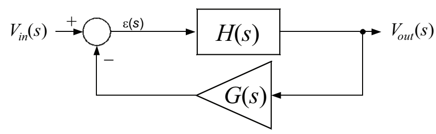



**Figure 3.1** An oscillator is actually a control system where the error signal does not oppose the output signal variations.

In this expression, the product $G(s)H(s)$ is called the *loop gain*, also noted $T(s)$. To transform our system into a self-sustained oscillator, an output signal must exist even if the input signal has disappeared. To satisfy such a goal, the following condition has to be met:

$$\lim_{V_{in}(s)\to 0}\left[\frac{H(s)}{1+G(s)H(s)}V_{in}(s)\right]\neq 0 \tag{3.4}$$

To verify this equation in which $V_{in}$ disappears, the quotient must go to infinity. The condition for this quotient to go infinite is that its *characteristic* equation, the denominator $D(s)$, equals zero:

$$1+G(s)H(s)=0 \tag{3.5}$$

To meet this condition, the term $G(s)H(s)$ must equal −1. Otherwise stated, the magnitude of the loop gain must be 1 and its sign should change to minus. A sign change with a sinusoidal signal is simply a 180° phase reversal. These two conditions can be mathematically noted as follows:

$$\left|G(s)H(s)\right|=1 \tag{3.6}$$

$$\arg G(s)H(s)=-180^\circ \tag{3.7}$$

When these two expressions are exactly satisfied, we have conditions for steady-state oscillations. This is the so-called Barkhausen criterion, expressed in 1921 by the eponymous German physicist. Practically speaking, in a control loop system, it means that the correction signal no longer opposes the output but returns *in phase* with the exact same amplitude as the excitation signal. In a Bode plot, (3.6) and (3.7) would imply a loop gain curve crossing the 0-dB axis and affected by a 180° phase lag right at this point. In a Nyquist analysis, where the imaginary and real portions of the loop gain are plotted versus frequency, this point corresponds to the coordinates $-1, j0$. Figure 3.2 displays these two curves where conditions for oscillations are met. Should the system slightly deviate from these values (e.g., temperature drift, gain change), output oscillations would either exponentially decrease to zero or diverge in amplitude until the upper/lower power supply rail is reached. In an oscillator, the designer strives to reduce as much as possible the gain margin so that conditions for oscillations are satisfied for a wide range of operating conditions.

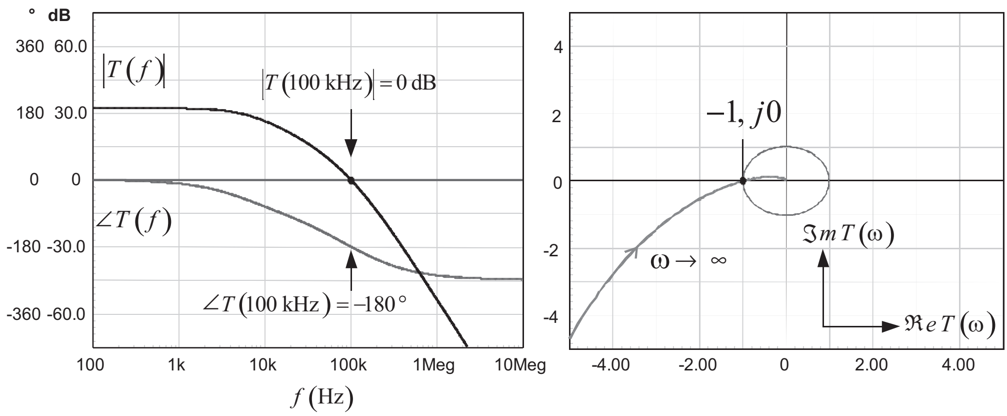



**Figure 3.2** Conditions for oscillations can be illustrated either in a Bode diagram or in a Nyquist plot.

### 3.1.1 Theory at Work

To understand how oscillations are formed, a simulation circuit using SPICE can help us. Figure 3.3 shows our oscillator structure, based on three cascaded *RC* networks. The ac response of such a third-order configuration appears in Figure 3.4. It shows a flat 0-dB magnitude up to 1 kHz with a slope of –60 dB per decade beyond. The phase starts to drop around 1 kHz and reaches –180° at a 39-kHz frequency. At this point, the magnitude of the *RC* network transfer function is –29.32 dB. Our experiment will consist of compensating this attenuation via a compensator block $G(s)$ made of a simple gain, independent of frequency. If we cascade a block $H$

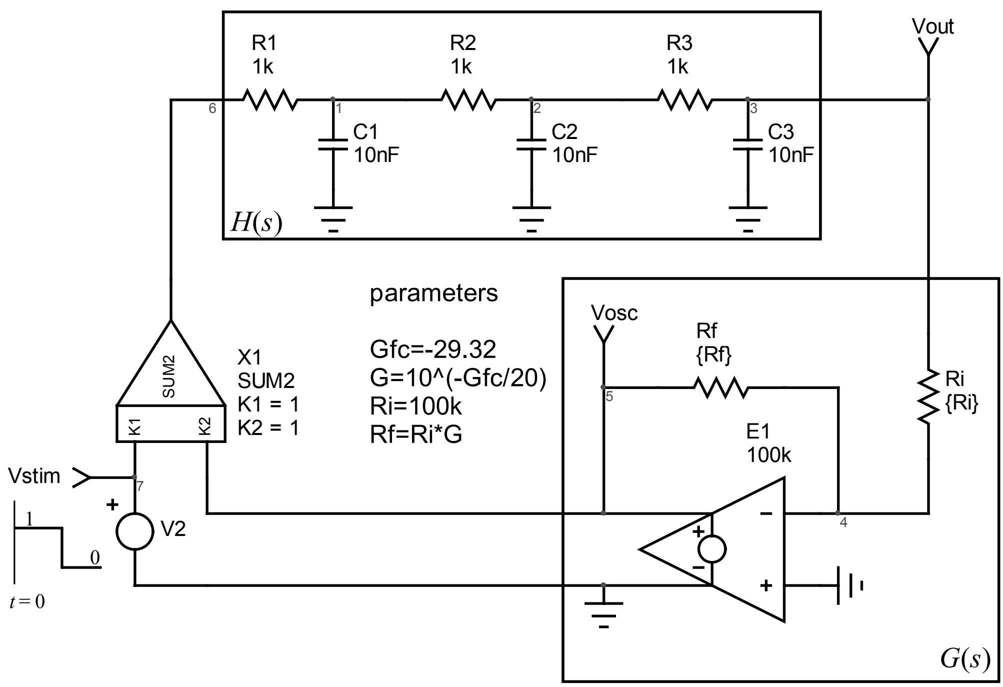



**Figure 3.3** We have formed a *RC* network whose cumulated phase lag goes beyond 180°. An operational amplifier compensates the attenuation to make the total loop gain approaching 1.

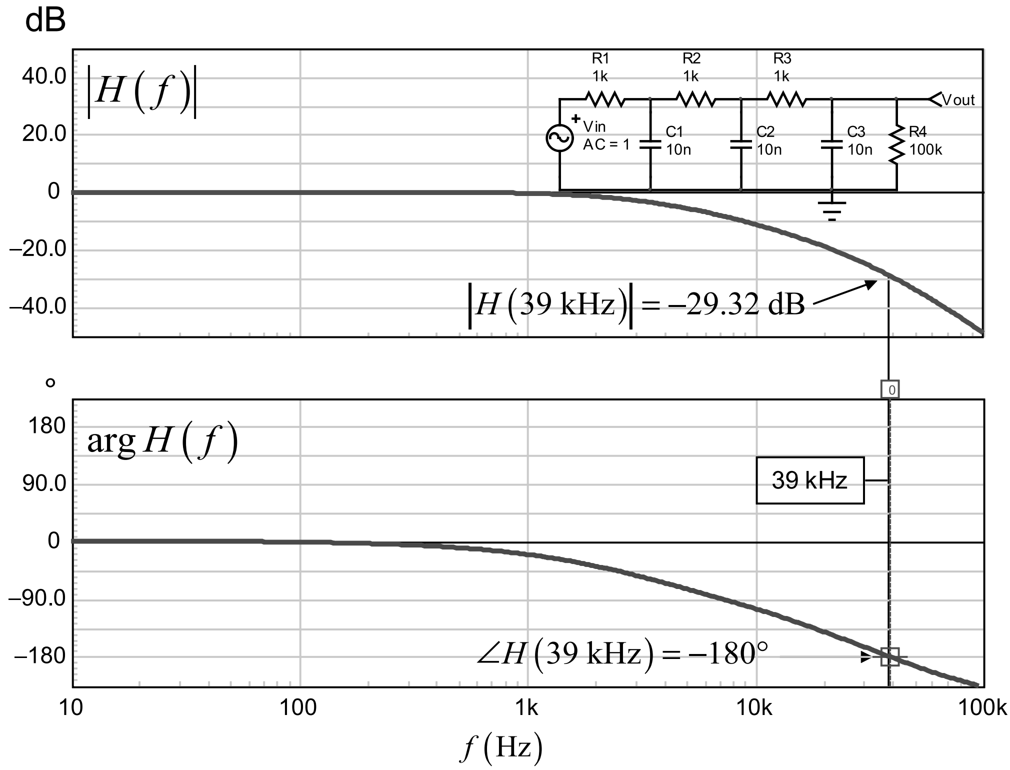



**Figure 3.4** The ac response of this network shows a phase lag of 180° at a frequency close to 39 kHz. At this point, the attenuation of the network is 29.32 dB.

featuring an attenuation of –29.32 dB at 39 kHz with another block $G$ showing a permanent gain of 29.32 dB, the resulting loop gain magnitude at 39 kHz will thus be 0 dB. Our conditions for oscillation can therefore be met where $|T(39\text{ kHz})| = 1$ and $\arg T(39\text{ kHz}) = -180°$. The compensation scheme uses a perfect operational amplifier (op amp) wired as an inverter and adjusted in gain by playing with resistor $R_f$. On the left side of the picture appears our adder subcircuit $X_2$ used to close the loop and inject the cranking voltage, $V_1$. This source equals 1 V for a short period of time and immediately goes to 0, obeying what (3.4) describes.

The simulation will then consist of adjusting the compensating gain provided by the op amp to create different operating conditions for the loop gain $T(s)$. The ac responses for three selected cases are plotted in Figure 3.5 at the 0-dB reference point; the phase lag is less than 180° ($G = 27$ dB), exactly 180° ($G = 29.32$ dB), or beyond 180° ($G = 33$ dB).

The simulated transient data covering each compensation scenario appear in Figure 3.6 and reveal different types of responses.

- In the upper section of the picture, the gain was adjusted to 27 dB. In this condition, the magnitude curve crosses the 0-dB axis at a 33.85-kHz frequency where the total phase lag is 171°. The conditions for oscillations are not met. The transient response is a damped oscillatory signal that decays to zero in an exponential envelope.

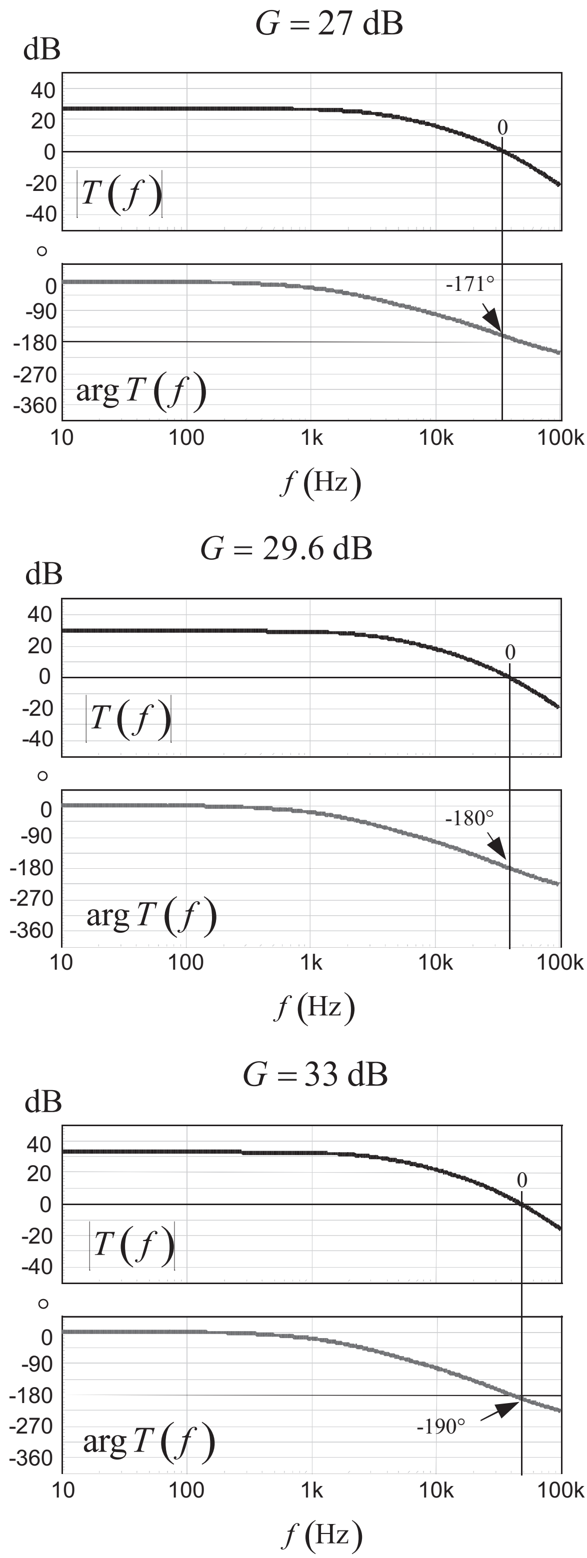



**Figure 3.5** Three different loop gain and phase scenarios are used to look at the corresponding transient responses.

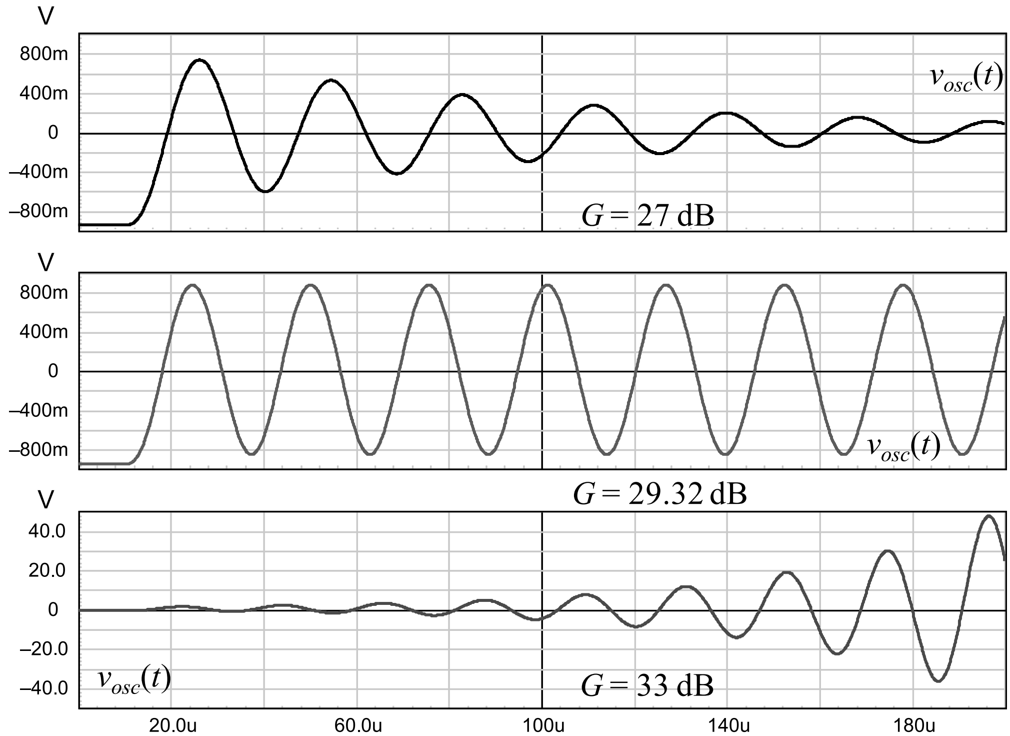



**Figure 3.6** Simulation results show a response that is decaying, self-sustained, or diverging, depending on the points at which crossover occurs.

- When the gain is exactly set to 29.32 dB, the 0-dB crossing point is reached at 39 kHz and the phase lag is precisely –180°. The oscillation criteria are now respected, and we have a nice self-sustained sinusoidal waveform at a 39-kHz frequency.
- In the last plot, the op amp gain is pushed to 33 dB, forcing the loop gain to cross over where the phase lag is –190°. The output diverges and exponentially grows until the op amp hits its upper limit voltage range or smoke pops up, preceded by a loud noise.

From these experiments, we can see that oscillations are obtained and self-sustained *only* when the Barkhausen criteria are met. Deviating from this point results in an oscillatory response either decaying or growing.

## 3.2 Stability Criteria

You understand that our goal with a control system is not to build an oscillator. We want a control system featuring speed, precision, and an oscillation-free response. We must therefore keep away from a configuration where conditions for oscillation or divergence are met. One way is to limit the frequency range within which our system will react. By definition, the frequency range, or the *bandwidth*, corresponds to a frequency value where the closed-loop transmission path from the input to the output drops by 3 dB. The bandwidth of a closed-loop system can be seen as a frequency range where the system is said to satisfactorily respond to its input (i.e., follows the setpoint or efficiently rejects the perturbations). As we will later see, during the design stage, we do not directly control the closed-loop bandwidth but the *crossover frequency* $f_c$, a parameter pertinent to an open-loop analysis. Both variables are often approximated as equal, and we will see that it is true in one condition only. However, they are not far away from each other, and both terms can be interchanged in the discussion.

We have seen that the open-loop gain represents an important parameter for our control system. When gain exists (i.e., $|T(s)| > 1$), the system works in dynamic closed-loop conditions and can compensate for incoming perturbations or react to setpoint changes. However, there is a limit to the system reaction: the system must offer gain at the frequencies involved in the perturbing signal. If the perturbations or the setpoint changes are too fast, the frequency content of the excitation signal is beyond the bandwidth of the system, implying the absence of gain at these frequencies: the system becomes slow and cannot react, operating as if the loop were unresponsive to varying waveforms. Is an infinite-bandwidth system desirable then? No, because increasing the bandwidth is like widening the diameter of a funnel: you are certainly going to collect more information and react faster to incoming perturbations, but the system will also accept spurious signals such as noise and parasitic data, self-produced by the converter in some cases (the output ripple in switching power supplies, for instance). For this reason, it is mandatory to limit the bandwidth to what your application really requires. Adopting too wide a bandwidth would be detrimental to the *noise immunity* of the system (e.g., its robustness to external parasitic signals).

How do we limit the bandwidth of a control system? By shaping the loop gain curve through the compensator block, $G$. This block will make sure that after a certain frequency $f_c$, the loop gain magnitude $|T(f_c)|$ drops and passes below 1 or 0 dB. As we explained, it is roughly the bandwidth of your control system once the loop is closed. The frequency at which this phenomenon occurs is called the *crossover frequency* noted $f_c$. Is this enough to obtain a robust system? No, we need to ensure another important parameter: the phase of $T(s)$ at the point where its magnitude is 1 must be less than –180°. From our experiments, we have seen that when the loop phase is less than –180° at the crossover frequency, we obtain a response converging toward a stable state. This is obviously a highly desirable characteristic of our control system. To make sure we stay away from the –180° limit at crossover, the compensator $G(s)$ must tailor the loop argument at the selected crossover frequency to build *phase margin*, PM or $\varphi_m$. The phase margin can be considered a design or a safety limit ensuring that despite external perturbations or unavoidable production spreads, changes in the loop gain will not put the stability in jeopardy. As we will later see, the phase margin also impacts the transient response of the system. Therefore, its choice does not exclusively depend on stability considerations but also on the type of transient response you want. Mathematically, the phase margin is defined as follows:

$$\varphi_m = 180 + \angle T(f_c) \text{ in degrees} \tag{3.8}$$

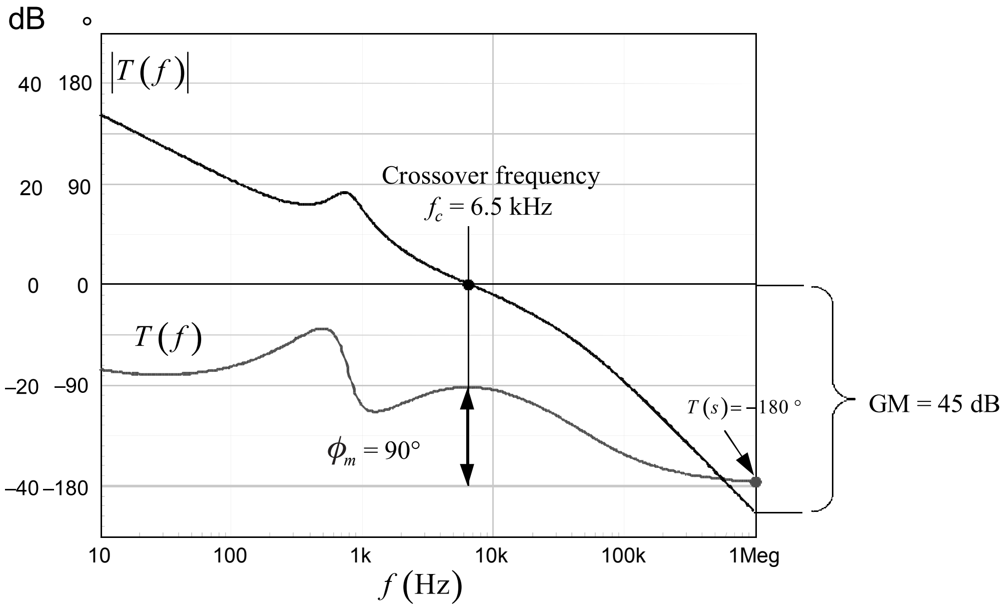



**Figure 3.7** In this example, the 0-dB crossover point is located at 6.5 kHz, where the total phase lag offers a phase margin of 90°.

where $T$ represents the open-loop gain made of the cascaded plant $H$ and compensator $G$ gains.

A typical compensated loop gain curve appears in Figure 3.7 and shows a crossover frequency of 6.5 kHz. At this point, the phase of $T(s)$ is –90°. If you start from the –180° line at a 6.5-kHz frequency and positively count the degrees until you cross the argument waveform, you have the phase margin: 90° in this example. This is an extremely robust system that is told to be *unconditionally stable*: despite moderate loop gain variations around the crossover point, there are no possibilities to cross over at a frequency where the phase margin is too small. By too small, we mean a phase margin approaching 30°, a limit below which the system gives an unacceptable ringing response. This is the reason why you have learned at school that 45° was the limit, giving an extra margin compared to 30°. We will later see that there is an analytical origin for these numbers.

### 3.2.1 Gain Margin and Conditional Stability

Figure 3.8 shows another typical frequency response of a compensated converter, highlighting the 0-dB crossover point as well as the phase margin. We know by experience that the elements constitutive of the converter will exhibit variations along the product life cycle. These variations can be linked to natural production spreads (for instance, resistors or capacitors affected by a lot-to-lot tolerance). The converter environmental operating conditions also have an impact on components. Among these variables, temperature plays an important role and affects passive or active component parameters. It can be capacitors or inductors equivalent series resistors (ESR), the optocoupler current transfer ratio (CTR), or the beta of bipolar transistors for instance. These variations impact the loop gain by shifting it up or down depending on the affected parameters. If the gain curve undergoes a shift, the 0-dB crossover frequency will transition to a new value imposing a different bandwidth to the converter. How can the converter stability be affected under these changes? Well, if the new crossover takes place at a point where the phase margin is weak, you may degrade the transient response so that the overshoot is no longer acceptable. It is thus your responsibility, as a designer, to ensure that these dispersions do not suddenly increase the gain at a frequency where you approach the –180° limit. You need sufficient gain margin as defined by

$$GM = \frac{1}{|T(f_\pi)|} \tag{3.9}$$

where $f_\pi$ corresponds to the frequency point where $\angle T(s)$ is exactly –180° or $-\pi$ radians (1 MHz in Figure 3.7).

Figure 3.8 portrays typical gain variations of ±10 dB due to production spreads in the selected components. It brings the crossover frequency from 1.5 kHz to 30 kHz. In this area, the phase margin changes from 70° to 45°, safe numbers according to theory. What is the worst case? It is when the new crossover frequency occurs where the total phase lag is 180°, matching the conditions for oscillations. This condition occurs at 1 MHz, implying a positive gain change of 35 dB.

Fortunately, deviations of 35 dB are unlikely to happen in modern electronics circuits. Time ago, when amplifiers or servomechanisms were driven by vacuum tubes-based circuits, warm-up times during the power-on sequence could induce large loop gain variations. Gain provisions were thus necessary to reject a second point where the stability could be in danger. This gain margin, identified on the loop gain curve at the frequency where the total phase lag reaches –180°, is noted GM in Figure 3.7. In modern electronic circuits, gain margins beyond 10 dB are

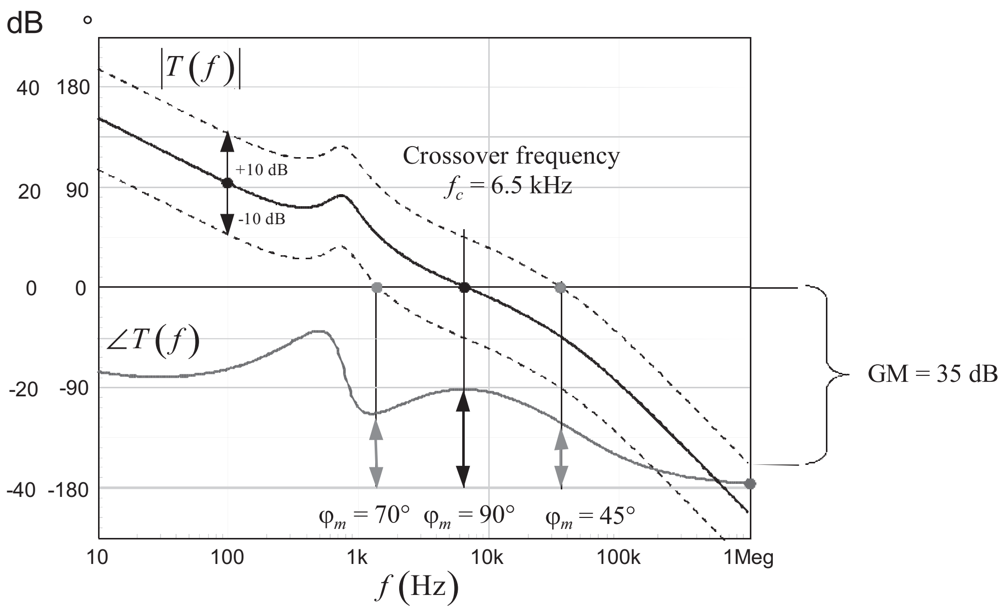



**Figure 3.8** The loop gain can show sensitivity to external parameters such as temperature. When a variation occurs, the phase margin must always stay within safe limits.

usually enough, unless your loop gain exhibits extreme sensitivity to an external parameter.

Another example of gain shift appears in Figure 3.9. It shows another compensated converter exhibiting a phase margin of 80° at 10 kHz. Based on what we discussed, we know that gain changes can occur, inducing ups or downs on the gain curve. In our example, we can identify an area around 2 kHz where the phase margin is as small as 18°. If a gain decrease of 20–25 dB occurs, you can end up with a control system showing a dangerously low phase margin around 2 kHz. It would lead to an oscillatory response, perhaps exceeding the overshoot specifications. This kind of system is told to be *conditionally stable*. Fortunately, as already said, a 25-dB variation of gain is unusual and such a system can be considered robust with this gain margin. However, I have seen design cases where the end user (your customer) clearly stated in the specifications that conditional designs were not acceptable, asking for a phase margin greater than 60° at all points below the crossover frequency. In this case, it becomes mandatory to compensate the converter so that no region of reduced phase margins below crossover ever exist whatever the operating conditions are.

It is often believed that a system where the phase curve dips below –180° before crossover is an unstable system. Such a response appears in Figure 3.10. The phase curve quickly drops after 1 kHz and passes the –180° limit at 1.5 kHz for a few kilohertz.

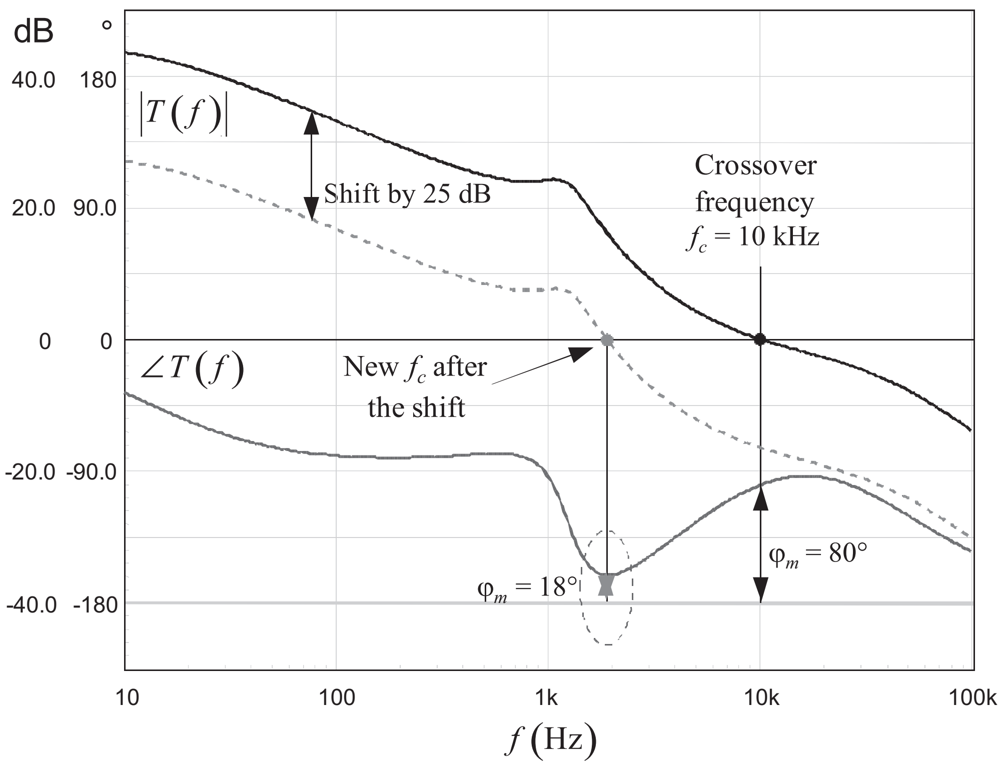



**Figure 3.9** In this example, if the gain shifts down by 25 dB, the curve crosses the 0-dB axis at a point where the phase margin is only 18°. Such a low phase margin will give a very oscillatory response, affected by a large overshoot. This is a case for conditional stability.

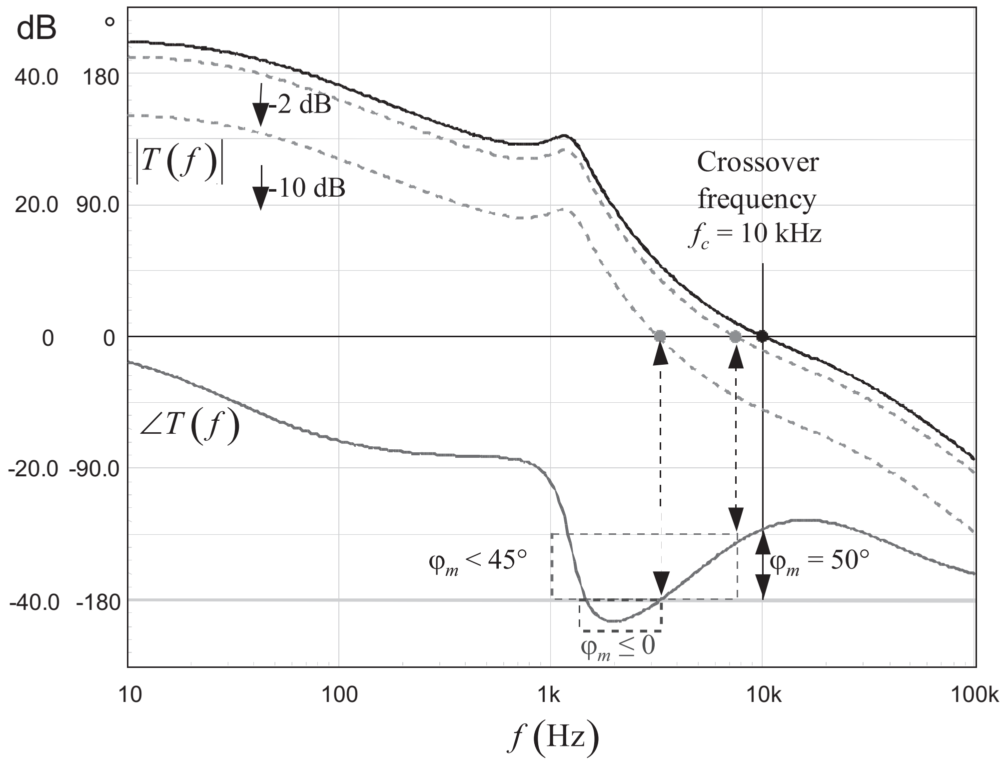



**Figure 3.10** The phase lags beyond 180° but in an area where the gain is larger than 1. This is not a problem, and the response is acceptable.

It then goes up again to offer a phase margin of 50° at 10 kHz. Yes, this system is stable simply because we do not satisfy (3.7) at 0 dB. Remember, to cancel the denominator of (3.3), you must have the gain magnitude exactly equal to 1 *and* a phase lag of 180° or beyond. In the graph, we can see that this condition is not satisfied at any point in the picture. However, it is worth noting that the loop is highly conditional. Should the gain reduce by a few decibels and your phase margin becomes less than 45°. Another 10-dB decrease and you enter a dangerous area of zero phase margin where, this time, the oscillation criteria would be met.

### 3.2.2 Minimum Versus Nonminimum-Phase Functions

In this book, we will essentially use Bode plots for stability analysis. Due to simplicity, the technique is widely adopted in the control engineering industry and, in particular, power conversion. However, the reader must be aware that Bode plots interpretations can sometimes mislead the designer when determining closed-loop stability. Caution must be taken when pure delays or right half plane poles or zeros appear in the transfer function. Here is why: intuitively, if you see a +1 slope in a Bode plot, a zero is in action and you expect the phase to increase and reach +90° at some frequency point. The same reasoning applies to a pole where its –1 slope should bring the phase to –90° at a given frequency. If you combine a pole and a zero, you would see the phase going toward –90° and returning to 0 when the zero starts to kick in. Mathematically, you can write that a minimum phase function has a high-frequency phase asymptote defined as

$$\varphi_{\max} = -90°(n - m) \tag{3.10}$$

where $n$ is the number of poles and $m$ the number of zeros: five poles and four zeros give an asymptote of –90°.

Figure 3.11 shows you two examples. Both functions share similar poles and zeros positions, except that the left curve combines a LHP zero and the right curve implements a RHP zero. If you look at the phase for the left-side function while masking the magnitude graph and knowing that a –90° phase corresponds to a –1 slope and 0° to a 0-slope (flat curve), you could easily reconstruct an asymptotic magnitude graph. On the right side, if you follow the same path, you see a phase curve starting from –90° implying a –1 slope and then touching –180°, meaning that a second –1 slope combines with the first one to become a –2 slope: it can be the result of two cascaded poles, but this is obviously wrong because of the RHP zero presence.

Mathematically, Hendrik Bode demonstrated that it is always possible to link the imaginary and real portions of a transfer function when its poles or zeros sit in the left-half plane: in this case, this transfer function is said to be of *minimum phase*. Practically, it means that you should be able to reconstruct a phase plot from a magnitude plot and vice versa as we shown via the Figure 3.11 left curve. After all, both phase and magnitude are a combination of the real and imaginary parts of the transfer function under study, and a path to reach either one must exist through the computed magnitude or phase values. Now, if RHP zeros or RHP poles appear in the transfer function, they distort the phase information as we showed. The Bode law is violated: this is a *nonminimum phase* function. The remark also applies if a pure delay is part of the transfer function. A delay does not change the magnitude of the function in which it appears but it makes its phase further lag as

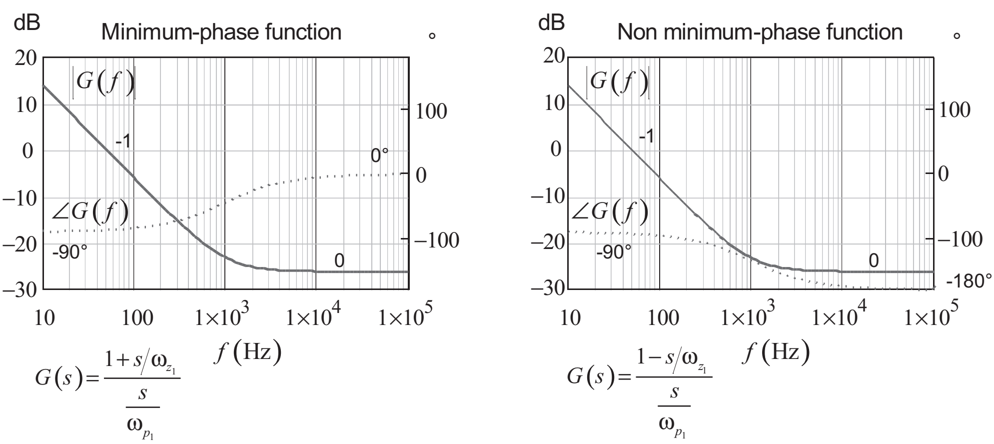



**Figure 3.11** With a minimum phase function, you should be able to reconstruct the phase from the magnitude plot and vice versa. On the left side, a –1 slope implies a phase lag of 90° and the 0 slope tells you that the phase has returned to 0°. On the right side, despite similar magnitude plots, the phase curve no longer fits the slopes values: this is a nonminimum-phase function.

the frequency increases, again distorting the phase response and violating the Bode law. In a many theory books, it is clearly stated that you should use Bode plots with caution with the presence of nonminimum-phase functions. In these specific cases, a Nyquist plot must be used, as it does not combine the real and imaginary parts to get magnitude and phase values but individually plots them on a dedicated chart. Even if RHP zeros or poles (also called *unstable* poles or zeros) are present, it does predict the stability or instability without ambiguity.

### 3.2.3 Nyquist Plots

Though Bode plots are widely used among the design community, it is important to understand Nyquist plots. As we just explained, in some cases, you must apply them to get the right answer, as Bode will fail to predict the instability. This paragraph is a very brief and necessarily incomplete introduction to Nyquist plots, but it should give you the basis in case you need to analyze a linear or switching regulator with this technique. The reader is encouraged to search the available literature and the Web to further learn the vast field of Nyquist plots applications.

A Nyquist graph is a plot where you mark the loci of points in which $x$,$y$ coordinates are respectively the real and imaginary parts of the function under study as the angular frequency increases from 0 to infinity. Linking all these marks together gives birth to the Nyquist plot. A typical example appears in Figure 3.12. The curve starts from the right where the angular frequency is low (this is your starting frequency in the ac sweep, dc or 0 in theory). Then, as the angular frequency increases, you mark points of $x$ and $y$ coordinates, made of the real and imaginary parts of the loop gain function $T$, respectively:

$$T(j\omega) = x + jy = \operatorname{Re} T(j\omega) + j \operatorname{Im} T(j\omega) \tag{3.11}$$

A point on the plot corresponds to a certain angular frequency $\omega$. You can reconstruct the transfer function magnitude and phase related to this point by

$$|T(\omega)| = \sqrt{\operatorname{Re}^2 T(\omega) + \operatorname{Im}^2 T(\omega)} \tag{3.12}$$

$$\angle T(\omega) = \tan^{-1} \frac{\operatorname{Im} T(\omega)}{\operatorname{Re} T(\omega)} \tag{3.13}$$

In this graph, there is one particular point where the magnitude of the function is 1 and its argument is $-\pi$. This point, called the "–1" point, appears in the graph at the position $-1,j0$: when the open-loop gain curves exactly passes through this point, we have conditions for oscillations as described by (3.6) and (3.7) with an illustration in Figure 3.2. Coming close to this point also means that the denominator $D(s)$ in (3.3) dangerously approaches zero.

Now suppose you sit on a motorcycle and drive along the path in the prescribed direction as shown in Figure 3.13. All the points you see on your right hand are said to be *enclosed*. For instance, the point positioned at $1,-j$ is enclosed while the "–1" point is not.

For a minimum-phase system, theory teaches us that stability is ensured if the "–1" point is *not* enclosed while sliding the path along the prescribed direction.

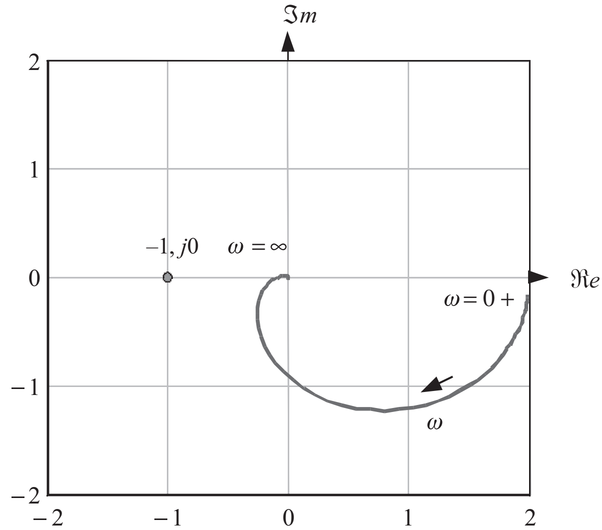



**Figure 3.12** A Nyquist plot represents the loci of the loop gain real and imaginary parts as $\omega$ increases from 0 to infinity.

Practically speaking, if you leave the "–1" point on your left while going in the prescribed direction, the system is stable. This rule is also known as the Nyquist *left-hand criterion*. An application example is given in Figure 3.14 with two functions *a* and *b*.

When going along the prescribed direction, function *b* encloses the "–1" point and is unstable. On the contrary, function *a* leaves the point on the left while sliding along the path: the function is stable. This simple stability criterion for the Nyquist plot is valid for minimum-phase functions only. For nonminimum phase functions, you will need to apply the more complex Cauchy argument principle whose usage goes beyond this simple introduction.

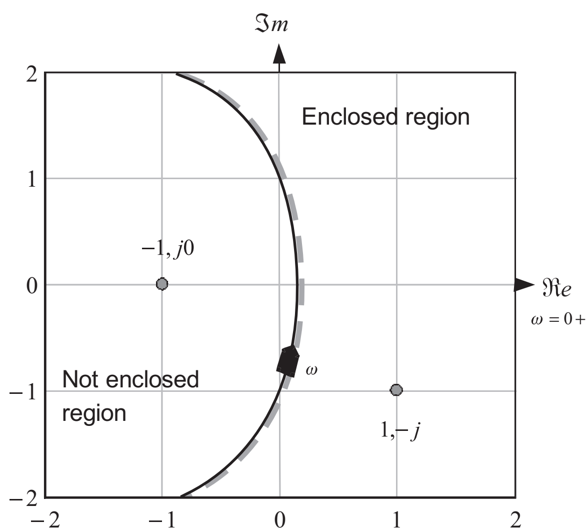



**Figure 3.13** All points situated in the right of a path driven along the prescribed direction are said to be enclosed.

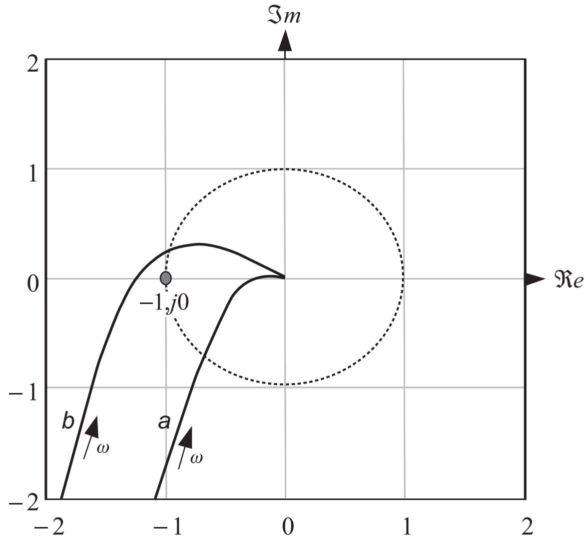



**Figure 3.14** In this example, curve b encloses the "–1" and is unstable, while curve a leaves the "–1" on the left when going along the prescribed direction: the function is stable.

### 3.2.4 Extracting the Basic Information from the Nyquist Plot

From a Nyquist plot, the first information you can extract is the crossover angular frequency, $\omega_c$. Looking at Figure 3.15, we can see that from every coordinates pair $\mathrm{Re}\,T(\omega)$ and $\mathrm{Im}\,T(\omega)$, we can form a triangle whose hypotenuse $h$ starts at the graph origin. Considering triangle geometry and in particular Pythagoras theorem, we can write:

$$h(\omega) = \sqrt{\mathrm{Re}^2\, T(\omega) + \mathrm{Im}^2\, T(\omega)} = \left|T(\omega)\right| \tag{3.14}$$

This is the loop gain magnitude definition. When the curve crosses an origin-centered circle of radius 1, the hypotenuse equals 1: we have our crossover angular frequency at $\omega = \omega_c$. The drawback with Nyquist is that the crossover angular frequency is not displayed on the graph. You need to check the point at which the circle crossing occurs. This is admittedly less convenient than with a Bode plot.

Equation (3.8) tells us that the phase margin is the distance from the open-loop gain argument to the –180° or $-\pi$ limit at the crossover frequency (Figure 3.7). On the Nyquist graph, the line –180° or $-\pi$ is the left end of the $x$-axis when turning clockwise from the right end of this axis. The argument of the open-loop gain $T(s)$ is thus the angle made between the positive section of the $x$-axis and the hypotenuse $h$ at the crossover angular frequency when turning clockwise. Figure 3.16 details how to form this angle. Finally, the phase margin $\varphi_m$ is the remaining angle formed by the hypotenuse $h$ and the negative section of the $x$-axis. The sum of $\angle T(\omega_c)$ and $\varphi_m$ is well equal to –180°, as defined by (3.8).

The gain margin requires a little more attention. The gain margin concerns the observation of the loop gain magnitude at a frequency $f_\pi$ where the loop-gain argument is –180° or $-\pi$ (see Figure 3.7 for instance). How much do we need to shift the

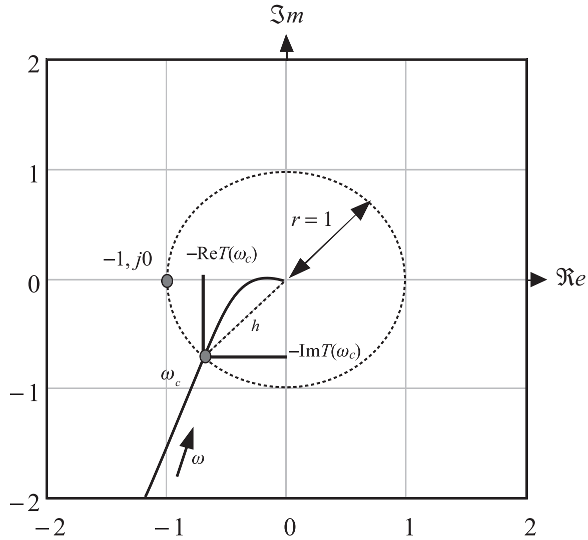



**Figure 3.15** When the curve crosses a circle of radius 1, this is the point where the magnitude of $T$ is 1.

curve up (or down) to obtain a new 0-dB crossing point and conditions for oscillations? To satisfy this condition, we can write:

$$\left|T(f_\pi)\right| \cdot x = 1 \tag{3.15}$$

The multiplicand $x$ is nothing else than our gain margin GM previously defined:

$$\mathrm{GM} = \frac{1}{\left|T(f_\pi)\right|} \tag{3.16}$$

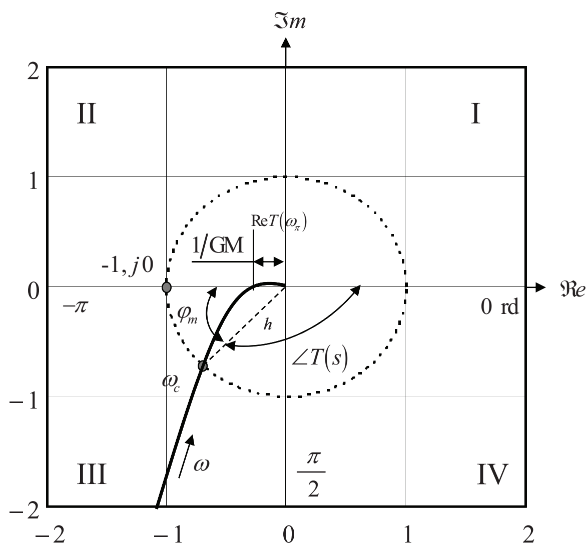



**Figure 3.16** The open-loop gain argument is found as the angle measured from the $x$-axis to the triangle hypotenuse at the point where the curve crosses the circle of radius 1.

Where in the Nyquist plot do we see the point where the loop-gain argument is –180°? This is exactly when the curve crosses the $x$-axis in the left side of the plot in Figure 3.16. At this particular point, the imaginary portion of $T(s)$ is exactly 0. The magnitude expression thus simplifies to:

$$\mathrm{GM} = \frac{1}{\left|\mathrm{Re}\,T(f_\pi)\right|} \tag{3.17}$$

The real part value is directly measurable on the graph as shown in Figure 3.16. Suppose you find the crossing point with the $x$-axis is 0.25; then, the gain margin is simply:

$$\mathrm{GM} = \frac{1}{0.25} = 4 = 12\ \mathrm{dB} \tag{3.18}$$

### 3.2.5 Modulus Margin

The gain margin alone is not a sufficient condition to qualify the system robustness. By robustness, we mean its ability to efficiently reject the incoming perturbations such as the input voltage or the output current in a voltage regulator case (linear or switching). In the opening paragraph related to Nyquist, we said that the curve must not enclose the "–1" point and should always stay away from it. Our interest is now to check the shortest distance from the curve to the "–1" point at any angular frequency value. This distance is represented by the new hypotenuse $h$ drawn in Figure 3.17.

Applying again Pythagorean geometry and considering the height of the triangle as the imaginary part of the loop gain $\mathrm{Im}\,T(\omega)$, we can write:

$$h = \sqrt{[1 + \mathrm{Re}\,T(\omega)]^2 + [\mathrm{Im}\,T(\omega)]^2} \tag{3.19}$$

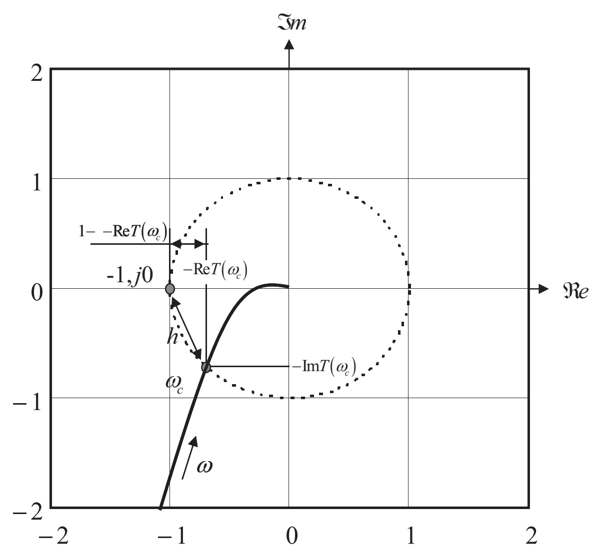



**Figure 3.17** The distance between the "–1" point and the curve is represented by the hypotenuse $h$.

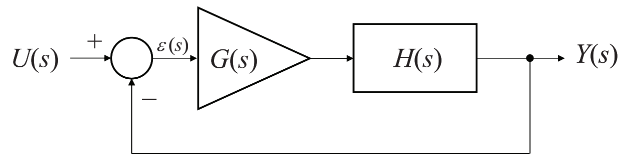



**Figure 3.18** The ratio between the error $\varepsilon$ and the setpoint $U$ is called the sensitivity function $S$.

This expression is nothing other than the magnitude of $D(s)$ in our closed-loop gain expression as defined in (3.3):

$$h = |1 + T(s)| \tag{3.20}$$

When this hypotenuse length diminishes, the curve dangerously approaches the "–1" point and the ability to reject the perturbations can be at stake. If we look at Figure 3.18, where a simple unity-return closed-loop system appears, we can derive the relationship linking the error variable $\varepsilon$ to the input setpoint $u$:

$$\frac{\varepsilon(s)}{U(s)} = \frac{1}{1 + T(s)} = S \tag{3.21}$$

This function is called the *sensitivity* function, noted $S$. Ideally, $S$ should be extremely small, meaning that $\varepsilon$, the error, becomes negligibly low and the output matches the setpoint.

Now, assume a perturbation input is inserted into our system, as described by Figure 3.19. This perturbation $u_2$ could be the input voltage for our converters or the delivered output current.

The transfer function of such a closed-loop system can be derived applying the superposition theorem since we deal with a linear system. First, if we ground the perturbation input $U_1$, we classically have:

$$Y(s)\big|_{U_1=0} = U_2(s)\frac{T(s)}{1 + T(s)} \tag{3.22}$$

Now, if we ground $U_2$, we obtain the following expression:

$$Y(s)\big|_{U_2=0} = -U_1(s)\frac{1}{1 + T(s)} \tag{3.23}$$

The complete expression is thus the sum of (3.22) and (3.23):

$$Y(s) = U_2(s)\frac{T(s)}{1 + T(s)} - U_1(s)\frac{1}{1 + T(s)} \tag{3.24}$$

As you can read in the previous equation, the perturbation rejection depends on $S$, which is the inverse of $|1 + T(s)|$. As $|1 + T(s)|$ becomes smaller, the ability to

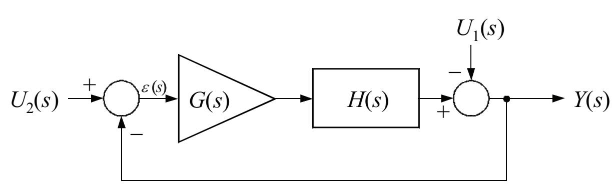



**Figure 3.19** When a perturbation is inserted in the control chain, its rejection depends on the sensitivity function.

reject the perturbation weakens and system robustness suffers. Beyond crossover, where the loop gain is below 1, the system loop no longer fights the incoming perturbations. Ideally, the natural gain reduction over frequency must be smooth, without peaking. However, if a hidden resonance suddenly makes $|1+T(s)|$ go to a very small value, much less than one, then the sensitivity function peaks and the perturbation is amplified rather than attenuated. The peak of the sensitivity function or the minimum of $|1+T(s)|$ corresponds to the shortest value of the hypotenuse $h$ in Figure 3.17. It is also the shortest distance between our Nyquist curve and the "–1" point. It is thus always advisable to maintain a minimum value for $h$. This value is the *modulus margin*, noted $\Delta M$. The term *modulus* can be replaced by the word *magnitude*, if you prefer. It is mathematically defined as the smallest acceptable radius of a circle centered at the point $-1, j0$ and tangent to the loop gain loci. It is usual to adopt a radius of 0.5 for this circle. You can show by calculating the intersection between the 0.5-radius circle and the 1-radius circle that keeping $\Delta M > 0.5$ ensures a gain margin better than 6 dB and a phase margin greater than 30°.

Any points of the loci entering the circle violate the minimum modulus margin we stated. This principle appears as a graphical illustration in the left side of Figure 3.20. As long as no points belonging to the Nyquist loci enter the circle, the modulus margin principle is respected and the gain margin criteria is always fulfilled. As shown on the right side of the same picture, the gain margin can very well be correct; it is not a sufficient condition for stability: a hidden resonance briefly routes the loci into the circle, while returning to its trajectory afterward. You might think the gain margin criterion is satisfied, but the system robustness would be at stake with this fugitive excursion. As a general statement, if the modulus margin is satisfied, the gain margin will also be. In most textbooks, the modulus margin, not the gain margin, is considered a robustness criterion.

For the sake of the practical example, we simulated a buck converter compensated by a structure combining a double pole/double zero arrangement and purposely left the subharmonic poles untreated (weak slope compensation). The corresponding Nyquist plot from the control input to the error amplifier output

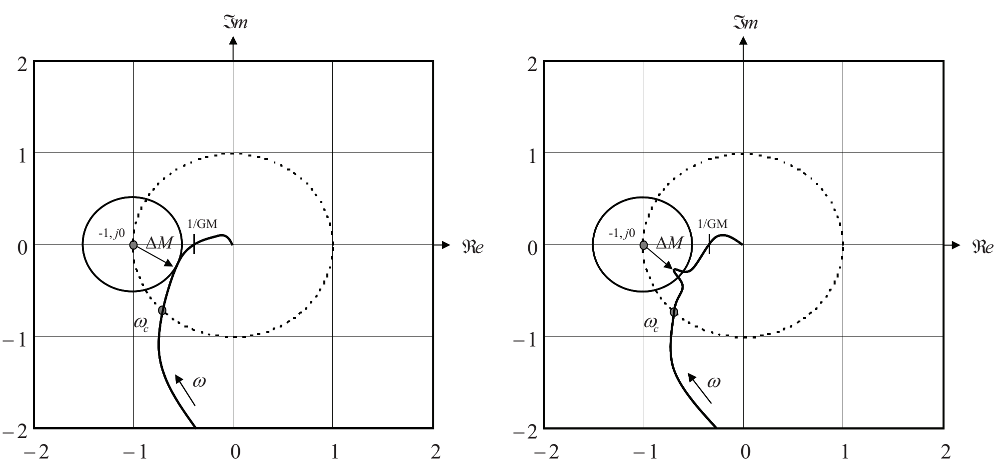



**Figure 3.20** Any point entering the 0.5-radius circle violates the 6-dB modulus margin rule.

appears in the upper side of Figure 3.21. We can see the gain margin is within the circle, implying a value less than 6 dB (4.4 dB)—not very good. However, this is not enough to safely characterize the system as the trajectory approaches the "–1" point even closer at a 26.3-kHz frequency, where the distance is further reduced. The modulus margin is way too small here.

A Bode plot can also be used to check the modulus margin criteria. You can derive the sensitivity function $S$ and plot its magnitude in decibels. You can do that in SPICE by combining the real and imaginary parts of the loop gain $T$:

$$|S| = \frac{1}{|1+T(s)|} = \frac{1}{\sqrt{(1+\operatorname{Re}T(s))^2 + (\operatorname{Im}T(s))^2}} \tag{3.25}$$

Some graphical analysis tools accept programming lines. For instance, with Intusoft's IntuScope graphical viewer, the code could look as follows:

```
* Sensitivity for a transfer function
assertvalid vin vout
phase = phaseextend(phase(vout)–phase(vin)
gain = db(vout) – db(vin)
mag = 10^(gain/20)
real = –mag*cos(phase)
imag = –mag*sin(phase)
```

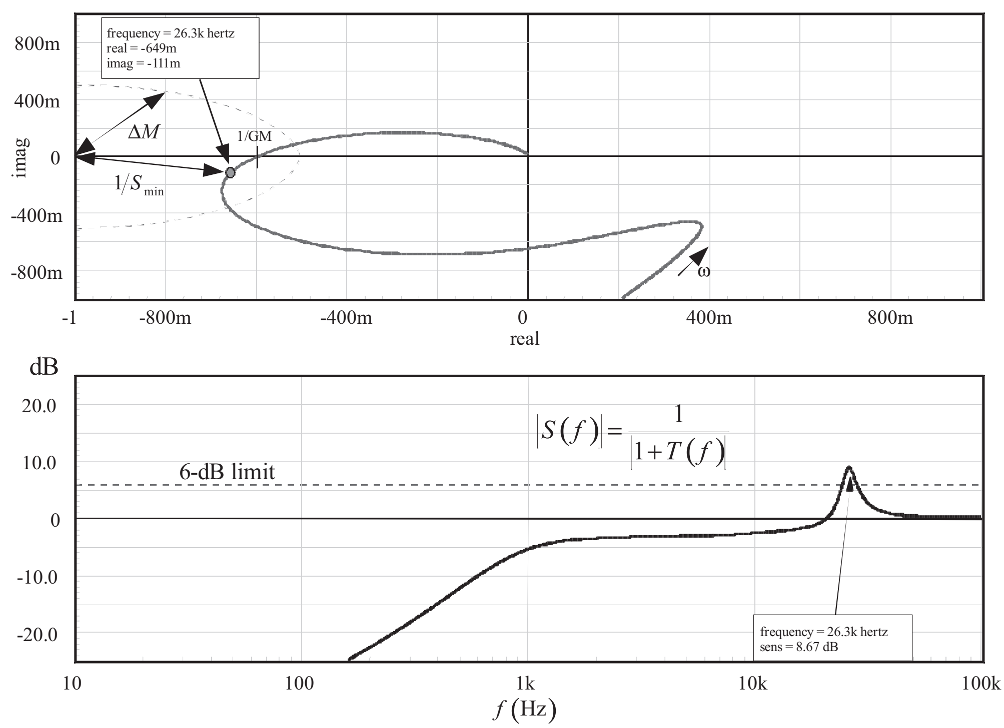



**Figure 3.21** The Nyquist plot shows a violation of the 0.5-radius circle, confirmed by the sensitivity function that clearly peaks beyond 6 dB. It confirms the design problem: the modulus margin is too small.

*Re2 = (real+1)^2*
*Imag2 = (imag)^2*
*D = sqrt(Re2+Imag2)*
*sens = db(1/D)*
*plot sens*

If a peak exceeds the 6-dB limit, the modulus margin is violated. You can clearly see this peak in the lower side of Figure 3.21.

## 3.3 Transient Response, Quality Factor, and Phase Margin

We have seen how the gain and modulus margins must be characterized to build robust systems. What design strategy must we now adopt for the other important parameter, the phase margin? In other words, what is the minimum phase margin we should tweak our design to? In most of the available textbooks, and this is what I learned when I was a student, it is often stated that the phase margin should never be less than 45° at the crossover frequency. Period. However, as underlined by Dr. David Middlebrook in his design-oriented analysis course, this is neither the right question nor the correct answer. The phase margin selection depends on the type of transient response you expect from your control system. If you look back to Figure 3.6, in the upper section, you can see a waveform whose amplitude is decaying further to an excitation stimulus. What does this ringing signal remind you? Yes, the response delivered by an *RLC* circuit subjected to a step input. We know that depending on the *quality factor Q* or the *damping ratio* $\zeta$ (pronounced "zeta") of the network, the ringing signal will decay more or less quickly. At a certain point, if the damping ratio is cancelled, you have permanent oscillations as in the middle of Figure 3.6. Let us first explore the response delivered by a second-order system, a *RLC* network, to smoothly unveil its relationship to the closed-loop quality factor and the open-loop phase margin.

### 3.3.1 A Second-Order System, the RLC Circuit

An *RLC* circuit appears in Figure 3.22 where the input is stepped from 0 V to 1 V in a very short time while the output is observed.

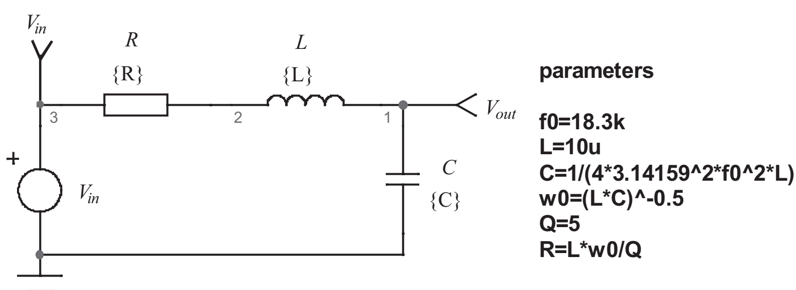



**Figure 3.22** A *RLC* network receives an input voltage step and delivers an oscillatory response for high *Q* values.

To have an idea of the time-domain response of such a circuit, we can first evaluate its transfer function. Let's use the classical approach using impedance ratios, as the result is straightforward:

$$V_{out}(s) = V_{in}(s)\frac{\frac{1}{sC}}{R + sL + \frac{1}{sC}} \tag{3.26}$$

Rearranging this equation leads to the following transfer function:

$$\frac{V_{out}(s)}{V_{in}(s)} = \frac{1}{LCs^2 + RCs + 1} \tag{3.27}$$

This expression can be put in the form of a second-order system using either the *quality factor* $Q$ or the *damping ratio* $\zeta$:

$$\frac{V_{out}(s)}{V_{in}(s)} = \frac{1}{\frac{s^2}{\omega_0^2} + 2\zeta\frac{s}{\omega_0} + 1} = \frac{1}{\frac{s^2}{\omega_0^2} + \frac{s}{\omega_0 Q} + 1} \tag{3.28}$$

Identifying $\omega_0$, $Q$, and $\zeta$ within (3.27), we have the following relationships:

$$\omega_0 = \frac{1}{\sqrt{LC}} \tag{3.29}$$

$$f_0 = \frac{\omega_0}{2\pi} = \frac{1}{2\pi\sqrt{LC}} \tag{3.30}$$

$$\zeta = \frac{1}{2Q} = \frac{R}{2}\sqrt{\frac{C}{L}} = \frac{RC\omega_0}{2} = \frac{R}{2L\omega_0} \tag{3.31}$$

$$Q = \frac{1}{2\zeta} = \frac{1}{R}\sqrt{\frac{L}{C}} = RC\omega_0 = \frac{L\omega_o}{R} \tag{3.32}$$

An $LC$ network can be seen as the linear combination of two storage elements, a capacitor $C$ and an inductor $L$: this is a second-order system. When a voltage step appears on the input $V_{in}$, a current circulates in the network. This current induces energy storage in $L$, while this is the voltage developed across $C$ that dictates the energy it stores. In our example, when the input signal no longer changes, the energy keeps circulating between both elements as a pendulum swinging back and forth: an oscillatory process takes place, implying, without losses, an endless energy transfer between the two storage elements, $C$ and $L$. If a resistor $R$ is introduced in the path, part of the energy is lost in heat while circulating from $C$ to $L$ and vice versa. As the stored energy diminishes when swinging from one element to the other, the resulting oscillatory signal exponentially decays in amplitude until it completely ceases. In the $RLC$ circuit, the quality factor—or its counter part the damping ratio—quantifies the amount of losses in the network via ohmic paths:

- $Q$ is high or $\zeta$ is low $\rightarrow$ low losses, weak damping, high ringing;
- $Q$ is low or $\zeta$ is high $\rightarrow$ high losses, high damping, low or no ringing.

The circuit is *damped* by the resistor presence in the Figure 3.22 circuit.

To gain further insight into the $RLC$ network response, it is interesting to study the denominator $D(s)$ of (3.28). It is of the following form:

$$D(s) = \frac{s^2}{\omega_0^2} + \frac{s}{\omega_0 Q} + 1 \tag{3.33}$$

The poles or the roots of the equation can be found by solving the 2$s$ values, which satisfy $D(s) = 0$:

$$\frac{s^2}{\omega_0^2} + \frac{s}{\omega_0 Q} + 1 = 0 \tag{3.34}$$

Applying classic algebra formulas, we have the following roots, already introduced in Chapter 2:

$$s_1, s_2 = \frac{\omega_0}{2Q}\left(\pm\sqrt{1 - 4Q^2} - 1\right) \tag{3.35}$$

Considering the damping ratio instead:

$$s_1, s_2 = \omega_0 \zeta\left(\pm\sqrt{1 - \frac{1}{\zeta^2}} - 1\right) \tag{3.36}$$

This expression is close to the one we found in Chapter 2, when looking at the closed-loop compensated converter. Actually, the discussion about $Q$ is similar as we deal with a second-order system:

- $Q < 0.5$ or $\zeta > 1$: the expression below the square root is positive and the roots are negative, real, and separate: there is no imaginary number in the solutions. We have a fully nonringing response; the system is said to be *overdamped*. In that case, theory tells us that (3.28) can be reformulated considering two cascaded $RC$ filter.
- $Q = 0.5$ or $\zeta = 1$: the square root returns 0 and both roots are real, negative, and coincident. The system is said to be *critically damped*. We still have a nonoscillatory response, as no imaginary numbers appear in the roots.

$$s_{1,2} = -\omega_0 \tag{3.37}$$

- $Q > 0.5$: the expression below the square root becomes negative and both roots welcome imaginary numbers: we have an oscillatory response, damped by the presence of the real part, the ohmic losses in the circuit. The poles are complex conjugates with negative real parts:

$$s_{1,2} = -\frac{\omega_0}{2Q} \pm j\omega_0\sqrt{1-\frac{1}{4Q^2}} \tag{3.38}$$

also equal to:

$$s_{1,2} = -\zeta\omega_0 \pm j\omega_0\sqrt{1-\zeta^2} \tag{3.39}$$

In the previous expression, the terms $1/\zeta\omega_0$ or $\dfrac{2Q}{\omega_0}$ define the time constant of the system.

- As $Q$ approaches infinity, or $\zeta$ reaches zero, the real portion fades away (the ohmic losses diminish) until the roots become pure imaginary numbers: we have a fully *undamped* system, also called an oscillator.

To help figure out what is exactly going on, it is interesting to represent the roots loci on the $s$-plan as introduced in the previous chapter. This is what Figure 3.23 represents. If we follow the arrows, we can see that a low-$Q$ condition sets the roots apart, without imaginary portion at all. As the quality factor increases, the roots

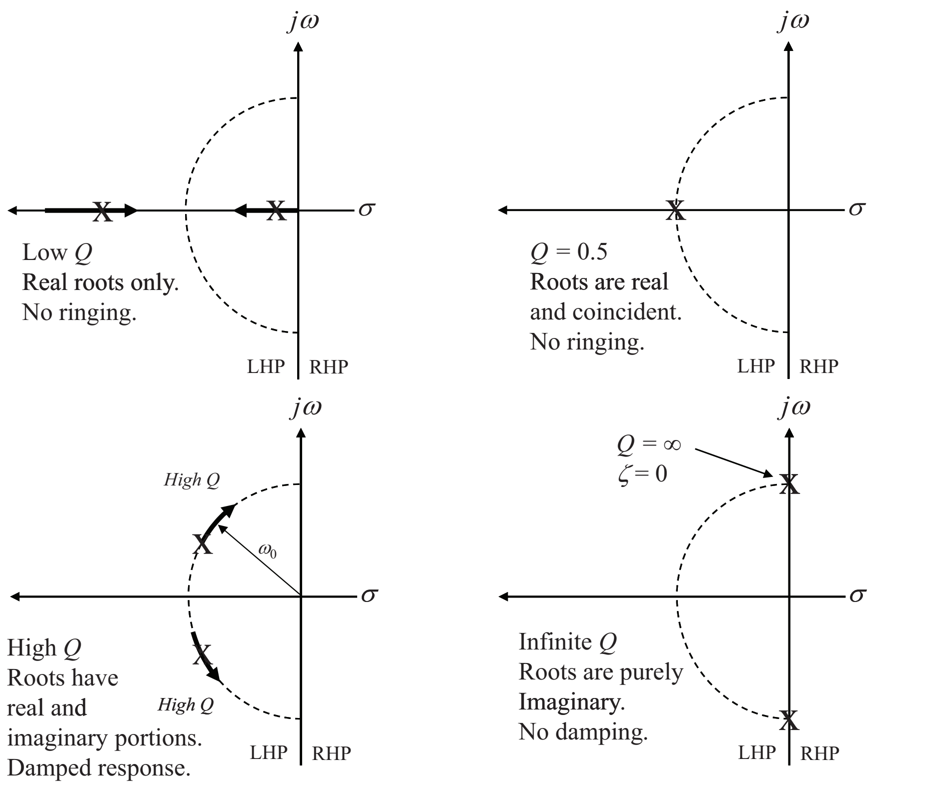



**Figure 3.23** The $s$-plane helps to locate the position of the roots and see the contribution of their real or imaginary parts.

move along the $x$-axis until they meet at $Q$ equals 0.5. At this point, the roots are coincident and still purely real. Then, as $Q$ increases, the imaginary portion starts to show up while the real contribution begins to diminish. The roots slide along a semicircle of radius $\omega_0$. As $Q$ reaches infinity, the real portion—representative of damping or losses—has left and the roots are pure imaginary numbers: the system starts oscillating and never stops.

### 3.3.2 Transient Response of a Second-Order System

With the transfer function of our $RLC$ network in the $s$-domain being known, we can derive the delivered response in the time domain. Such an exercise requires the selection of a stimulus applied to the system input. In our case, since we want the response to a step, such a stimulus expression in the Laplace domain is $1/s$. We have

$$v_{out}(t) = \mathcal{L}^{-1}\left\{\frac{1}{s}\frac{1}{\frac{s^2}{\omega_0^2}+2\zeta\frac{s}{\omega_0}+1}\right\} \tag{3.40}$$

The result of this calculus for $\zeta < 1$ appears next:

$$v_{out}(t) = 1 - \frac{1}{\sqrt{1-\zeta^2}}e^{-\zeta\omega_0 t}\sin(\omega_d t + \theta) \tag{3.41}$$

With a damped angular frequency equal to

$$\omega_d = \omega_0\sqrt{1-\zeta^2} \tag{3.42}$$

$$\theta = \cos^{-1}(\zeta) \tag{3.43}$$

Note that both the damped angular frequency ($\omega_d$) and the natural angular frequency ($\omega_0$) are equal when the damping ratio is 0 or the quality factor is infinite. Equation (3.41) can be plotted using a dedicated solver like Mathcad or simply by simulating the Figure 3.22 schematic. The results appear in Figure 3.24 for various $Q$ values.

The right side of (3.41) shows a sinusoidal response in which amplitude is affected by an exponential term. As $t$ increases, provided the exponential exponent remains negative, the sinusoid waveform decays in amplitude to become 0 at $t = \infty$. For certain quality factor values, the response is fully nonoscillatory (e.g., $Q = 0.1$ and $Q = 0.5$). For $Q = 0.1$, the response is extremely slow and slightly faster for $Q = 0.5$. In both cases, note that there is no overshoot. For $Q$ greater than 0.5, an oscillation superimposes on the signal and brings large over-/undershoots as the system is less damped ($Q$ from 1 to 5). Please note that in spite of large oscillations, the signal eventually stabilizes to 1 V, the input setpoint signal: the response is oscillatory but stable. An unstable system would have a diverging output when responding to an input step (e.g., when the damping disappears and the quality factor becomes infinite as in the lower right side of Figure 3.23).

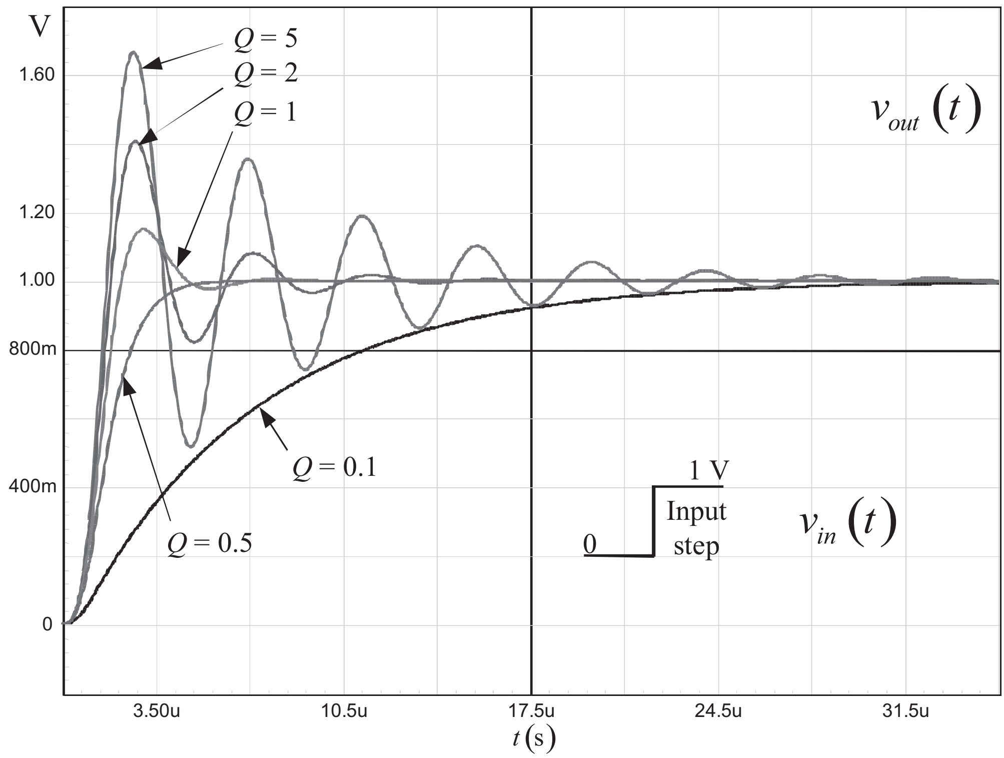



**Figure 3.24** Time-domain response of the *RLC* network from Figure 3.22.

Given the impact of the quality factor on the step response, it is important to select the right $Q$ depending on the criteria we want to meet with our control system. Figure 3.25 portrays the typical response of a second-order system where critical parameters affected by $Q$ appear. These parameters are the following:

- The rise time, $t_r$: this is the time needed for the output to rise from 10 percent to 90 percent of the steady-state value for an overdamped response. For an undamped circuit, whose response appears in Figure 3.25, we consider the time needed to reach 100 percent of the steady-state value, 1 V. After the picture, this time is 17 µs.
- The peak time, $t_p$: for a $Q$ greater than 0.5, at this moment of time, the signal exceeds the steady-state level and peaks to a maximum before decreasing again.
- The maximum percent overshoot, $M_p$: this is the maximum value the signal can take at the time $t_p$, expressed as a percentage of the steady-state level.
- The settling time, $t_s$: at this moment of time, the output signal is considered to be within the band of tolerance that you or your customer has set. At the time $t_s$, the output is considered to be steady state.

As detailed in Figure 3.24, depending on the value of $Q$, different responses are possible. For instance, you might want a response where absolutely no overshoot is accepted. In that case, a $Q$ below 0.5 must be selected. A specification could also be

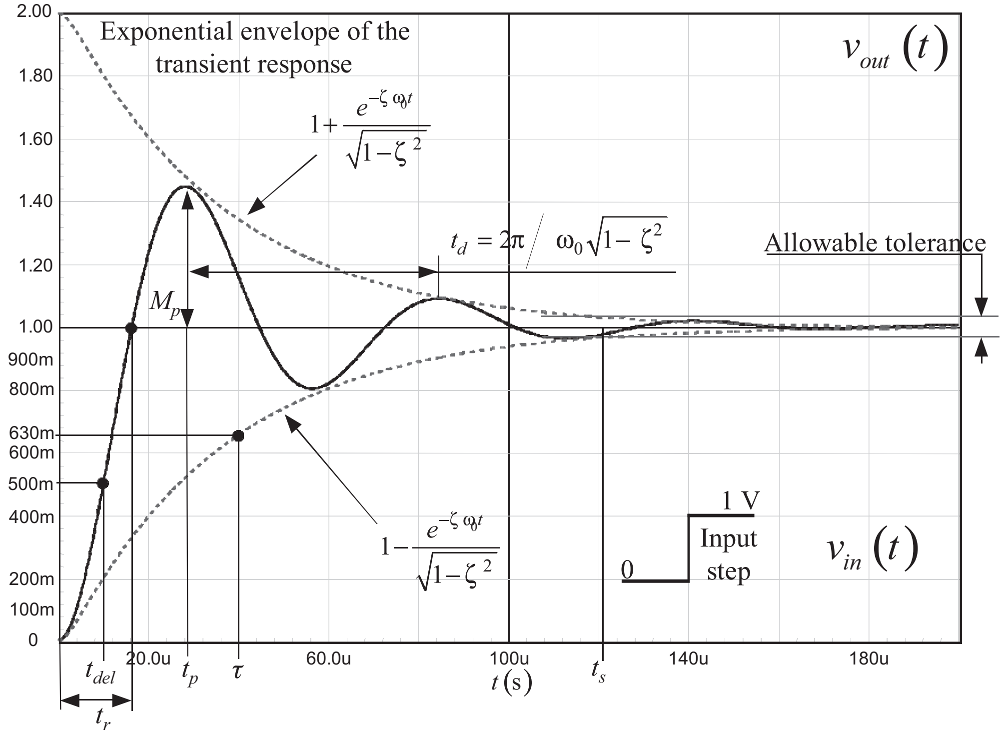



**Figure 3.25** This is the classical second-order response found in textbooks where critical parameters appear. In this example, $Q = 2$ or $\zeta = 0.25$.

that a certain amount of overshoot is accepted as long as the signal rises at a pace sufficiently high, specified by the customer. For all of these reasons, it is important to analytically derive some of these important timings. Please note that these equations are derived for a $Q$ greater than 0.5 where the roots are complex (real and imaginary coefficients).

The delay time $t_{del}$ corresponds to the time needed by the output signal to reach 50 percent of the steady-state value, 0.5 V. From (3.41), the equation can be written as follows:

$$0.5 = \frac{1}{\sqrt{1-\zeta^2}} e^{-\zeta\omega_0 t_{del}} \sin(\omega_d t_{del} + \theta) \tag{3.44}$$

Solving this equation symbolically requires a certain amount of mathematical dexterity that I don't have! Happily, an approximated expression has already been derived and appears next:

$$t_{del} \approx \frac{1+0.7\zeta}{\omega_0} \tag{3.45}$$

To plot Figure 3.25, we used a network whose natural frequency was 18.3 kHz with a quality factor of 2. When applying numerical values to (3.45), it gives 10.22 $\mu$s, whereas Mathcad® numerical solver delivers 10.064 $\mu$s.

For an overdamped system, the rise time $t_r$ is the time needed for the response signal to rise from 10 percent to 90 percent of the final value. For an undamped second-order system, let's say with 10 to 30 percent overshoot, the rise time is defined as the time needed to reach 100 percent of the final value or 1 V in the example. The equation to obtain the rise time definition in this last case is the following:

$$v_{out}(t_r) = 1 = 1 - \frac{1}{\sqrt{1-\zeta^2}} e^{-\zeta\omega_0 t_r} \sin(\omega_d t_r + \theta) \tag{3.46}$$

Otherwise stated:

$$0 = \frac{1}{\sqrt{1-\zeta^2}} e^{-\zeta\omega_0 t_r} \sin(\omega_d t_r + \theta) \tag{3.47}$$

The exponential term $\frac{1}{\sqrt{1-\zeta^2}} e^{-\zeta\omega_0 t_r}$ represents the waveform envelope and cannot be null in the previous equation. Therefore, we need to solve

$$0 = \sin(\omega_d t_r + \theta) = \sin\left(\omega_0 t_r \sqrt{1-\zeta^2} + \cos^{-1}(\zeta)\right) \tag{3.48}$$

Solving for the value of $t_r$ that can cancel the sinus term will give us the result we are looking for:

$$\sin\left(\omega_0 t_r \sqrt{1-\zeta^2} + \cos^{-1}(\zeta)\right) = 0 \tag{3.49}$$

Sin($x$) is zero if $x = n\pi$. Considering the case $n = 1$ and solving for $t_r$, we have

$$t_r = \frac{\pi - \cos^{-1}(\varsigma)}{\omega_0\sqrt{1-\varsigma^2}} \tag{3.50}$$

For a quality factor of 2 or a damping ratio of 0.25, we find a rise time of 16.4 $\mu$s, in agreement with Figure 3.25. For small damping ratios, the previous expression simplifies to

$$\lim_{\zeta\to 0} \frac{\pi - \cos^{-1}(\varsigma)}{\omega_0\sqrt{1-\varsigma^2}} \approx \frac{\pi - \frac{\pi}{2}}{\omega_0} \approx \frac{1.6}{\omega_0} \tag{3.51}$$

In our case with a 18,300-Hz natural frequency, we have an approximate rise time of

$$t_r \approx \frac{1.6}{2\times\pi\times 18300} \approx 14\ \mu\text{s} \tag{3.52}$$

Let us now calculate the time $t_p$ at which the output is maximum. To obtain this number, we will derive $v_{out}(t)$ and calculate the time value for which the result equals zero. Mathematically, it means

$$\frac{d}{dt}\left(1 - \frac{1}{\sqrt{1-\zeta^2}} e^{-\zeta\omega_0 t} \sin(\omega_d t + \theta)\right) = 0 \tag{3.53}$$

The derivative of a product *uv* is *vu*'-*uv*', thus,

$$\frac{dv_{out}(t)}{dt} = \frac{\zeta\omega_0 e^{-\zeta\omega_0 t}\sin(\omega_d t+\theta)}{\sqrt{1-\zeta^2}} - \frac{\omega_d e^{-\zeta\omega_0 t}\cos(\omega_d t+\theta)}{\sqrt{1-\zeta^2}} = 0 \tag{3.54}$$

Simplifying, we have

$$\zeta\omega_0\sin(\omega_d t+\theta) - \omega_d\cos(\omega_d t+\theta) = 0 \tag{3.55}$$

$\zeta$ can be extracted from (3.43) to obtain

$$\zeta = \cos(\theta) \tag{3.56}$$

If we substitute this value in (3.42), we obtain a new definition for $\omega_d$, valid for $0 \le \theta \le \pi$:

$$\omega_d = \omega_0\sqrt{1-\cos^2(\theta)} = \omega_0\sin(\theta) \tag{3.57}$$

Combining (3.56) and (3.57) into (3.55),

$$\sin(\omega_d t+\theta)\cos(\theta)\omega_0 - \omega_0\cos(\omega_d t+\theta)\sin(\theta) = 0 \tag{3.58}$$

If we divide everything by $\omega_0$, we have an equation of the form

$$\sin a\cos b - \cos a\sin b = \sin(a-b) \tag{3.59}$$

Otherwise stated,

$$\sin(\omega_d t+\theta-\theta) = 0 \tag{3.60}$$

Sin $(x)$ can go to zero for $x = 0$ or $x = n\pi$. In our case, $t = 0$ is the starting time; therefore, the next possible value is $\pi$:

$$\omega_d t = \pi \tag{3.61}$$

If we solve for $t_p$, we obtain

$$t_p = \frac{\pi}{\omega_d} = \frac{\pi}{\omega_0\sqrt{1-\zeta^2}} = \frac{1}{2f_0\sqrt{1-\zeta^2}} \tag{3.62}$$

The numerical value for a $Q$ of 2 or a damping ratio $\zeta$ of 0.25 is $t_p \approx 28.2\ \mu$s. As you can see on Figure 3.25, this value corresponds to half of the oscillating period. The damped period $t_d$ is thus obtained by a simple multiplication by two, in agreement with (3.42):

$$t_d = \frac{2\pi}{\omega_0\sqrt{1-\zeta^2}} = \frac{1}{f_0\sqrt{1-\zeta^2}} \tag{3.63}$$

Now that the time at which the function peaks is known, we can substitute its value into (3.41) while replacing $\omega_d$ by its definition with (3.42):

$$v_{out}(t_p) = 1 - \frac{1}{\sqrt{1-\zeta^2}} e^{-\zeta\omega_0\left(\frac{\pi}{\omega_0\sqrt{1-\zeta^2}}\right)} \sin\left(\omega_0\sqrt{1-\zeta^2}\left(\frac{\pi}{\omega_0\sqrt{1-\zeta^2}}\right) + \theta\right)$$

$$= 1 + \frac{e^{-\frac{\zeta\pi}{\sqrt{1-\zeta^2}}}}{\sqrt{1-\zeta^2}} \sin(\theta) \tag{3.64}$$

From (3.56), the right term in the denominator of (3.64) can be reformulated, again considering $0 \leq \theta \leq \pi$:

$$\sqrt{1-\zeta^2} = \sqrt{1-\cos^2\theta} = \sqrt{\sin^2\theta} = \sin\theta \tag{3.65}$$

Therefore, (3.64) can be further simplified to

$$v_{out}(t_p) = 1 + e^{-\frac{\zeta\pi}{\sqrt{1-\zeta^2}}} \tag{3.66}$$

The mathematical definition for the maximum percent overshoot $M_p$ is the following one:

$$M_p = \frac{V_{out}(t_p) - V_{out}(\infty)}{V_{out}(\infty)} \tag{3.67}$$

With our steady-state value being 1 V, the right-side extra term in (3.66) is simply the overshoot:

$$M_p(\%) = e^{-\frac{\zeta\pi}{\sqrt{1-\zeta^2}}} \times 100 \tag{3.68}$$

If we express the damping ratio in relationship to the quality factor $Q$, the previous formula becomes

$$M(\%) = e^{-\frac{\pi}{\sqrt{4Q^2-1}}} \times 100 \tag{3.69}$$

From (3.68) and (3.69), we can plot the relationship between the overshoot and the quality factor (or the damping ratio) of our second-order system. It appears in Figure 3.26.

For the quality factor of 2 used to plot Figure 3.25, we have a 44 percent overshoot. What is the maximum overshoot value the output can be affected by? Well, when the damping ratio reaches zero or when the quality factor goes to infinity, we have

$$\lim_{Q\to\infty} e^{-\frac{\pi}{\sqrt{4Q^2-1}}} = 100\% \tag{3.70}$$

In this case, the input step is simply doubled. This is what was already shown in Figure 3.25, where the top of the envelope is 2 V, corresponding to a 100 percent overshoot.

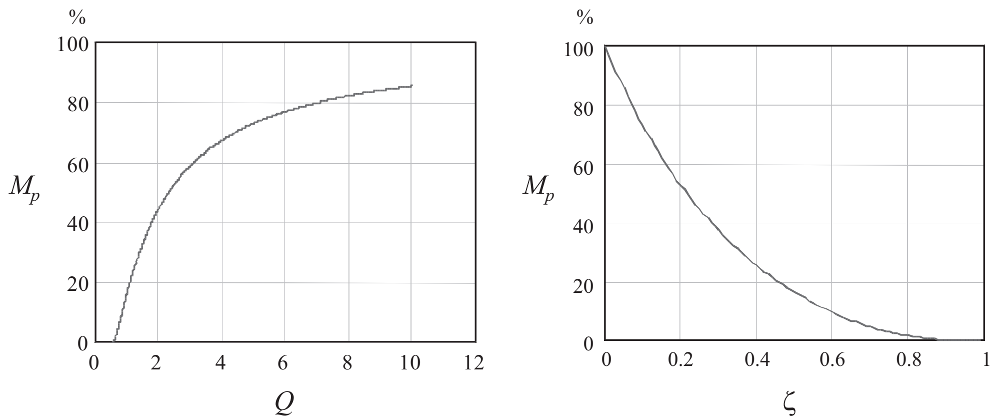



**Figure 3.26** The overshoot depends on the selected quality factor when above 0.5 or on the damping ratio when lower than 2.

By applying these formulas, we can find the various peaks values. The first one is equal to

$$V_{peak1} = \left(1 + e^{-\frac{\pi}{\sqrt{4Q^2 - 1}}}\right) \times 1 = (1 + 0.444) \times 1 = 1.44 \text{ V} \tag{3.71}$$

in line with Figure 3.25 results.

The rest of the peaks can be found by selecting different $n\pi$, with $n = 2$ (negative), $n = 3$ (positive), and so on. For instance, the second peak, which is an undershoot, can be computed as follows given the 1-V step input:

$$V_{peak2} = \left(1 - e^{-\frac{2\pi}{\sqrt{4Q^2 - 1}}}\right) \times 1 = (1 - 0.197) \times 1 = 802 \text{ mV} \tag{3.72}$$

This is what we can read in Figure 3.25. The third peak is found by letting $n = 3$:

$$V_{peak3} = \left(1 + e^{-\frac{3\pi}{\sqrt{4Q^2 - 1}}}\right) \times 1 = (1 + 87.7m) \times 1 = 1.087 \text{ V} \tag{3.73}$$

This is exactly the displayed value.

It is sometimes interesting to derive the damping ratio (or the quality factor) and the resonant frequency from an observed waveform such as the one presented in Figure 3.27. The best way to obtain this parameter is to measure the attenuation level between two consecutive peaks as those represented in Figure 3.27. From the picture, the ratio $\alpha$ between the two positive overshoots is

$$\alpha = \frac{V(t_0)}{V(t_0 + t_d)} = \frac{e^{-\frac{\pi}{\sqrt{4Q^2 - 1}}}}{e^{-\frac{3\pi}{\sqrt{4Q^2 - 1}}}} \tag{3.74}$$

From the picture, we extracted a value of 2.413. Please make sure you considered only the overshoot; do not forget to remove the final value from your readings

(5 V here). From this equation, if we take the natural logarithm of both sides, we have a definition for $\delta$ (delta) known as the *logarithmic decrement*:

$$\delta = \ln(\alpha) = \ln\left(\frac{e^{-\frac{\pi}{\sqrt{4Q^2-1}}}}{e^{-\frac{3\pi}{\sqrt{4Q^2-1}}}}\right) = \ln\left(e^{-\frac{\pi}{\sqrt{4Q^2-1}}}\right) - \ln\left(e^{-\frac{3\pi}{\sqrt{4Q^2-1}}}\right)$$

$$= \frac{1}{\sqrt{4Q^2-1}}(-\pi + 3\pi) = \frac{2\pi}{\sqrt{4Q^2-1}} \tag{3.75}$$

Otherwise written:

$$\delta = \frac{2\pi}{\sqrt{4Q^2-1}} \tag{3.76}$$

From this value, the quality factor $Q$ is a few lines of algebra away:

$$Q = \sqrt{\frac{4\pi^2 + \delta^2}{4\delta^2}} = \sqrt{\left(\frac{\pi}{\delta}\right)^2 + \frac{1}{4}} \tag{3.77}$$

If you prefer the damping ratio instead, the formula becomes

$$\zeta = \frac{1}{\sqrt{\left(\frac{2\pi}{\delta}\right)^2 + 1}} \tag{3.78}$$

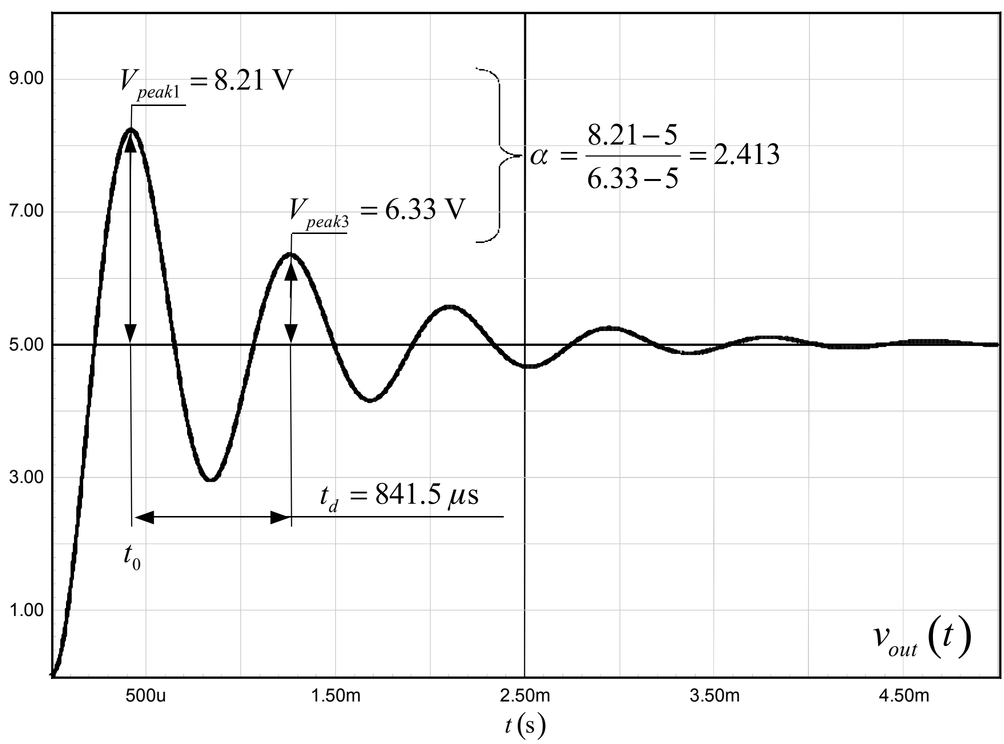



**Figure 3.27** The typical response of a second-order filter to a 5-V step with a quality factor of 3.6.

If we apply the numerical value found in Figure 3.27, we have

$$Q = \sqrt{\left(\frac{3.14}{\ln(2.413)}\right)^2 + 0.25} = 3.6 \tag{3.79}$$

and

$$\zeta = \frac{1}{\sqrt{\left(\frac{6.28}{\ln(2.413)}\right)^2 + 1}} = 0.139 \tag{3.80}$$

The resonant frequency is easily obtained via (3.63) once the time distance between two peaks has been measured. We have 841.5 $\mu$s from the graph

$$f_0 = \frac{1}{t_d\sqrt{1-\zeta^2}} = \frac{1}{841.5u\sqrt{1-0.139^2}} = 1.2 \text{ kHz} \tag{3.81}$$

As the calculation is based on the identification of peaks of two consecutive positive amplitudes, you realize that the parameter extraction can become approximated if not impossible when the response is really damped (e.g., for quality factors below 1). When $Q$ is 0.5, in any case, the two poles are coincident and the overshoot is gone.

We almost have all our timing definitions except the settling time. The settling time $t_s$ represents the time needed for the output to be within its allowable tolerance. Let's say this tolerance is 2 percent of $V_{out}$. According to (3.41), the transient response is a sinusoidal signal whose envelope amplitude is a decaying exponential. This envelope can be put under the following expression:

$$1 \pm e^{-\zeta\omega_0 t} \tag{3.82}$$

At $t = 0$, the exponential term is 1 and the signal starts from 0 as confirmed in Figure 3.25. At $t = t_s$, the envelope must have reduced to a point where it stays within the "tunnel" defined by the allowable tolerance. If we consider a 2 percent tolerance and a 1-V signal, we can write

$$1 \pm e^{-\zeta\omega_0 t} = 1 \pm 0.02 \tag{3.83}$$

which, further to simplification, leads to solving for $t$ when

$$e^{-\zeta\omega_0 t} = 0.02 \tag{3.84}$$

By using logs, we have

$$t_s = -\frac{\ln 0.02}{\omega_0\zeta} \approx \frac{3.9}{\omega_0\zeta} \tag{3.85}$$

If we apply the numerical values of the circuit used to plot Figure 3.25 ($\omega_0$ = 114.4 krad/s):

$$t_s = \frac{3.9}{0.25 \times 114.4k} = 136 \ \mu\text{s} \tag{3.86}$$

**Table 3.1** Summary of Parameter Definitions for an Under-Damped Second-Order System

| *Parameter* | *Definition* | *Value* |
|---|---|---|
| $\omega_d$ | The damped angular frequency in relationship to the natural angular frequency $\omega_0$ | $\omega_d = \omega_0\sqrt{1-\zeta^2}$ |
| $t_{del}$ | Delay time, defined for $v_{out}(t)$ equals 50 percent of the final value | $t_{del} \approx \frac{1+0.7\zeta}{\omega_0}$ |
| $\delta$ | Delta, the logarithmic decrement | $\delta = \ln\left(\frac{x(t_0)}{x(t_0+t_d)}\right) = \frac{2\pi}{\sqrt{4Q^2-1}}$ |
| $t_r$ | Rise time, defined when $v_{out}(t)$ reaches 100 percent of the final value | $t_r = \frac{\pi - \cos^{-1}(\varsigma)}{\omega_0\sqrt{1-\varsigma^2}}$ |
| $t_p$ | Peak time, the time at which $v_{out}(t)$ reaches a maximum | $\frac{\pi}{\omega_0\sqrt{1-\zeta^2}}$ |
| $M_p(\%)$ | Overshoot value | $e^{-\frac{\zeta\pi}{\sqrt{1-\zeta^2}}} \times 100$ |
| $t_s$ | Settling time, the time needed for $v_{out}(t)$ to be within $x$ percent of the final value | $-\frac{\ln x}{\omega_0\zeta}$ |

From these equations, we can see the quality factor (or the damping ratio) affects a lot of parameters. Its choice clearly depends on the final set of specifications you have: is overshoot acceptable, what settling time is acceptable, and so on.

Table 3.1 offers a summary of the variables we have derived for an under-damped system ($Q > 0.5$).

We have seen in Figure 3.6 that the phase margin changes the response of the system, exactly as the $Q$ of the $RLC$ system does. Therefore, there must be a relationship between the phase margin as measured on an *open-loop* system and the quality factor observed after the loop is *closed*. If we can find it, we will be able to pick a phase margin design goal based on the type of transient response we expect.

### 3.3.3 Phase Margin and Quality Factor

Assume a simple converter as the one appearing in Figure 3.28. We can see a plant featuring a transfer function $H(s)$ and a compensator $G(s)$ processing the error signal $\varepsilon$.

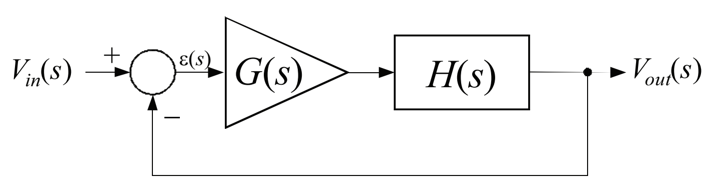



**Figure 3.28** A compensated converter featuring a return path of a unity gain.

The closed-loop transfer function of such a system can be easily derived with a few lines of algebra:

$$V_{out}(s) = \left[V_{in}(s) - V_{out}(s)\right]G(s)H(s) \tag{3.87}$$

As we have seen, the loop gain $G(s)H(s)$ is also noted $T(s)$. Substituting and rearranging (3.87), we have:

$$\frac{V_{out}(s)}{V_{in}(s)} = \frac{T(s)}{1+T(s)} \tag{3.88}$$

This is the closed-loop gain expression of Figure 3.28 unity feedback system. Now, let's assume that its open-loop response $T(s)$ is plotted in Figure 3.29. We know that a control system reacts to incoming perturbations as long as some gain exists at the frequencies of concern. We also know that stability is ensured if we limit the bandwidth by forcing the gain to roll off as the frequency increases. To limit the phase lag at crossover, and thus obtain a good phase margin, we bend the loop gain to cross the 0-dB axis with a –1 slope, implying a single pole response slightly before and after $f_c$. We also learned that at some point, when the phase lag reaches 180°, the loop gain must be sufficiently low to ensure a good gain margin. To speed-up the gain decrease beyond $f_c$ and ensure sufficient gain margin, a second pole is generally installed after the crossover point. If we use a magnifier to observe

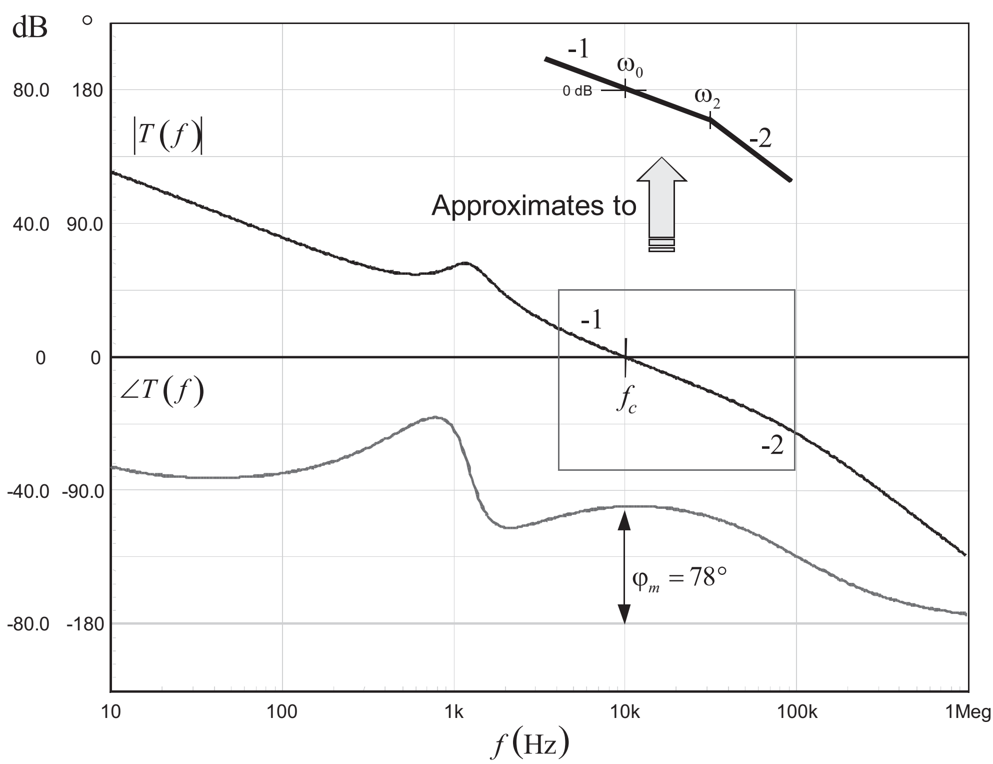



**Figure 3.29** The typical open-loop response of a compensated converter can be approximated to a second-order system in the vicinity of the crossover frequency.

the curve around the crossover frequency (see the framed area in Figure 3.29), we can see a two-pole configuration. This two-pole configuration is constructed with a 0-dB crossover pole, $\omega_0$, and a high-frequency pole, $\omega_2$. The transfer function of these cascaded poles can be put under the following form:

$$T(s) \approx \frac{1}{\left(\frac{s}{\omega_0}\right)\left(1+\frac{s}{\omega_2}\right)} \tag{3.89}$$

In this approximated expression, we clearly do not consider extra poles and zeros away from $f_c$, naturally eliminating their impact on the transfer function. However, our interest lies in the approximate response the converter is going to deliver once its loop is closed. In other terms, let us identify the closed-loop transfer function derived from (3.89) and rearranged according to (3.88):

$$\frac{T(s)}{1+T(s)} = \frac{\frac{1}{\left(\frac{s}{\omega_0}\right)\left(1+\frac{s}{\omega_2}\right)}}{1+\frac{1}{\left(\frac{s}{\omega_0}\right)\left(1+\frac{s}{\omega_2}\right)}} = \frac{1}{\frac{s^2}{\omega_0\omega_2}+\frac{s}{\omega_0}+1} \tag{3.90}$$

The right term of (3.90) looks familiar. It actually recalls the form already derived in (3.28) from which we can establish a relationship with the coefficients of (3.90):

$$\frac{1}{\frac{s^2}{\omega_0\omega_2}+\frac{s}{\omega_0}+1} = \frac{1}{\frac{s^2}{\omega_r^2}+\frac{s}{\omega_r Q_c}+1} \tag{3.91}$$

The identification of the closed-loop quality factor $Q_c$ and the resonant frequency $\omega_r$ is straightforward:

$$Q_c = \sqrt{\frac{\omega_0}{\omega_2}} \tag{3.92}$$

$$\omega_r = \sqrt{\omega_0\omega_2} \tag{3.93}$$

We now have an equation that describes the approximated closed-loop response of our converter and it includes a quality factor $Q_c$. The next step is to derive a relationship between $Q_c$ and the key design parameter, the open-loop phase margin $\varphi_m$. First, based on (3.89), let us calculate the crossover frequency brought by the location of the 0-dB crossover pole $\omega_0$ and its associated high frequency pole $\omega_2$. At the crossover point $f_c$, we know that the magnitude of $T(s)$ equals 1 or 0 dB. Therefore, if we assume a harmonic excitation at the crossover frequency, we can replace $s$ by $j\omega_c$ in (3.89) and write:

$$\left|\frac{1}{\left(\frac{j\omega_c}{\omega_0}\right)\left(1+\frac{j\omega_c}{\omega_2}\right)}\right| = 1 \tag{3.94}$$

Otherwise stated:

$$\left|\left(\frac{j\omega_c}{\omega_0}\right)\left(1+\frac{j\omega_c}{\omega_2}\right)\right| = \left|\left(\frac{j\omega_c}{\omega_0}\right)\right|\left|\left(1+\frac{j\omega_c}{\omega_2}\right)\right| = 1 \tag{3.95}$$

Extracting the individual magnitudes gives us

$$\frac{\omega_c}{\omega_0}\sqrt{1+\left(\frac{\omega_c}{\omega_2}\right)^2} = 1 \tag{3.96}$$

To get rid of the square root, we square both sides of the expression:

$$\left(\frac{\omega_c}{\omega_0}\right)^2\left(1+\left(\frac{\omega_c}{\omega_2}\right)^2\right) = 1 \tag{3.97}$$

From (3.92), we extract $\omega_0$ and plug it in (3.97):

$$\frac{\omega_c^2\left(\omega_2^2+\omega_c^2\right)}{\left(Q_c\omega_2\right)^4} = 1 \tag{3.98}$$

Solving for $\omega_c$, we obtain

$$\omega_c = \frac{\omega_2\sqrt{\left(\sqrt{1+4Q_c^4}-1\right)}}{\sqrt{2}} \tag{3.99}$$

Equation (3.99) shows us how the closed-loop quality factor and the open-loop crossover frequency are linked. It is important for this remark to be well understood: $Q_c$ represents the resulting *closed-loop* response quality factor based on the pole/zero arrangement describing the approximated *open-loop* compensated transfer function $T(s)$ near the crossover frequency in (3.89).

To continue further with our analysis, we can evaluate the argument of $T(s)$ defined in (3.89) at the crossover point $\omega_c$:

$$\arg T(\omega_c) = -\left(\tan^{-1}\frac{\omega_c/\omega_0}{0}+\tan^{-1}\frac{\omega_c}{\omega_2}\right) = -\frac{\pi}{2}-\tan^{-1}\frac{\omega_c}{\omega_2} \tag{3.100}$$

Looking at Figure 3.7, the phase margin $\varphi_m$ represents the distance between the total phase lag at crossover and the –180° limit. We can write

$$\arg T(\omega_c) = -180+\varphi_m \tag{3.101}$$

Rearranging and introducing $\pi$ rather than 180°, we obtain

$$\varphi_m = \pi+\arg T(\omega_c) \tag{3.102}$$

Substituting (3.100) in (3.102), we obtain

$$\varphi_m = \pi-\tan^{-1}\frac{\omega_c}{\omega_2}-\frac{\pi}{2} = \frac{\pi}{2}-\tan^{-1}\frac{\omega_c}{\omega_2} \tag{3.103}$$

Remembering our (far away) trigonometric classes, we know that

$$\tan^{-1} x + \tan^{-1} \frac{1}{x} = \frac{\pi}{2} \tag{3.104}$$

Replacing $\pi/2$ in (3.103) by its expression in (3.104), we obtain

$$\varphi_m = \tan^{-1} \frac{\omega_c}{\omega_2} + \tan^{-1} \frac{\omega_2}{\omega_c} - \tan^{-1} \frac{\omega_c}{\omega_2} \tag{3.105}$$

Collecting the remaining terms, we finally obtain a simpler definition based on the position of the second pole in relationship to the crossover frequency:

$$\varphi_m = \tan^{-1} \frac{\omega_2}{\omega_c} \tag{3.106}$$

We have already defined the crossover angular frequency versus the closed loop quality factor in (3.99). If we capitalize on this definition in (3.106) we have

$$\varphi_m = \tan^{-1}\left(\sqrt{\frac{2}{\sqrt{\left(1+4Q_c^4\right)}-1}}\right) \tag{3.107}$$

The next step is to extract the closed-loop quality factor from (3.107) and simplify the result:

$$Q_c = \frac{\sqrt[4]{1+\tan^2(\varphi_m)}}{\tan(\varphi_m)} \tag{3.108}$$

Again, looking back at what we learned at school, we remember that

$$1+\tan^2 \varphi_m = \frac{1}{\cos^2 \varphi_m} \tag{3.109}$$

Replacing the expression under the root in (3.108), we have

$$Q_c = \frac{\sqrt[4]{\frac{1}{\cos^2(\varphi_m)}}}{\tan(\varphi_m)} = \frac{1}{\sqrt{\cos\varphi_m}} \frac{\cos\varphi_m}{\sin\varphi_m} = \frac{\sqrt{\cos\varphi_m}}{\sin\varphi_m} \tag{3.110}$$

This expression also works backward if you already have measured the closed-loop quality factor:

$$\varphi_m = \cos^{-1}\left(\frac{\sqrt{4Q_c^4+1}-1}{2Q_c^2}\right) \tag{3.111}$$

This is it! We now have a relationship between our main design criterion, the open-loop phase margin $\varphi_m$, and the closed-loop quality factor $Q_c$. Figure 3.30

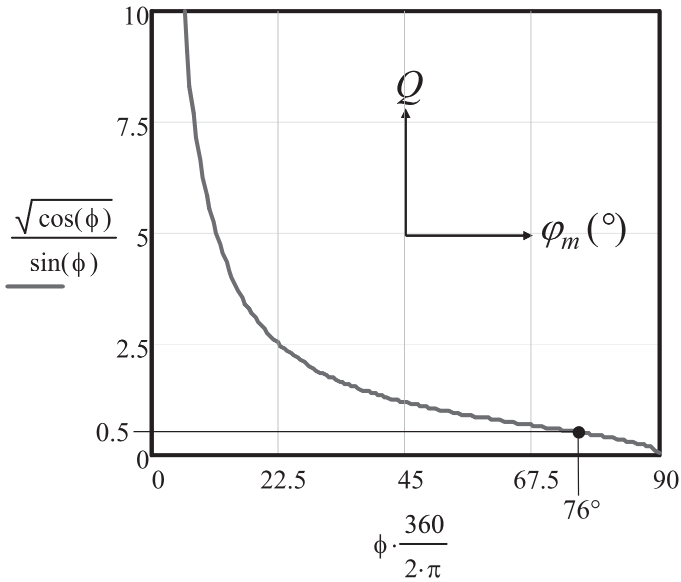



**Figure 3.30** The graph shows the evolution of the closed-loop quality factor $Q$ as you select different phase margins $\varphi_m$.

graphs the quality factor your closed-loop system will exhibit depending on the phase margin you have selected at the design stage.

We have drawn the ac closed-loop response of a second-order system compensated to give different open-loop phase margins. The result appears in Figure 3.31 and clearly show the peaking as with a *RLC* filter.

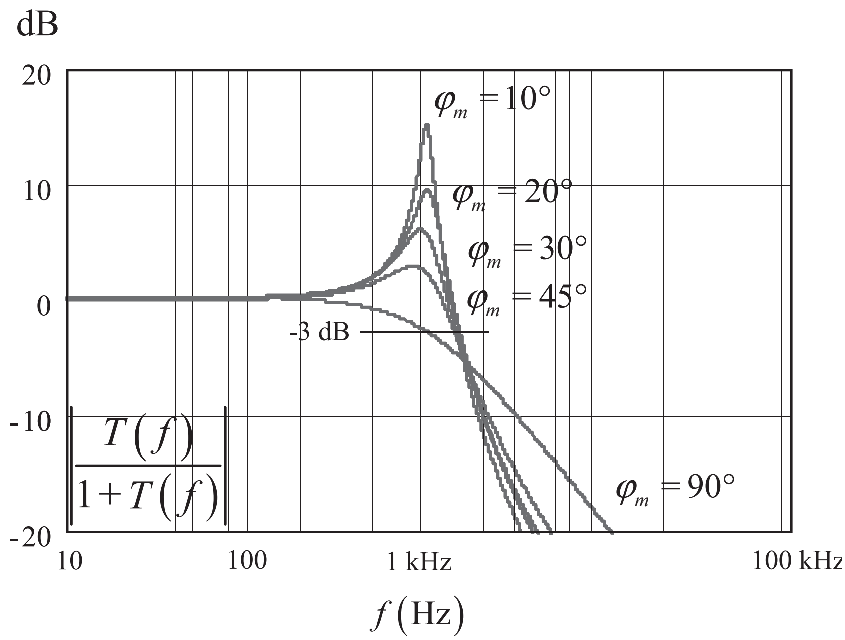



**Figure 3.31** By selecting different open-loop phase margin values, we modify the ac response of the closed-loop system.

The peaking amplitude noted $M_m$ can also be analyzed. Going back to the transfer function of a second-order system as defined by (3.28), we first express its magnitude by replacing $s$ with $j\omega$:

$$T(j\omega) = \frac{1}{\frac{(j\omega)^2}{\omega_0^2} + 2\zeta\frac{j\omega}{\omega_0} + 1} = \frac{1}{1 - \left(\frac{\omega}{\omega_0}\right)^2 + j2\zeta\frac{\omega}{\omega_0}} \tag{3.112}$$

The magnitude can be extracted as follows:

$$\left|T(j\omega)\right| = \frac{1}{\sqrt{\left[1 - \left(\frac{\omega}{\omega_0}\right)^2\right]^2 + \left(2\zeta\frac{\omega}{\omega_0}\right)^2}} \tag{3.113}$$

To obtain the frequency at which the expression peaks, the resonant frequency, we derive the previous function and solve for $\omega$, which makes the result go to zero:

$$\frac{d\left|T(j\omega)\right|}{d\omega} = 0 \tag{3.114}$$

Solving for the equation leads to

$$\omega_m = \omega_0\sqrt{1 - 2\zeta^2} \tag{3.115}$$

defined for $\zeta < \frac{1}{\sqrt{2}}$.

The value of the overshoot itself is obtained by plugging (3.115) into (3.113). Once done, we obtain

$$M_m = \frac{1}{2\zeta\sqrt{1 - \zeta^2}} \tag{3.116}$$

If you want to combine speed and lack of overshoot, Figure 3.23 suggests a $Q$ of 0.5 where both poles are coincident, without imaginary contributions. Reading the corresponding phase margin in Figure 3.30, we can see that a recommended value of 76° satisfies this request for such a $Q$. We are far away from the 45° target found in the majority of textbooks! What does it mean then? It means that you select the phase margin based on the transient response you need. In the response to a load step, once the loop is closed, the open-loop phase margin mostly affects the recovery speed and a little the undershoot depth. If a fast recovery is needed and some overshoot accepted, then reducing the phase margin can be an option. On the contrary, if absolutely no overshoot is tolerated, you have no other choice than to increase the phase margin to the detriment of the recovery speed. Whatever solution you select, you must ensure that, despite changing operating conditions such as input/output voltage or load, ambient temperature, and production spreads (equivalent series resistors of capacitors, for instance), the phase margin never goes below 45°: the obtained ringing would be unacceptable. Some customers like military or space agencies ask for a minimum phase margin of 90° and must be backed by Monte Carlo analysis to prove that dispersions are under control. For the general case, aiming for a typical value around 70° should become a good design practice.

### 3.3.4 Opening the Loop to Measure the Phase Margin

The ac response of a closed-loop system can be deduced from its transient response but it is complex to extract parameters such as crossover or phase margin. To study a closed-loop system, we must physically open the loop at some appropriate point to access the transfer functions of interest: $H(s)$, the plant transfer function, $G(s)$ the compensator we designed, and $T(s)$ the open-loop gain. The first transfer function, $H(s)$, is the starting point of control loop study: we must know the frequency response of the system we want to stabilize. The compensator block $G$ is then tailored based on the collected data and the desired response: we want to force the loop gain crossover at the selected frequency while ensuring enough phase margin at this salient point. The study of the loop gain on the hardware prototype will eventually tell us if $G$ was well designed, revealing the obtained crossover frequency and the associated phase margin. As we will later see, the study of the whole system can also be undertaken analytically or by using a SPICE simulator.

A typical loop opening appears in Figure 3.32: this is the classical example found in textbooks. A signal is injected in the left input while its effects are observed on outputs A or B depending on what you want to plot, $H(s)$, $G(s)$, or $T(s)$. The drawing represents a control system where the output must follow the input signal. We have seen in Chapter 1 that other types of control systems exist and are called regulators. These systems work with a fixed control input and ensure a constant

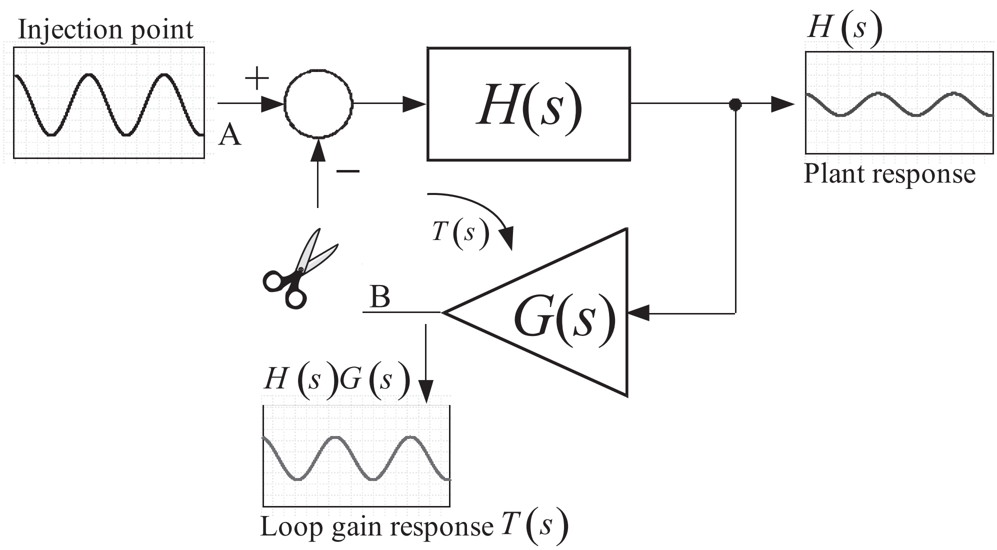



**Figure 3.32** The loop is opened just prior the inverted input on the adder. Please note that the negative sign is not included in the path under study.

output variable regardless of the operating conditions. A dc-to-dc converter is a regulator: it delivers a constant voltage even if the input level changes or if the current drawn by the load varies.

In such a regulator, a reference source, often noted $V_{ref}$, imposes the setpoint while the system strives to ensure a constant output (usually a multiple of $V_{ref}$) whatever the perturbations. The perturbations are the input voltage changes, the delivered current or the temperature if it needs to be accounted for. In such an arrangement, the loop gain $T(s)$ is given by

$$T(s) = H(s)G(s) \tag{3.117}$$

As you can see, the reversal sign brought by the inverting input is not included in the formula since we open the loop at point B. The system becomes unstable when the signal injected in A becomes lagged by 180° in B. The phase margin is thus measured by reference to the –180° distance and the condition for instability is noted:

$$\angle T(j\omega) = -180° \tag{3.118}$$

To reflect an implementation closer to reality, the Figure 3.32 sketch must be updated by what is proposed in Figure 3.33.

In this drawing, the varying setpoint is replaced by a fixed reference voltage $V_{ref}$. The output is subtracted to this voltage level and the error voltage $\varepsilon$ enters the compensation block $G$. It then further controls the power $H$ to deliver the right output level. As explained, this level can be affected by perturbations such as the input voltage and the output current. In this representation, these perturbations are modeled as sources that subtract from the plant output. We will later see, in particular for the output impedance, that different arrangements can be made. In

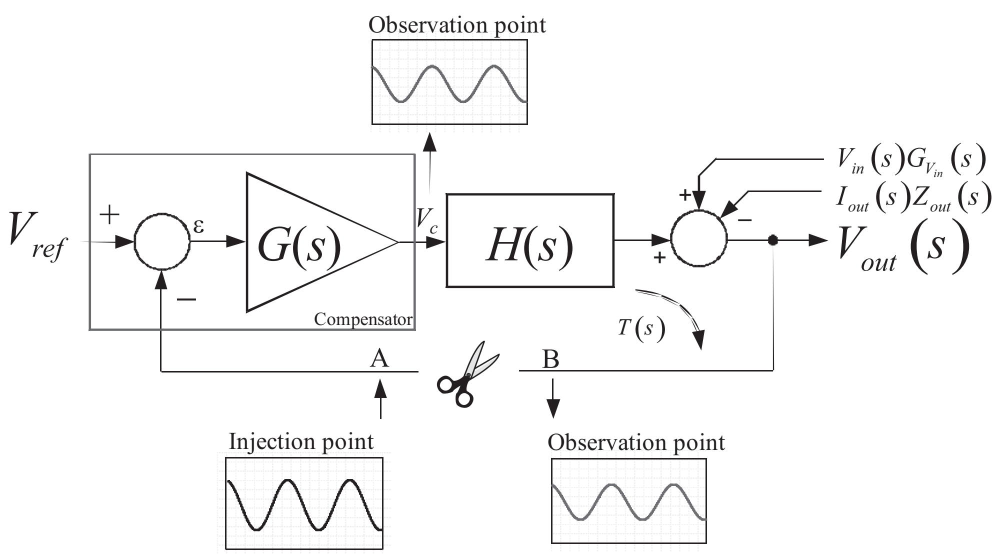



**Figure 3.33** A regulator maintains a constant output regardless of external perturbations. The opening of the loop can take place in the return path, and several signals can be observed.

the example, the return path is broken and used to inject the modulating signal. If we inject in A, we can observe $V_c$ and check that the transfer function of $G$ is the one we were looking for:

$$G(s) = \frac{V_c(s)}{V_A(s)} \tag{3.119}$$

Now, if we inject in A but observe $V_c$ and $V_{out}$, we have the plant transfer function $H(s)$:

$$H(s) = \frac{V_{out}(s)}{V_c(s)} \tag{3.120}$$

Finally, if we inject in A and observe B, we have the complete open-loop gain, $T(s)$:

$$T(s) = \frac{V_B(s)}{V_A(s)} \tag{3.121}$$

In this particular configuration, the reversal action of the compensator is now included and the loop gain definition becomes

$$T(s) = -H(s)G(s) \tag{3.122}$$

The instability condition is now that the signal injected in A returns in phase in B (with equal amplitude, of course). This complete lag implies an instability condition redefined as

$$\angle T(j\omega) = -360° \text{ or } 0° \tag{3.123}$$

The phase margin is thus measured by reference to –360° or 0°. For this reason, all plots in Figure 3.34 refer to a similar phase margin. Network analyzers or SPICE graphical tools usually display the open-loop phase as in the upper graph.

It is important to note that this ac analysis makes sense only if the perturbation such as the input voltage or the input current is kept constant during the ac sweep. You select them to fix the operating point at which the converter must be analyzed, and they no longer change during the harmonic sweep.

In the given examples, the loop has been physically open, meaning the operating point of the closed-loop system is lost. In order to put the system in conditions representative of its closed-loop working mode, you will need to recreate the operating point during the ac sweep. If the converter is compensated to deliver 19 V/3 A while the input voltage is 100 V, then physically opening the loop implies the addition of an external bias keeping a 19 V output while it delivers 3 A. If it can sometimes work to extract the plant transfer function of a given converter, it is almost impossible for the open-loop gain. As the dc gain of the loop is purposely made extremely high, a very small variation in the dc signal (e.g., noise, thermal drift of the source) causes the output to jump into its lower or upper bounds. When manipulating high-power systems, this is not something you want to undertake. Fortunately, alternative methods exist and authorize open-loop measurements without opening the loop. We will come back to the technique when studying practical cases in Chapter 9.

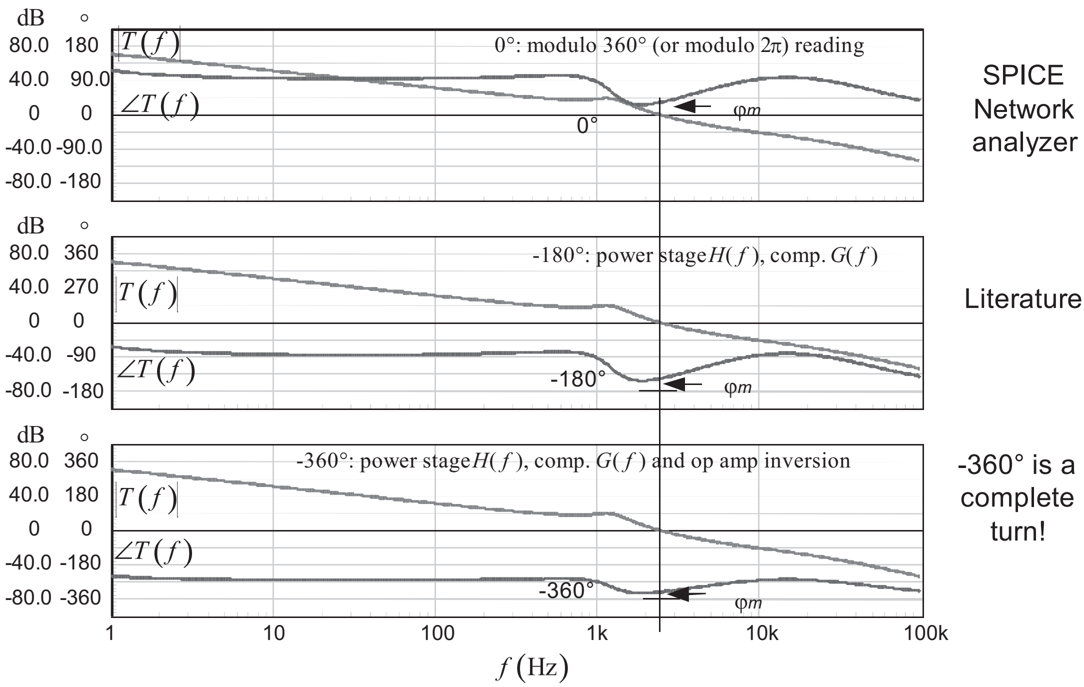



**Figure 3.34** All these curves refer to the same phase margin value.

### 3.3.5 The Phase Margin of a Switching Converter

Further to these theoretical derivations, it is time to look at a working example. Figure 3.35 shows an average model described in [1], wired in a voltage-mode buck configuration. The compensation elements around the operational amplifier are automatically evaluated using the k-factor technique, a method that will be described in an upcoming chapter. Thanks to this automated template, we can easily select the phase margin of our choice at a constant crossover value (10 kHz) and check for the transient response to a sudden load change.

The output is subjected to a current step ranging from 1 A to 2 A in 1 $\mu$s. The results appear in Figure 3.36. The 76° phase margin gives a little overshoot of 0.05 percent whereas the 49° margin triples that overshoot, still reasonable though, given the vertical axis scale of 20 mV per division. However, you can observe a faster recovery in the 49° phase case (70 $\mu$s) versus the 76° case (227 $\mu$s). Why do we still have overshoot with 76° when theory states there should be none? It is because (3.89) is a simplified view of the transfer function, captured only in the vicinity of the crossover frequency, with a two-pole configuration, without zeros. If you have extra zeros in the transfer function or a lower open-loop gain, the $Q$ factor approximation we have been through does not work anymore and extra work is required as detailed in [2]. So what is the interest of the derivation we just went through? It is to analytically show how the phase margin in the open-loop transfer function affects the transient response once the loop is closed. Its choice is capital to meet the specifications. As confirmed by Figure 3.36, a small phase margin leads to a peaky closed-loop response and a large phase margin implies a sluggish but nonovershooting response. As usual, you will need to trade one parameter versus

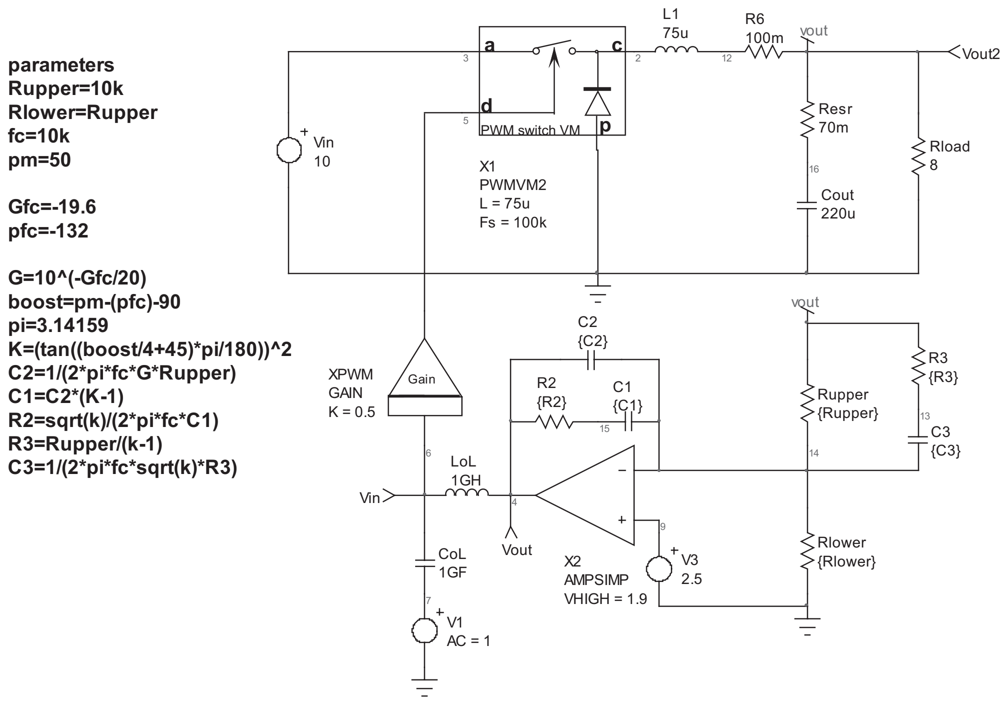



**Figure 3.35** A buck converter operated in voltage mode is used to illustrate the impact of the phase margin on the transient response.

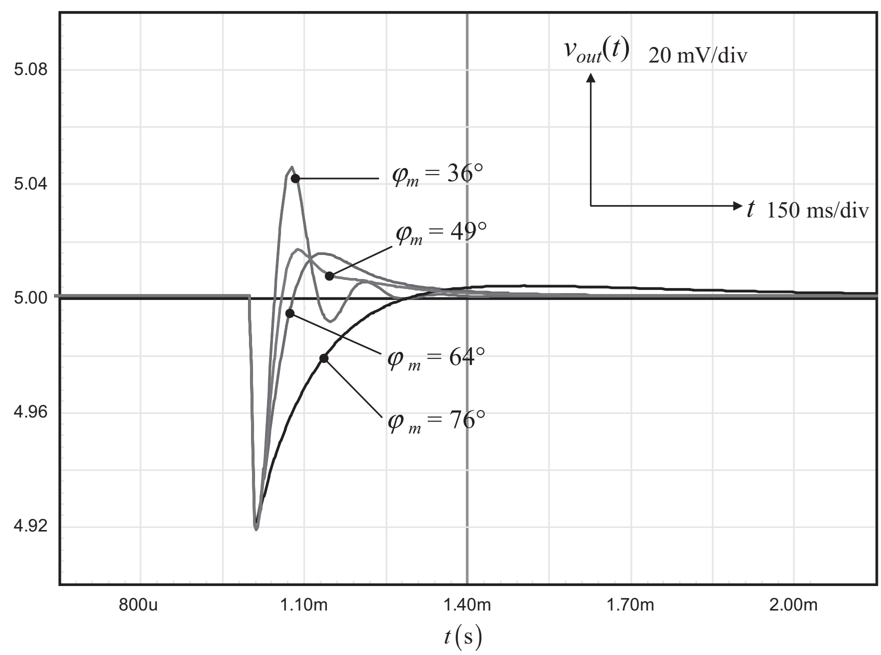



**Figure 3.36** The phase margin has been adjusted at different values—$f_c$ is constant—and it clearly affects the transient response in both the recovery time and the overshoot above the 5-V target.

the other one: a sluggish but nonringing response or a fast settling time with some overshoot?

### 3.3.6 Considering a Delay in the Conversion Process

Phase margin is an important parameter in the design of control systems. Not only because it must be selected based on the desired transient response as we just explained, but also because the transfer function ac response will change with time and production spreads. Changes can be imputed to temperature drifts, aging, production shifts, or simply a component replacements. It is your responsibility as a design engineer that these variations do not jeopardize the stability of the converter. In other words, you must give the total loop lag some freedom to safely move within a certain range, hence the term margin.

A loop is made of various components, regardless of whether they belong to the compensator $G$, the plant $H$, or the sensors. Some components are passive and can undergo the deterioration/variation of some of their stray elements. The equivalent series resistor of a capacitor is a typical example, as it creates a zero in the transfer function that affects the small-signal response of the converter. As the ESR moves, the zero position will be affected by temperature or lots-to-lots variations. Therefore, provisions must be made to shield the control loop against these variations. Some other blocks are active like comparators, analog-to-digital converters, logic gates, and so on. A simple example of such an active function is the pulse width modulator (PWM) or the PWM block in a switching converter. Its purpose is to drive the power switch conduction time in relationship to the power demand. A typical circuit implementation appears in Figure 3.37.

The circuit works by comparing the error voltage $v_{err}(t)$ to a ramp of a period $T_{sw}$. When the error voltage (pin +) is above the ramp signal (pin –), the comparator is in a high state and drives the switch on. This high state is called the on-time

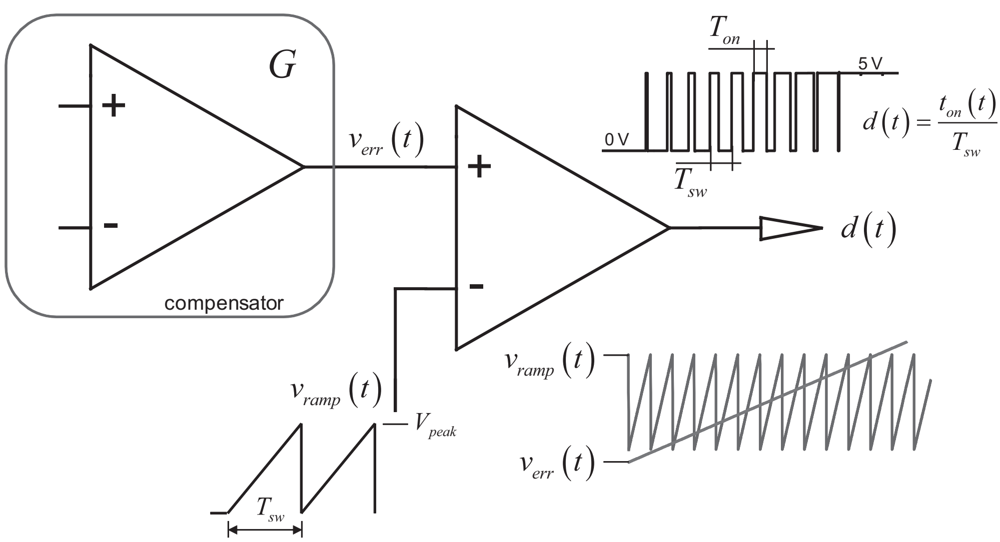



**Figure 3.37** A pulse-width modulator uses a simple comparator that toggles every time the error signal crosses an artificial ramp.

(noted $T_{on}$) and lasts until the ramp signal exceeds the error voltage, where the comparator returns to its low state. The duration $T_{on}$ over the switching period $T_{sw}$ is called the duty ratio and is noted $D$:

$$D = \frac{T_{on}}{T_{sw}} \tag{3.124}$$

As $T_{on}$ changes over time, the duty ratio can also be expressed as an instantaneous variable:

$$d(t) = \frac{t_{on}(t)}{T_{sw}} \tag{3.125}$$

By controlling the duty ratio via the error voltage, the control system has a means to adjust the power transfer of the converter. In order to perform ac analysis, this modulating block must be modeled to study the loop gain of our converter, $V_c(s)$ to $V_{out}(s)$. The time-domain equation is simple: the transition occurs when the sawtooth and the error levels meet. The sawtooth equation is that of a line going up to $V_{peak}$ in a time duration of $T_{sw}$:

$$v_{ramp}(t) = V_{peak}\frac{t}{T_{sw}} \tag{3.126}$$

When $t$ equals $T_{on}$, $v_{err}$ has crossed the sawtooth level, hence

$$v_{err}(t) = V_{peak}\frac{T_{on}(t)}{T_{sw}} = d(t)V_{peak} \tag{3.127}$$

where $T_{on}(t)$ is the on-time value at the instant $v_{err}$(t) equals the sawtooth value. Rearranging gives us

$$d(t) = \frac{v_{err}(t)}{V_{peak}} \tag{3.128}$$

If we average this equation over a switching period, we obtain a large-signal equation:

$$D(V_{err}) = \frac{V_{err}}{V_{peak}} \tag{3.129}$$

The ac transfer function of this equation is simply obtained by a differentiation: you will differentiate $D(V_{err})$ with respect to $V_{err}$ by checking its sensitivity to $V_{err}$. This sensitivity is actually the small-signal gain:

$$\hat{d} = \frac{dD(V_{err})}{dV_{err}}\hat{v}_{err} = \frac{1}{V_{peak}}\hat{v}_{err} \tag{3.130}$$

We obtain the small-signal gain of the PWM block noted $K_{PWM}$:

$$K_{PWM} = \frac{1}{V_{peak}} \tag{3.131}$$

Suppose the peak value of the sawtooth is 2 V. Then the small-signal gain of the modulator is 0.5 or –6 dB. When studying the transfer function of a power stage (a buck, for instance), you will have to add the pulse-width modulator small-signal

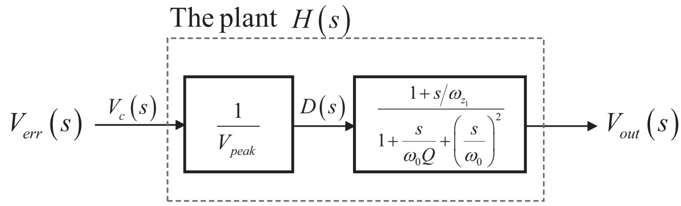



**Figure 3.38** The plant transfer function $H(s)$ also includes the PWM modulator gain.

gain to the control input to account for its own ac transfer function. This is what has been done in Figure 3.35 with the insertion of the subcircuit XPWM. If we want to model the power stage alone, including the modulator gain, we obtain the drawing shown in Figure 3.38.

As we all know, the comparator is affected by a response time. When a voltage difference exists between both inputs, the decision to actually toggle the output requires a processing time: charging parasitic capacitances, switching branches, and so on. For a comparator like the LM311, it can be as high as 250 ns. Practically, it means that the compensator instructs a change in the duty ratio because the operating conditions impose it, but this change will only occur 250 ns later; this is a *delay*, also called a *transport* or *propagation* delay, in the control loop. Intuitively, if a delay appears in slow-bandwidth system (e.g., 1 kHz), 250 ns of delay is a relatively small contribution to the control chain. Expand the bandwidth to 100 kHz (very common in miniature high-frequency dc-dc converters for wireless applications), and these 250 ns start to trouble the whole control loop.

Figure 3.39 depicts a signal $u(t)$ entering a block affected by a delay $\tau$.

The output signal $y(t)$ appears after a delay of $\tau$ seconds has elapsed. When you observe the output signal $y(t)$, you can say that what you see is actually a signal that occurred $\tau$ seconds before (i.e., at the instant $t_2$-$\tau$ on the picture). Mathematically, this observation can be written as follows:

$$y(t) = u(t - \tau) \tag{3.132}$$

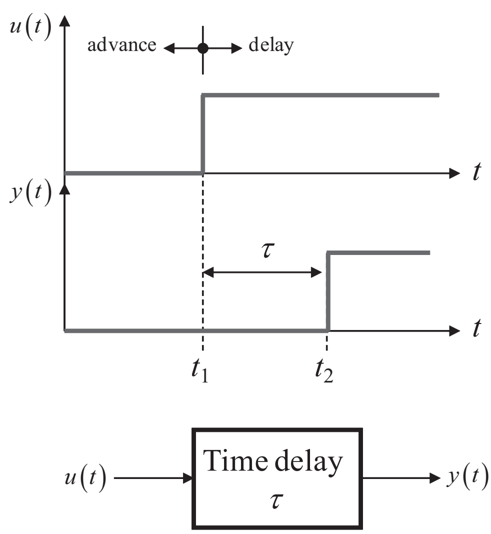



**Figure 3.39** A delay can be represented by a time-domain shift to the input signal.

Now, to include this delay block in the Laplace domain and thus be able to run small-signal analysis, we will first derive the Laplace transform of (3.132):

$$\mathcal{L}\left[y(t)\right] = \mathcal{L}\left[u(t-\tau)\right] \tag{3.133}$$

At first glance, we do not know the result of this equation, but we can think of a different approach to solve it. Assume $u(t)$ is a sinusoidal waveform with an amplitude $A$. Thanks to Euler, we can express this signal as

$$u(t) = Ae^{j\omega t} \tag{3.134}$$

Once this signal passes through the delay block, its amplitude is unaffected, but it becomes delayed as (3.132) describes. The term $t$ in (3.134) simply becomes $t$-$\tau$:

$$y(t) = Ae^{j\omega(t-\tau)} = Ae^{j\omega t - j\omega\tau} \tag{3.135}$$

This is of the form $e^a e^b = e^{a+b}$. Otherwise stated:

$$y(t) = Ae^{j\omega(t-\tau)} = Ae^{j\omega t}e^{-j\omega\tau} \tag{3.136}$$

The first term is $u(t)$, and the second term is the delay affecting the input signal. If we take the Laplace transform of the previous equation, we obtain

$$Y(s) = U(s)e^{-s\tau} \tag{3.137}$$

The transfer function of the time delay block is thus

$$\frac{Y(s)}{U(s)} = e^{-s\tau} \tag{3.138}$$

Going back to the frequency domain, the ac response of the delay block is

$$\frac{Y(j\omega)}{U(j\omega)} = e^{-j\omega\tau} = e^{j\varphi} \tag{3.139}$$

The form $Ae^{j\varphi}$ is the Euler expression of a sinusoidal waveform under the form of

$$Ae^{j\varphi} = Ae^{j\omega t} = A(\cos\omega t + j\sin\omega t) \tag{3.140}$$

Because of the negative sign in the exponent, $\varphi = -\omega\tau$. Therefore,

$$\arg e^{-j\omega\tau} = \tan^{-1}\left(\frac{\sin(-\omega\tau)}{\cos(\omega\tau)}\right) = -\omega\tau \tag{3.141}$$

and

$$\left|e^{-j\omega\tau}\right| = \sqrt{\cos^2\omega\tau + \sin^2\omega\tau} = 1 \tag{3.142}$$

To check the frequency response of the block, we can implement a SPICE simulation using an arrangement suggested in [3]. The setup appears in Figure 3.40 and shows a buffered delay line, loaded by its characteristic impedance (50 Ω in our example).

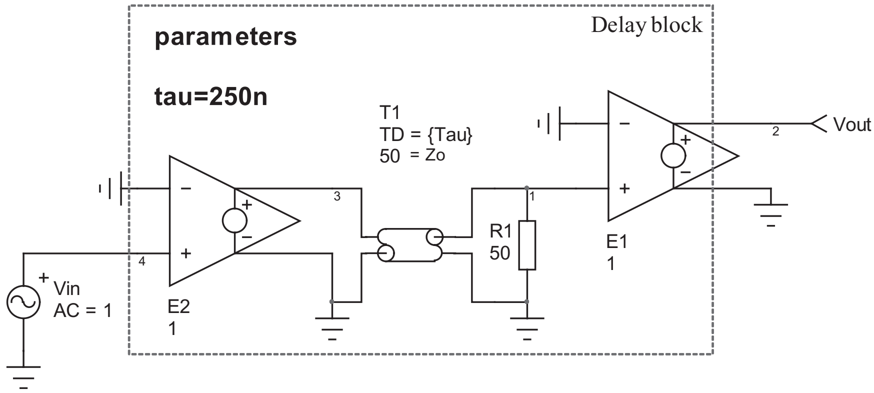



**Figure 3.40** A simple delay line helps to model a delay that appears in a power converter.

The ac response of such a subcircuit with a 250-ns delay is given in Figure 3.41. The magnitude is 1 over the whole frequency sweep, but the phase lags as we go down the horizontal axis. At 1 kHz, the phase lag contributed by the delay is negligible. At 100 kHz, (3.141) predicts a phase lag of

$$\varphi = -\omega\tau = -100k \times 6.28 \times 250n \times \frac{180}{\pi} = -9° \tag{3.143}$$

This is what the graph reads, while it goes down to –18° at 200 kHz. A 250-ns delay does not seem much, but if we consider the whole transmission chain, it can

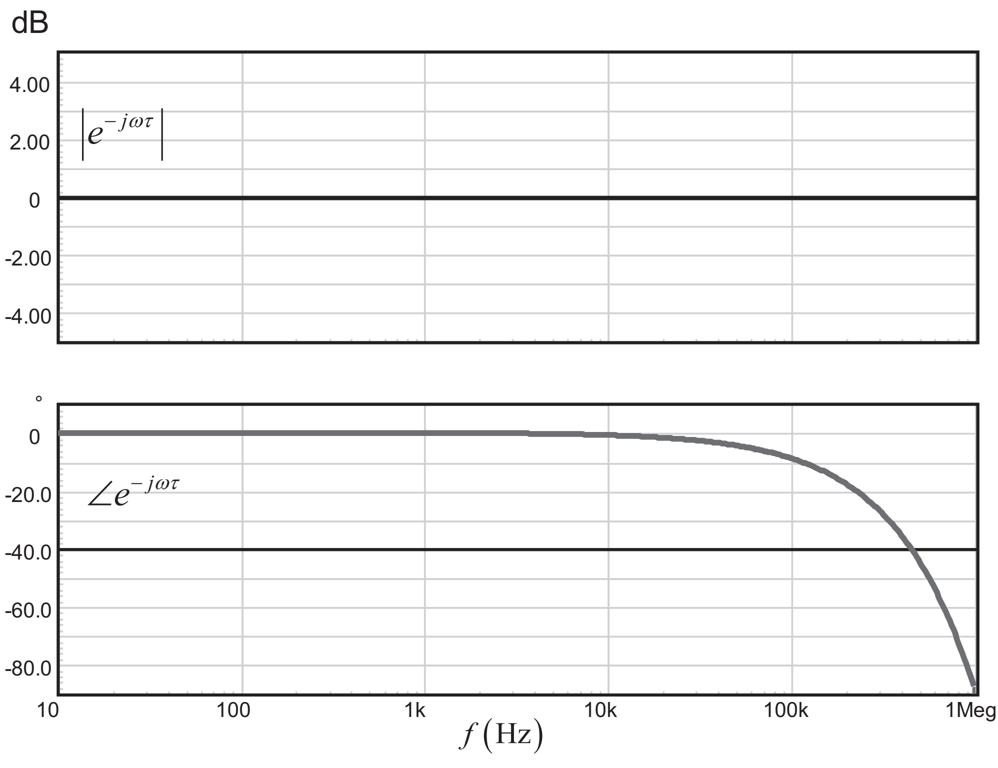



**Figure 3.41** As expected, the magnitude is 0 dB all over the spectrum, but the phase lags as the frequency increases.

be much longer. The delay does not only depend on the PWM block but also on the internal logic arrangement, including the way you turn off the power switch. The logic propagation delay can be very small, but if you slowly turn off the MOSFET for EMI considerations, this is another delay. A total transport delay of 300–400 ns is thus not uncommon at all.

### 3.3.7 The Delay in the Laplace Domain

Now that we know a delay exists in the transmission chain, we can update Figure 3.38 representation to make it appear in Figure 3.42:

The new plant transfer function for our voltage-mode buck converter can thus be expressed as

$$H(s) = \frac{e^{-s\tau}}{V_{peak}} \frac{1 + s/\omega_{z_1}}{1 + \dfrac{s}{\omega_0 Q} + \left(\dfrac{s}{\omega_0}\right)^2} \tag{3.144}$$

The question now is how to run pole/zero analysis or root locus tests with the extra term $e^{-s\tau}$? We clearly need to replace this expression with a pole/zero combination capable of reproducing the Figure 3.41 ac response. What do we see there? A phase lag increasing as we go down the $x$-axis. What transfer function brings phase lag as frequency increases? A pole, indeed. Unfortunately, the magnitude of a pole is not flat over frequency. To counteract the magnitude decrease versus frequency, why not including a zero then? A zero would certainly cancel the magnitude decrease of the pole but would also cancel its phase lag. Unless we purposely insert a RHP zero! This zero will have the same magnitude as a left half plane zero, but the phase will lag, cumulating with that of the pole. Amplitude will be flat, but not the phase. The simplified equality could thus look as follows:

$$e^{-st} \approx \frac{1 - s/\omega_\tau}{1 + s/\omega_\tau} \tag{3.145}$$

What value shall we assign to $\omega_\tau$ to match the delay block? As both left- and right-side expressions arguments must be equal, we can write

$$\arg(e^{-s\tau}) = \arg\left(\frac{1 - s/\omega_\tau}{1 + s/\omega_\tau}\right) \tag{3.146}$$

According to (3.141), we can further write

$$-\omega\tau = \arg(1 - s/\omega_\tau) - \arg(1 + s/\omega_\tau) \tag{3.147}$$

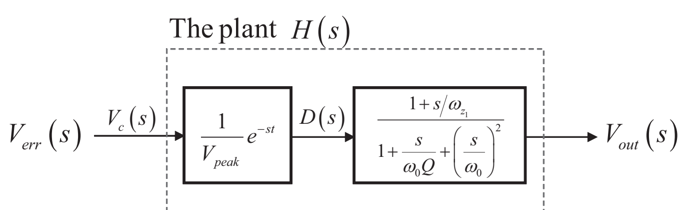



**Figure 3.42** The PWM delay now appears in the whole transmission chain.

Replacing $s$ by $j\omega$ and applying complex number formulas:

$$-\omega\tau = \tan^{-1}\left(-\frac{\omega}{\omega_\tau}\right) - \tan^{-1}\left(\frac{\omega}{\omega_\tau}\right) = -2\tan^{-1}\left(\frac{\omega}{\omega_\tau}\right) \tag{3.148}$$

The Taylor series of $\tan^{-1}(x)$ is $x - \frac{x^3}{3} + \frac{x^5}{5} + \ldots$. Applied to the right side of (3.148), we have

$$-\omega\tau \approx -2\left[\frac{\omega}{\omega_\tau} - \frac{\left(\frac{\omega}{\omega_\tau}\right)^3}{2} + \frac{\left(\frac{\omega}{\omega_\tau}\right)^5}{5}\right] \tag{3.149}$$

If we consider that $\omega_\tau \gg \omega$ along our ac analysis, then all the terms in cube and above can be neglected. The above equation simplifies to

$$\omega\tau \approx 2\frac{\omega}{\omega_\tau} \tag{3.150}$$

From which we can easily extract $\omega_\tau$:

$$\omega_\tau = \frac{2}{\tau} \tag{3.151}$$

Equation (3.145) can now be updated as

$$e^{-s\tau} \approx \frac{1 - \frac{s\tau}{2}}{1 + \frac{s\tau}{2}} \tag{3.152}$$

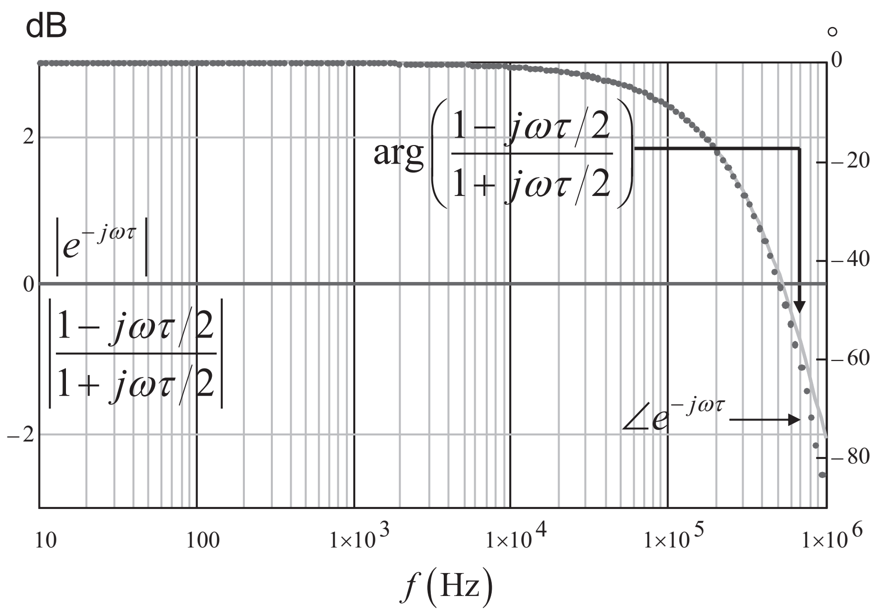



**Figure 3.43** The ac response of the exponential expression versus its simplified version is very good. $\tau$ is 250 ns in this example.

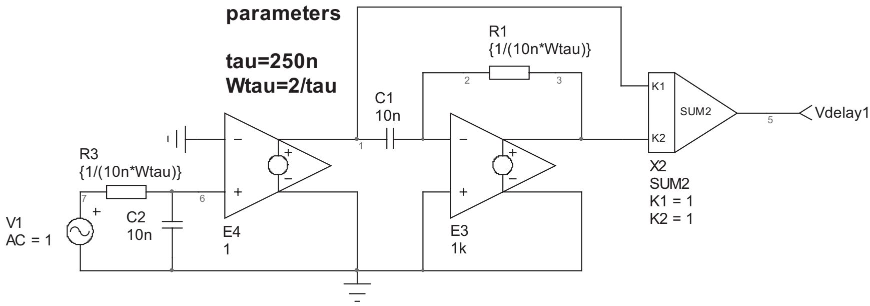



**Figure 3.44** An op amp realizing the RHPZ function to which a pole is added makes a simple and efficient delay block.

This expression is nothing more that the first-order Padé approximation for the exponential

$$e^x \approx \frac{1+\dfrac{x}{2}}{1-\dfrac{x}{2}} \tag{3.153}$$

Given the approximation that we made, $\omega_\tau \gg \omega$, the simplified formula will deviate from the original expression as the frequency increases. If the deviation is too wide, a higher-order Padé expression must be selected. However, the discussion goes well beyond the scope of this book. For those interested by the subject, a comprehensive coverage of the subject appears in [4].

As shown in Figure 3.40, the delay could be built in SPICE using a delay line. A delay line in SPICE usually increases the computational time and can sometimes

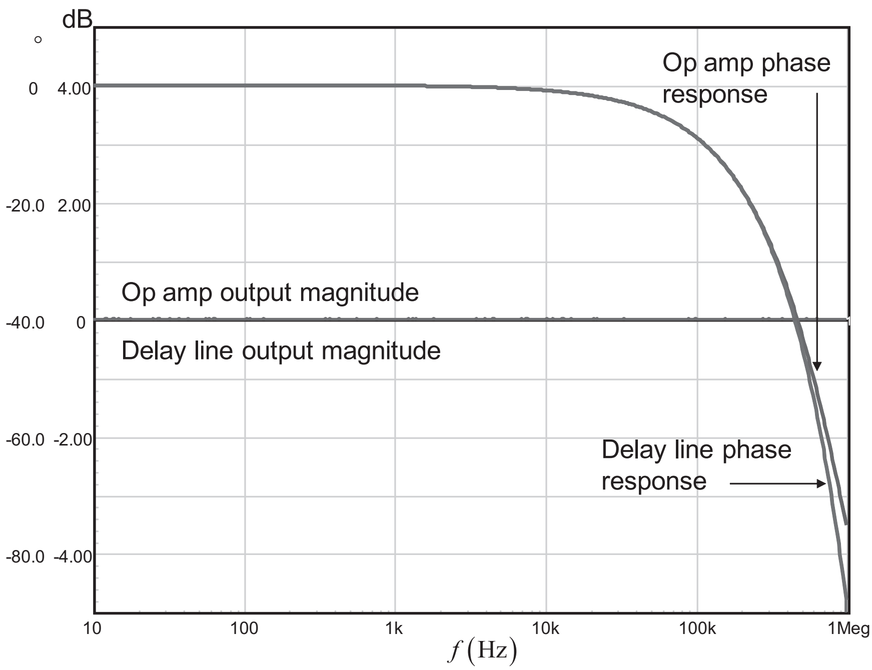



**Figure 3.45** The ac response of the op amp-based delay agrees quite well with that of the delay line. It still remains a first-order model though.

lead to convergence issues. Also, some simple simulators do not include delay lines in their primitives list. No problem, Chapter 2 taught us how to build a RHP zero with an op amp and an adder. If we add a simple pole coincident with the zero, we have our delay!

The ac response of the block appears in Figure 3.45 for a 250-ns transport delay. It is in excellent agreement with the response plotted in Figure 3.43 but still remains a first-order approximation.

### 3.3.8 Delay Margin versus Phase Margin

The definition we gave for the delay block shows that its insertion in the loop gain equation does not affect its magnitude, only its phase. For stability analysis in a unity feedback system, the classical characteristic equation $1+T(s)=0$ must be updated to account for the delay block presence:

$$\chi(s) = 1 + e^{-s\tau}T(s) = 0 \tag{3.154}$$

What matters now is to check how much the delay $\tau$ can safely vary without jeopardizing the system stability. If we call the maximum delay $\tau_{\max}$, (3.154) can be rewritten:

$$\chi(s) = 1 + e^{-s\tau_{\max}}T(s) = 0 \tag{3.155}$$

The conditions for which $\chi(s)$ equals zero are still the same. At crossover, we have

$$\left|e^{-s\tau_{\max}}T(\omega_c)\right| = 1 \tag{3.156}$$

since $\left|e^{-s\tau_{\max}}\right| = 1$, (3.156) is similar to

$$\left|T(\omega_c)\right| = 1 \tag{3.157}$$

Regarding the argument, this is where the change takes place:

$$-\pi = \arg(e^{-s\tau_{\max}}) + \arg T(\omega_c) = -\omega\tau_{\max} + \arg T(\omega_c) \tag{3.158}$$

Now, remembering the definition for the phase margin given in (3.8):

$$\varphi_m = \pi + \arg T(\omega_c) \tag{3.159}$$

we extract $T$ from the previous expression and substitute it into (3.158), solving for $\omega\tau_{\max}$:

$$\tau_{\max} = \frac{\varphi_m}{\omega_c} \tag{3.160}$$

where $\omega_c$ represents the crossover frequency of the compensated open-loop gain, and $\varphi_m$ the phase margin (in radians) measured without delay. If the system works well with the actual delay $\tau$, then the delay margin $\Delta\tau$ is defined as

$$\Delta\tau = \tau_{\max} - \tau \tag{3.161}$$

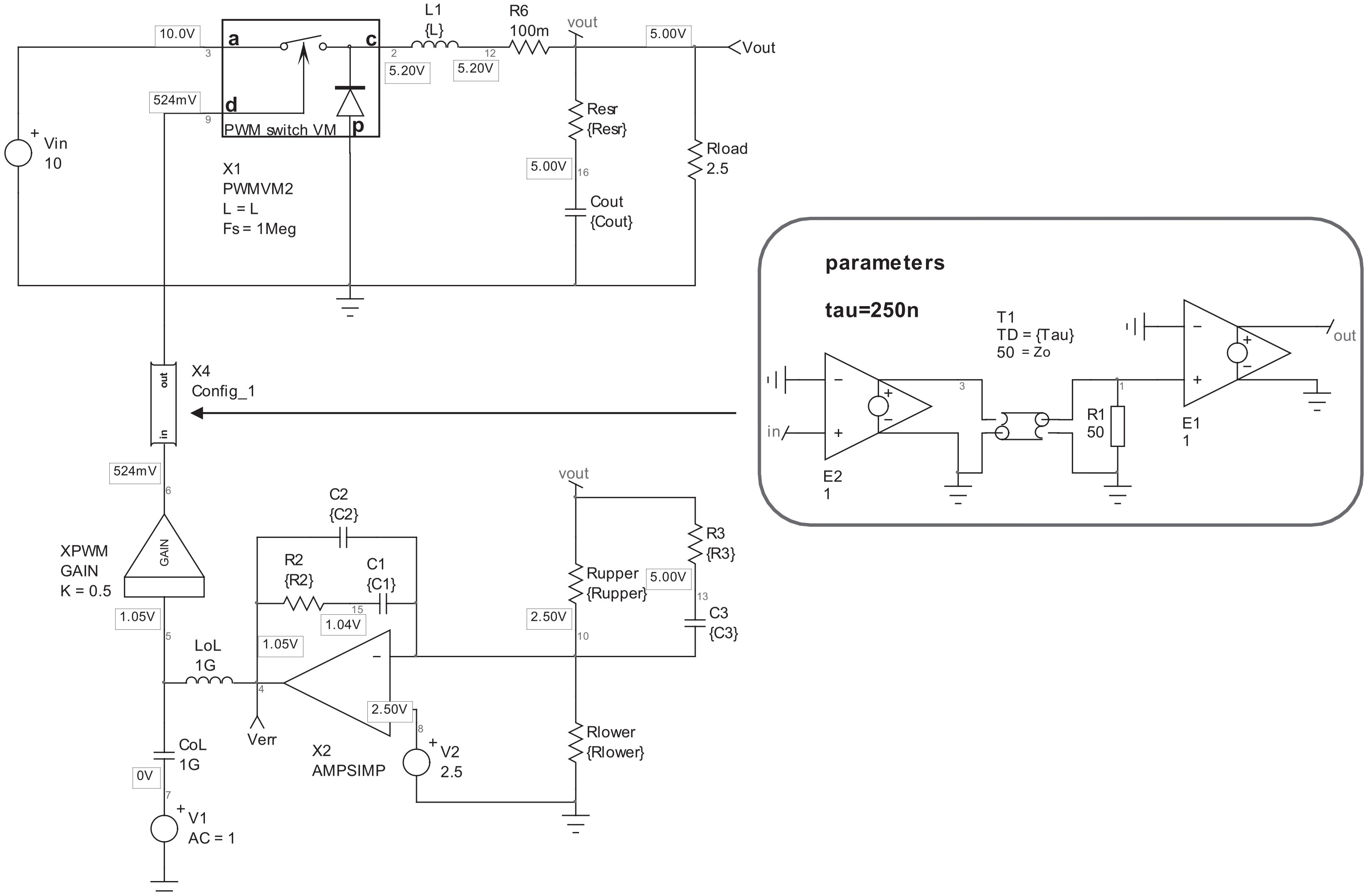



**Figure 3.46** The modulating section shows the addition of the delay block in series with the PWM subcircuit.

Let's take an example of a buck converter where we now consider a delay in the PWM block. The previous buck converter schematic shown in Figure 3.35 has been updated by inserting the delay block in series with the PWM block. The whole circuit appears in Figure 3.46. The initial conversion delay is 250 ns and the switching frequency is 1 MHz.

The loop ac response $T(s)$ is given in Figure 3.47. The compensation imposes a 100-kHz crossover frequency. This value seems high but is not uncommon in small dc-dc converters used by cell phones makers. The compensation we made leads to a 49.5° phase margin at the 100-kHz crossover frequency. If we apply (3.160), as we already have a 250-ns delay when we measure $\arg T(s)$, we can calculate the delay we could further accept in the modulating chain or elsewhere in the conversion process:

$$\Delta\tau = \frac{\varphi_m}{\omega_c} = \frac{49.5}{2\pi \times 100k}\frac{\pi}{180} = 1.375\ \mu\text{s} \tag{3.162}$$

Given the 250-ns original delay, the total delay would therefore be

$$\tau_{\max} = \tau + \Delta\tau = 1.375u + 250n = 1.625\ \mu\text{s} \tag{3.163}$$

To check this result, we have purposely increased the delay to the 1.625 $\mu$s and the updated Bode plot appears in Figure 3.48. As expected, the phase margin has gone down to 0: the system is completely unstable.

As a conclusion on the delay margin, we must always compare the phase margin to the crossover frequency. The previous lines show that when the crossover point is high, a small added delay may perturb the system and make it unstable.

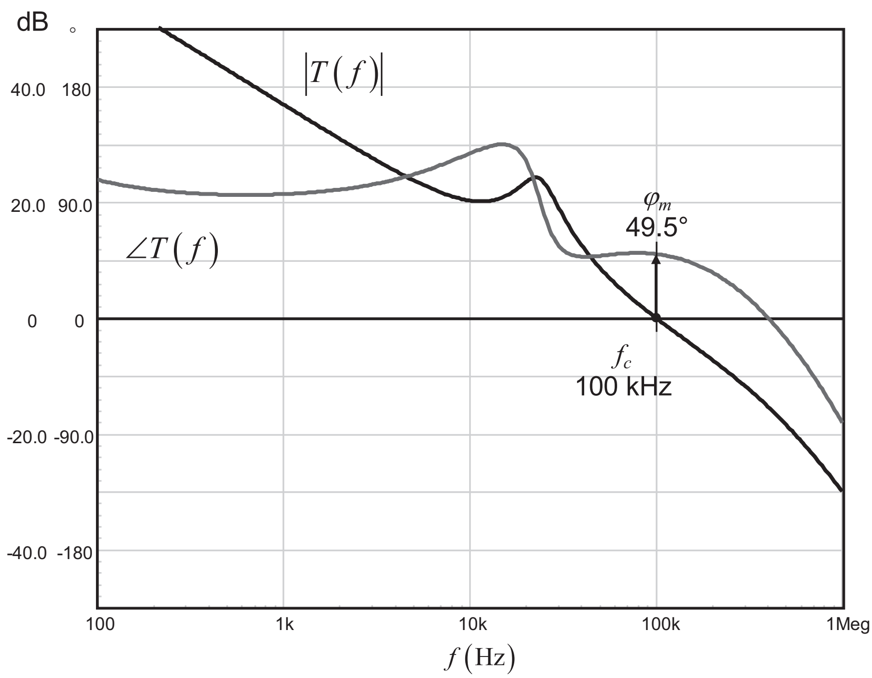



**Figure 3.47** Once compensated, this 1-MHz switching frequency buck converter exhibits a 100-kHz crossover frequency with a 49° phase margin. The PWM delay is 250 ns.

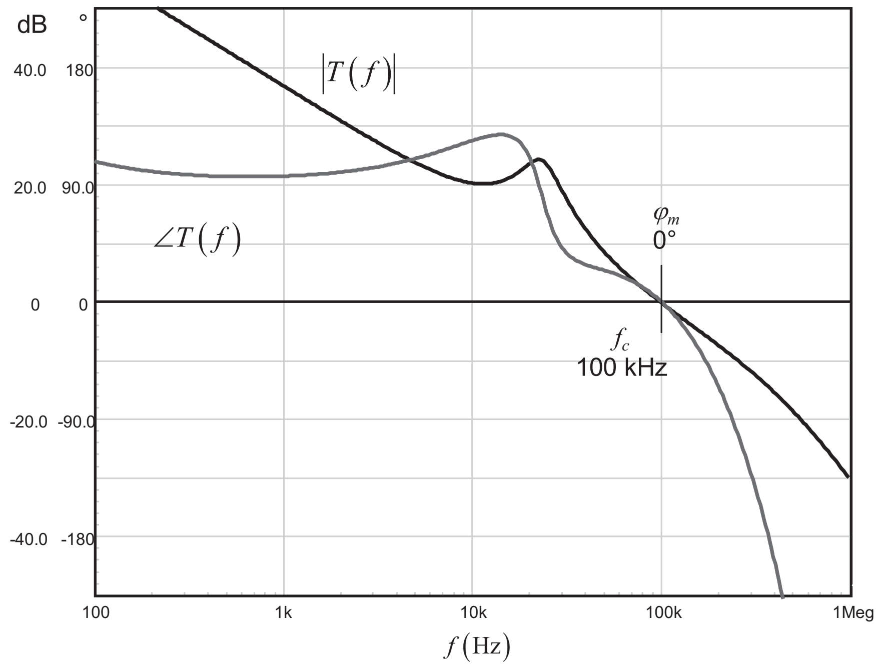



**Figure 3.48** With a delay pushed to the maximum value (1.62 μs), the phase margin vanishes to 0 degree as expected.

Phase margin alone can sometimes lead to an erroneous conclusion on the system robustness, especially in high-bandwidth systems. Delay margin will thus be preferred to phase margin in high-speed dc-to-dc converters.

## 3.4 Selecting the Crossover Frequency

With a closed-loop control system such as the one presented in Figure 3.28, we expect the output to exactly follow the input setpoint with a gain of 1 in this example. In a unity-gain system, we should have both input and output signals exactly matching each other in magnitude and phase. Can this matching be maintained at any frequencies? No, there is a physical limit beyond which the system reaction time starts to increase. This physical limit is called the *bandwidth*: if the setpoint signal changes too rapidly, the control system will not be able to follow it. Similarly, if the frequency of the perturbation or its spectrum content is higher than the system bandwidth, correction will no longer be ensured, and drift or distortion will appear in the output. A solution would be to extend the bandwidth as much as we can, but it would make the system sensitive to incoming noises as well as to its self-generated noise if we consider a switching converter output ripple. There must be a limit imposed to the closed-loop bandwidth, and this limit is set by the open-loop crossover frequency $f_c$.

We have seen that, in certain conditions, the closed-loop response of the system under study could be approximated to a second-order transfer function whose $Q$ depends on the open-loop phase margin. As with any transfer function, the system is affected by a bandwidth, exactly like a filter. By bandwidth, we mean a reduction

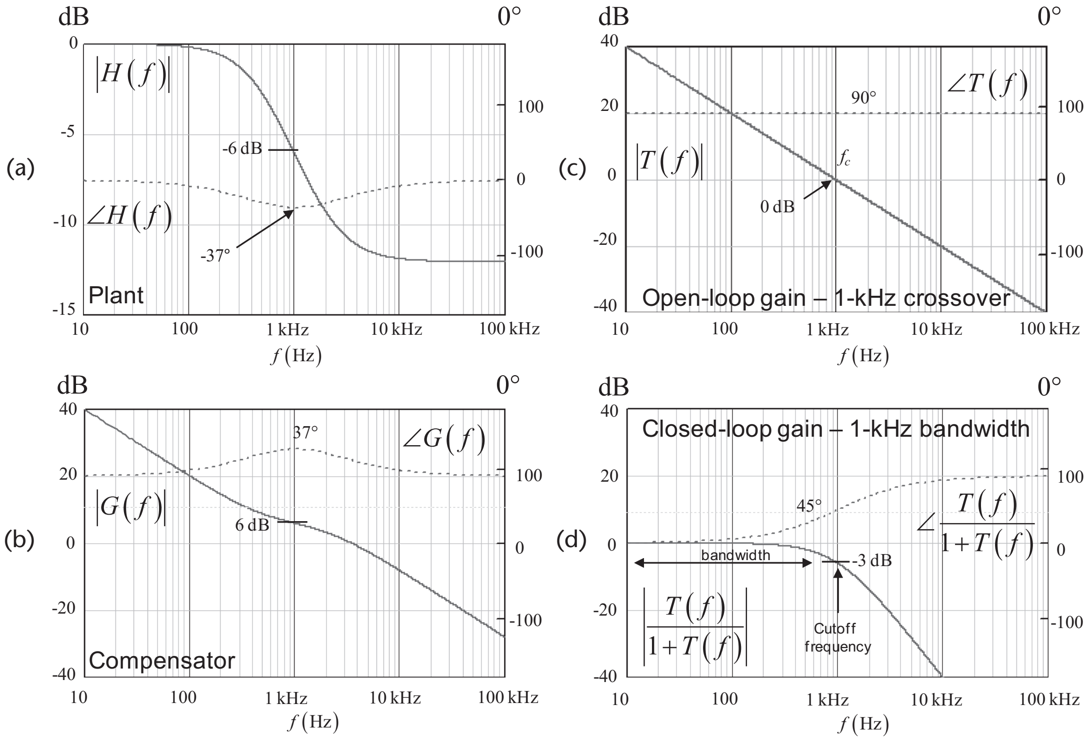



**Figure 3.49** The crossover is often approximated to the bandwidth of the system. They exactly match with each other when the phase margin is 90°.

by 3 dB from its zero-frequency magnitude value. If this value is 1 or 0 dB, as expected in the example, then the frequency at which the gain magnitude has fallen to 0.707 or – 3 dB is also called the cutoff frequency of the control system. How do we set the bandwidth or the cutoff frequency of our closed-loop system? By choosing a crossover frequency $f_c$ during the study of the open-loop transfer function. This crossover frequency will then be obtained by tailoring the transfer function of the compensator $G$ (i.e., placing poles, zeros, and gain to compensate the deficiencies of the plant transfer function $H$). This is exactly what Figure 3.49 shows you.

We start from the plant whose transfer function $H(s)$ appears in the Figure 3.49(a). As we want a 1-kHz crossover frequency, we observe the magnitude and the phase of the transfer function at this frequency point. We read an attenuation of –6 dB, together with a phase lag of 37°. The compensator frequency response appears in Figure 3.49(b) and shows a positive translation of +6 dB and a 37° phase increase at 1 kHz. When we plot $T(s) = H(s)G(s)$ in Figure 3.49(c), we measure a crossover frequency of 1 kHz, together with a phase margin of 90°. If we graph in Figure 3.49(d) the closed-loop gain as expressed by (3.90), we observe a flat response until 1 kHz, where the gain drops by 3 dB: this is our cutoff frequency. As you could see in Figure 3.31, changing the phase margin affects the cutoff frequency and the bandwidth of the system. The case where the crossover frequency exactly equals the cutoff frequency occurs only when the phase margin is 90°. For the rest of the cases, we can say that for undamped systems, the closed-loop bandwidth is roughly equal to 1.5 times the open-loop crossover frequency. This is confirmed by Figure 3.31 for phase margins different than 90°.

In most switching converters design examples, it is common to arbitrarily place the crossover frequency to one-fifth or one-tenth of the switching frequency. However, it is little known that the crossover frequency actually affects other parameters of the converter, such as its output impedance: a relationship exists between both variables. Therefore, once the output capacitor has been selected (based on its operating parameters, such as rms current, temperature, or acceptable voltage ripple), the designer can analytically select his crossover frequency to match the desired output undershoot. In the same way we learned how the phase margin did affect the transient response (recovery time and overshoot), we will explore the link between the crossover frequency and the output impedance.

For linear converters, the crossover frequency also affects the output impedance. If the output undershoot is the design criteria, it can be interesting to select the crossover frequency to minimize it. On the contrary, if the settling time matters, the crossover frequency can be chosen to match a certain design goal. We will come back to the linear case in some of the design examples.

### 3.4.1 A Simplified Buck Converter

A buck converter is a switching system that takes a voltage and decreases it to a lower, regulated value (e.g., a 5-V output obtained from a 10-V input). Basically, a buck converter can be modeled as a low-impedance square-wave generator followed by an $LC$ network. Such a simplified representation appears in Figure 3.50.

The output impedance of the network can be derived by shorting the input excitation. In this case, we have the parallel combination of the inductor and the capacitor networks:

$$Z_{out} = (sL_{out} + r_L) \,||\, \left( r_C + \frac{1}{sC_{out}} \right) \tag{3.164}$$

By inspection, we can see that the inductor resistive path $r_L$ dominates the impedance in dc ($L_{out}$ is shorted and $C_{out}$ is open). The inductor then enters the picture as the frequency increases. In the upper frequency portion of the spectrum, the capacitor impedance starts to take over the inductive section until it becomes a short circuit and leaves the impedance value to its series loss $r_C$. If we ac-sweep the

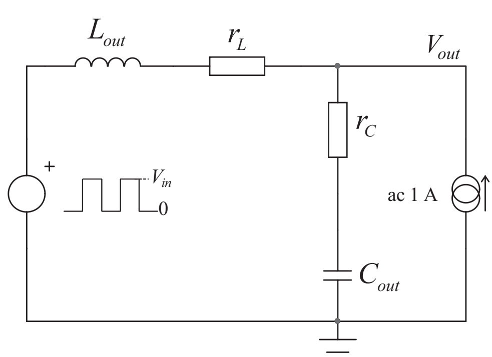



**Figure 3.50** A simplified buck representation where the current source ac sweeps the output impedance.

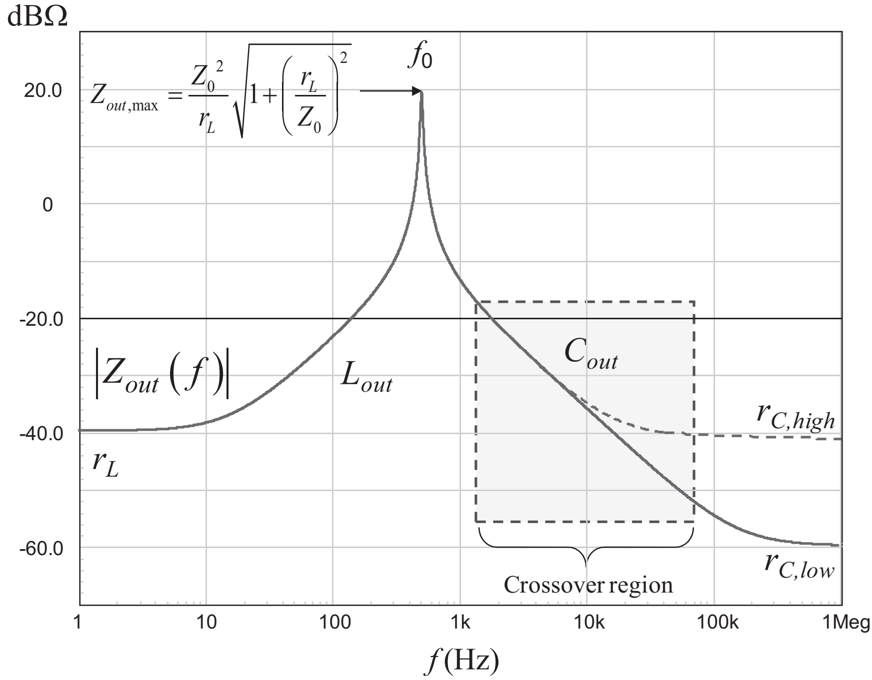



**Figure 3.51** As shown by (3.164), the ohmic losses dominate the output impedance at both extremes of the graph ($f = 0$ and $f = \infty$).

output impedance of this passive network using SPICE, we obtain a graph as the one appearing in Figure 3.51.

As observed, a peaking occurs at the resonant frequency $f_0$. The maximum of this peaking can be analytically derived if we neglect the capacitor ESR contribution $r_C$, as shown in [1]:

$$Z_{out,\max} = \frac{Z_0^2}{r_L}\sqrt{1+\left(\frac{r_L}{Z_0}\right)^2} \tag{3.165}$$

where $Z_0 = \sqrt{\frac{L_{out}}{C_{out}}}$ is the characteristic impedance of the filter and $r_L$ is the inductor series resistor. Such peaking is typical of a buck output impedance behavior where the $LC$ filter has been optimized to minimize the losses (as losses damp the filter). This situation induces a high quality factor, hence a severe peaking in the impedance graph. One of the feedback aims is to minimize the output impedance so that the output voltage drop is kept minimum when a load step occurs. On this plot, the natural output impedance of the filter dramatically peaks at the resonant frequency. Therefore, if we select a crossover frequency below the $LC$ filter resonance, we will not have enough gain to get rid of the resonance, and, despite a good phase margin, the system will severely ring. If we want to obtain a good transient response, we have to make sure the loop gain remains high enough to tame the peaking when it occurs. In other words, the crossover frequency $f_c$ must be selected well above $f_0$ so that some gain exists when the peaking appears. Usually, a ratio of five is enough, but a closed-loop impedance is important to reveal the presence of a peaking some-where. If it still exists, you will have to increase the gain at the resonant frequency or select a higher crossover point.

In Figure 3.51, if we select a crossover region beyond the resonance, we can see an impedance graph dominated by the output capacitor impedance $C_{out}$ and its resistive loss $r_C$, unless it is a very low value. At the crossover frequency, we consider an output impedance combining both elements:

$$|Z_{out}(f_c)| \approx \sqrt{\left(\frac{1}{2\pi f_c C_{out}}\right)^2 + r_C^2} \tag{3.166}$$

With this approach, we purposely ignored a more comprehensive model of the capacitor where its series parasitic inductance, the equivalent series inductor (ESL), is considered. This element plays a role in large $di/dt$ load steps and, given the extremely sharp transient it brings, the loop cannot fight it. The only way to reduce its effects is to select/combine low-ESL types of capacitors such as multilayer devices. In dc-dc converters supplying motherboards, current slopes of several tens of amperes per microsecond are not uncommon. In these cases, it is crucial to account for the ESL presence when computing the converter undershoot.

Figure 3.52 shows the typical response of a power supply submitted to a brutal load step. The output capacitor features both stray elements, ESR and ESL. The first drop is due to the ESL presence. Its amplitude is classically $L\frac{di_{out}(t)}{dt}$, where $L$ is the

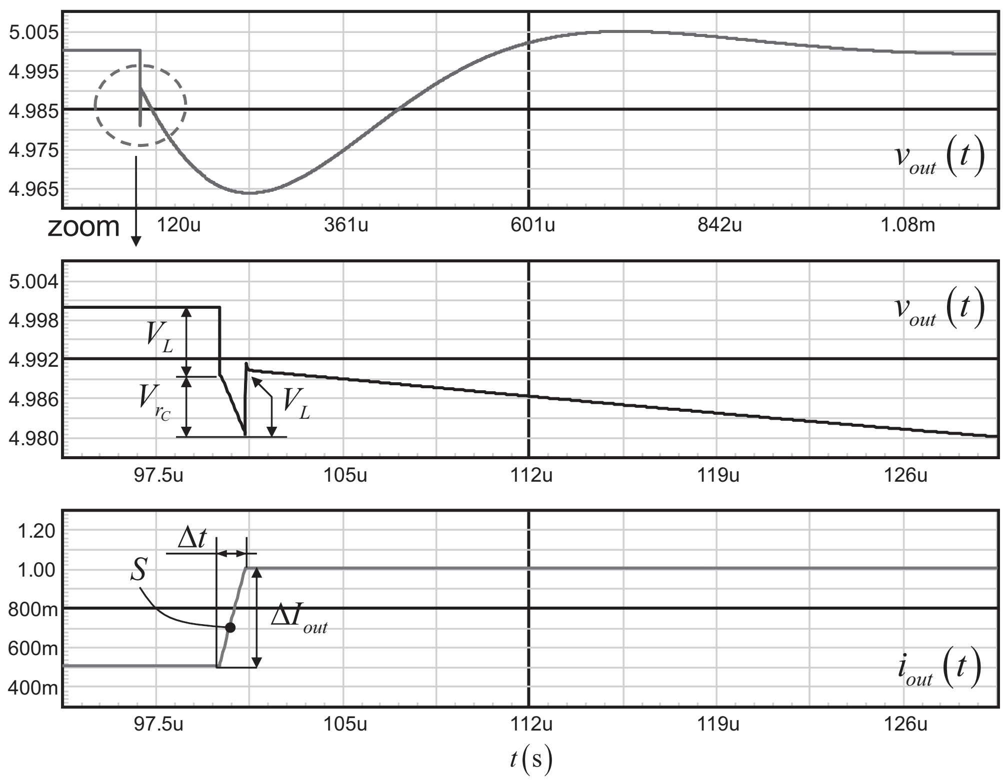



**Figure 3.52** When all parasitic elements are present, the response to a current step reveals different areas where each stray element plays a role.

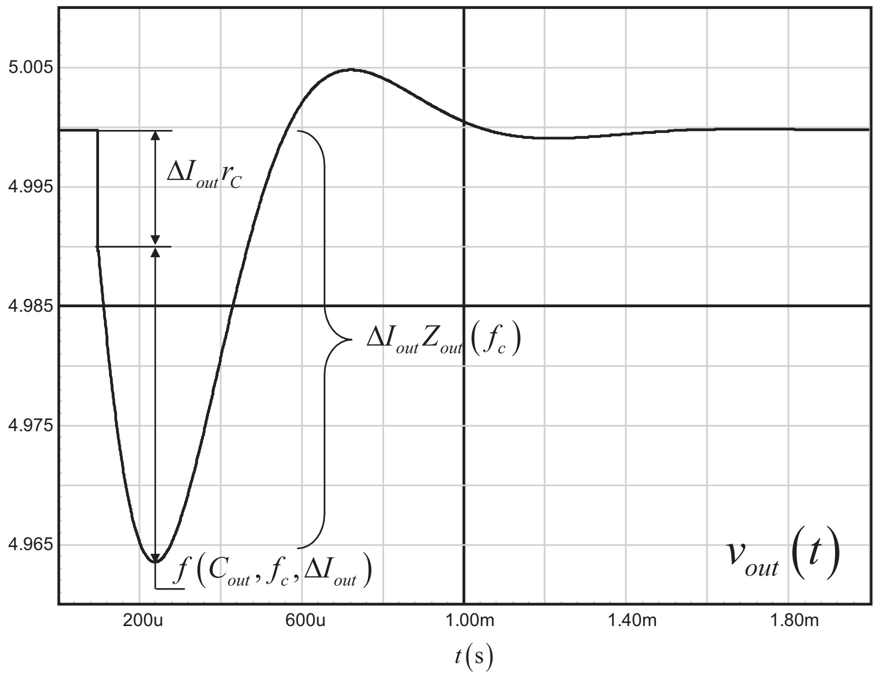



**Figure 3.53** The drop is made of a capacitive undershoot topping a resistive drop. The bandwidth can play on the capacitive contribution, but not on that from the ESR.

ESL of the output capacitor and $di_{out}(t)/dt$ is the output current slope $S$. On top of the ESL voltage, you have the resistive voltage brought by the ESR and simply equal to $r_C \Delta I_{out}$. The variation slope is that of the current. When the current has reached its maximum value and flattens, the ESL drop disappears. The voltage continues to undershoot until the loop eventually takes over. As we will see, the undershoot value depends on the output capacitor and the crossover frequency. Reference [5] studies in detail the implications of the stray elements in the design of dc-dc converters aimed at powering motherboards specifically.

If we now consider much slower slopes than before (e.g., slopes of 1 A/$\mu$s or so, a typical test value for offline converters and general-purpose power supplies), the ESL contribution is usually weak and can be neglected. The voltage response will thus combine a capacitive and a resistive contribution only as shown in Figure 3.53.

This time, in the absence of ESL, the first drop is inherent to the resistive component of the capacitor. It looks straight, but it has the same shape as in Figure 3.52 zoom. Again, besides selecting a low-ESR type, there is nothing you can do to counteract it. The second drop is the pure capacitive contribution. Its amplitude depends on several variables, such as the crossover frequency, the output capacitor itself, and the current step. This is the component on which we can play to make it reach the value we want via crossover frequency selection. Let us see how.

### 3.4.2 The Output Impedance in Closed-Loop Conditions

A power supply can always be represented by its Thevenin-equivalent circuit featuring a dc generator, $V_{th}$, accompanied by its output impedance, $R_{th}$. To obtain

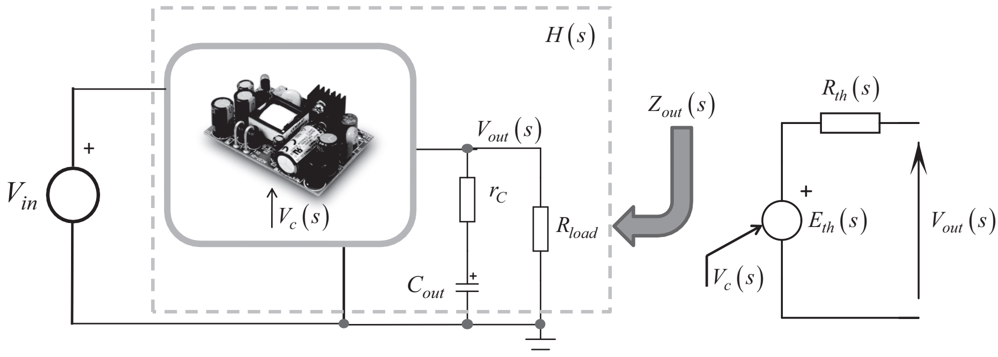



**Figure 3.54** A converter can always be replaced by its Thevenin-equivalent model.

these characteristics, we can take a converter operated in open-loop conditions and extract its parameters as suggested by Figure 3.54. The converter is controlled by a voltage $V_c$, and the power supply delivers current to a load. There is an output capacitor $C_{out}$ affected by an ESR $r_C$.

A simpler representation appears in Figure 3.55 where we can see the control voltage $V_c$ driving the power stage $H$. Please note that $H$ encompasses the total plant transfer function captured at a certain $V_{in}$, while delivering current to its load. The power stage output impedance is externally modeled via the addition of an output impedance—the $R_{th}$ equivalence—denominated $Z_{out,OL}$. $OL$ stands for open loop.

If we write a few lines of algebra, without surprise, we can express the output voltage as

$$V_{out}(s) = V_c(s)H(s) - I_{out}(s)Z_{out,OL}(s) \tag{3.167}$$

The second term expresses the drop incurred to the output impedance $Z_{out,OL}$. Now, let's assume that we want to transform this open-loop system into a closed-loop converter. We need to insert a subtraction block and a compensator box $G(s)$. The new circuit appears in Figure 3.56, featuring a unity-gain return path. This is a different representation—more physical—than that of Figure 3.19, but they are equivalent. The new control variable is no longer $V_c$ but $V_{ref}$.

Following the signals path, we can derive the transfer function $V_{out}(s)/V_{ref}(s)$ of this simple architecture:

$$V_{out}(s) = \left[V_{ref}(s) - V_{out}(s)\right]G(s)H(s) - Z_{out,OL}(s)I_{out}(s) \tag{3.168}$$

$$V_{out}(s)\left[1 + G(s)H(s)\right] = V_{ref}(s)G(s)H(s) - Z_{out,OL}(s)I_{out}(s) \tag{3.169}$$

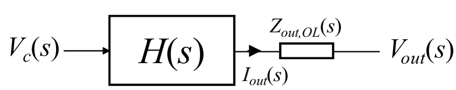



**Figure 3.55** This simplified schematic represents a converter affected by an output impedance $Z_{out,OL}$ which generates a voltage drop.

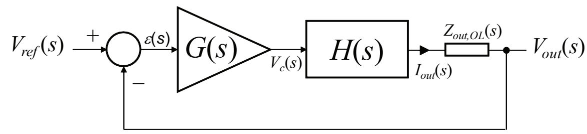



**Figure 3.56** This simplified schematic represents a converter affected by an output impedance $Z_{out,OL}$, which generates a voltage drop.

$$V_{out}(s) = V_{ref}(s)\frac{G(s)H(s)}{1+G(s)H(s)} - \frac{Z_{out,OL}(s)}{1+G(s)H(s)}I_{out}(s) \tag{3.170}$$

Identifying the loop gain $T(s)$ as $G(s)H(s)$, we can reformulate the closed-loop expression as

$$V_{out}(s) = V_{ref}(s)\frac{T(s)}{1+T(s)} - \frac{Z_{out,OL}(s)}{1+T(s)}I_{out}(s) \tag{3.171}$$

In this expression, we have two right-side members. The first one shows that, despite the precision you put in $V_{ref}$, the output voltage will never exactly match the setpoint even with a zero-output impedance converter. There will always be a small error, the dc *static error*, between the output voltage and the setpoint you want to reach. If we neglect the output impedance dc drop in closed-loop conditions, then the deviation between the theoretical setpoint ($V_{ref}$) and the measured output could be reformulated as follows:

$$\varepsilon_0 \approx V_{ref} - V_{ref}\frac{T_{(0)}}{1+T_{(0)}} = V_{ref}\left(1-\frac{T_{(0)}}{1+T_{(0)}}\right) = V_{ref}\left(\frac{1}{1+T_{(0)}}\right) \tag{3.172}$$

If the dc gain $T_{(0)}$ goes to infinity, the error $\varepsilon_0$ is nullified and the output exactly matches the setpoint $V_{ref}$. Therefore, as a good design practice, it is recommended to grow the dc gain to minimize the static error of your converter. A known means to bring the loop gain to infinity for $s = 0$ is to integrate the error voltage as seen in Chapter 1. What is the Laplace equivalent of integration? You introduce the term $\frac{1}{s}$ in the transfer function of $G(s)$. We will later see that the $s$ alone in the denominator is called an *origin pole* since the quotient goes to infinity in dc, when $s$ equals 0. Exactly what we are looking for!

A large dc gain has another good impact if you look at the second right-side member of (3.171). The open-loop impedance has been transformed and is now divided by the loop gain $T(s)$. Again, this second term is similar to that found in (3.24): the output current is a perturbation that is fought by the loop gain in the denominator. Practically, the expression of the new output impedance, the closed-loop output impedance $Z_{out,CL}$, is then

$$Z_{out,CL}(s) = Z_{out,OL}(s)\frac{1}{1+T(s)} \tag{3.173}$$

Calculating the magnitude of this expression, we obtain

$$\left|Z_{out,CL}(s)\right| = \left|Z_{out,OL}(s)\right|\left|\frac{1}{1+T(s)}\right| \tag{3.174}$$

The closed-loop output impedance is then the open-loop output impedance affected by a corrective term, dependent upon the loop gain value. If we have installed a pole at the origin in our compensating path $G(s)$, we know that the loop gain will go to infinity in dc (for $s = 0$), bringing the closed-loop output impedance to almost 0. In our *RLC* filter of Figure 3.50, the dc closed-loop output impedance will thus be much smaller than $r_L$, due to the control system operation. As the frequency increases, we learned that $T(s)$ must be tailored through $G(s)$ to roll off the gain at a certain point, forcing a 0-dB crossover at a selected frequency, $f_c$. If $T(s)$ now drops in magnitude as the frequency is increased, its action on the output impedance becomes weaker, until unity is reached at crossover. At this point, the closed-loop output impedance has come back to its open-loop value:

$$\left|Z_{out,CL}(f_c)\right| \approx \left|Z_{out,OL}(f_c)\right| \tag{3.175}$$

To illustrate this theory, we have inserted the buck converter featuring the open-loop output impedance curve of Figure 3.51 in closed-loop configuration. The compensated loop gain $T(s)$ and the resulting plots (open-loop and closed-loop) are displayed in Figure 3.57.

In this picture, we clearly see the impact of the decreasing loop gain, which eventually reaches unity at a 10-kHz frequency. At this point, the output impedance is almost that without feedback. Let's try to derive what hides behind the term *almost*.

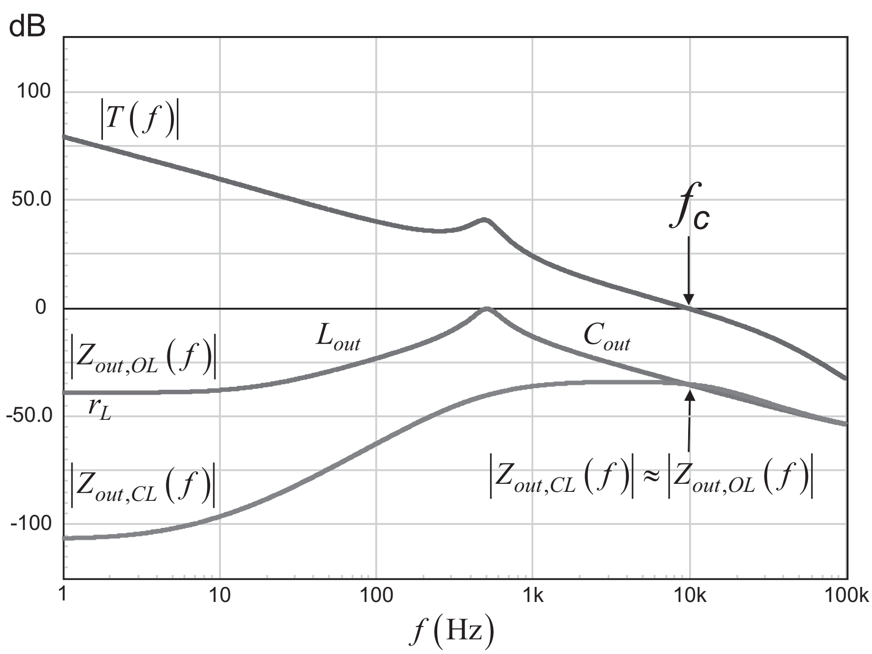



**Figure 3.57** The loop gain reduces in magnitude as the frequency increases. It clearly impacts the output impedance.

### 3.4.3 The Closed-Loop Output Impedance at Crossover

To obtain the exact definition of the output impedance at crossover, we will derive the magnitude of the right term in (3.174). Please note that we keep observing the loop gain in the vicinity of the crossover frequency, as already highlighted in Figure 3.29.

$$\left|\frac{1}{1+T(j\omega_c)}\right| = \left|\frac{1}{\dfrac{1}{j\dfrac{\omega_c}{\omega_0}\left(1+j\dfrac{\omega_c}{\omega_2}\right)}+1}\right| = \sqrt{\frac{(\omega_c\omega_2)^2+\omega_c^4}{(\omega_c\omega_2)^2+(\omega_c^2-\omega_0\omega_2)^2}} \tag{3.176}$$

In this equation, we can now substitute $\omega_c$ by its definition from (3.99) and $\omega_0$ from (3.92):

$$\begin{aligned}\left|\frac{1}{1+T(j\omega_c)}\right| &= \sqrt{\frac{Q^4\omega_2^2}{\omega_0^2+Q^4\omega_2^2+\omega_0\omega_2-\omega_0\omega_2\sqrt{4Q^4+1}}} \\ &= \frac{1}{2Q\sqrt{\dfrac{2Q^2+1-\sqrt{1+4Q^4}}{\left(1+\sqrt{1+4Q^4}\right)\left(\sqrt{1+4Q^4}-1\right)}}}\end{aligned} \tag{3.177}$$

Now replace $Q$ by its definition in (3.110) and have fun simplifying the result:

$$\left|\frac{1}{1+T(j\omega_c)}\right| = \frac{1}{\sqrt{\dfrac{1}{\cos(\varphi_m)}\left(2\cos(\varphi_m)+\left[\dfrac{1+\cos^2(\varphi_m)}{\sin^2(\varphi_m)}\right]\left[\cos^2(\varphi_m)-1\right]+1-\cos^2(\varphi_m)\right)}} \tag{3.178}$$

If everything goes well, you should find

$$\left|\frac{1}{1+T(j\omega_c)}\right| = \frac{1}{\sqrt{2-2\cos(\varphi_m)}} \tag{3.179}$$

This is the expression of the closed-loop gain magnitude. For a phase margin of 90°, we obtain 0.707 or exactly –3 dB. In that case, both the open-loop crossover frequency and the closed-loop cutoff frequency are equal. For different phase margin values, the closed-loop gain starts to peak and the cutoff frequency deviates from the open-loop crossover frequency. This is what has been shown in Figure 3.31.

Back to our output impedance equation in (3.166), the exact definition thus becomes

$$|Z_{out}(f_c)| = \sqrt{\left(\frac{1}{2\pi f_c C_{out}}\right)^2+r_C^2}\,\frac{1}{\sqrt{2-2\cos(\varphi_m)}} \tag{3.180}$$

From this expression, we are now able to link the selection of the crossover frequency with a design criterion, the output impedance. You could also link the crossover frequency with the closed-loop settling time if you wish. However, the voltage drop in a converter represents an important parameter, and we are going to use it in the following example to set the crossover frequency. Let's assume we have selected a 1000-μF capacitor based on several parameters such as its rms current capability at high temperature, its size, and also its cost. From the manufacturer data sheet, we read that its ESR is 30 mΩ. Now, assume the specification imposes a maximum undershoot $V_p$ of 90 mV when the output undergoes a current step of 2 A. First, let's calculate the contribution of the ESR alone:

$$V_{ESR} = \Delta I_{out} r_C = 2 \times 30m = 60 \text{ mV} \tag{3.181}$$

It is obvious that if this drop approaches the maximum undershoot, there is nothing you can do beyond selecting a bigger capacitor featuring a smaller ESR or associate capacitors in parallel. Here, the approach is different: we will try to select the crossover frequency to match the undershoot specifications. At this point, from the simplified impedance definition value, we can extract the crossover frequency so that the 1000-μF addition to the 60-mV ESR drop makes the whole undershoot stay below 90 mV. With a 2-A step, the closed-loop output impedance must stay below

$$Z_{out}(f_c) < \frac{V_p}{\Delta I_{out}} < \frac{90m}{2} < 45 \text{ m}\Omega \tag{3.182}$$

Equation (3.180) is really too comprehensive to be used as is. Keep in mind that capacitors and ESRs are affected by dispersions going up to ±30 percent or ±40 percent, so it is really meaningless to run calculations down to the third decimal. Experience shows that the original equation (3.175) gives acceptable practical results. Adopting this approach, the way to derive the crossover frequency is to solve for $f_c$ in the following equation:

$$\sqrt{\left(\frac{1}{2\pi f_c C_{out}}\right)^2 + r_C^2} \approx Z_{out}(f_c) \tag{3.183}$$

From which we can extract the crossover frequency $f_c$ target:

$$f_c > \frac{1}{2\pi C_{out}\sqrt{Z_{out}(f_c)^2 - r_C^2}} > \frac{1}{6.28 \times 1m \times \sqrt{45m^2 - 30m^2}} > 4.7 \text{ kHz} \tag{3.184}$$

Obviously, this is an approximation and practical experiments in the bench will be needed to confirm this choice. However, experience shows that the final result is often not too far from what is expected. The previous derivation is a good starting point for the crossover frequency selection.

### 3.4.4 Scaling the Reference to Obtain the Desired Output

In most of our drawing examples for the sake of simplicity (e.g., in Figure 3.56), the output is directly compared to the setpoint. In reality, this setpoint is made of a reference source whose value can be quite low compared to the observed output. For instance, the vast majority of stable sources are built on a bandgap voltage reference whose level is 1.25 V. Therefore, in high-voltage applications where the output voltage can be much higher than 1.25 V (400 V in power factor correction circuits), we do not directly observe Vout but a reduced image of it. This is done by installing a resistive divider between the observed variable and the input of the error amplifier. Figure 3.58 shows this typical arrangement with an op amp.

This divider introduces a scaling term equal to

$$\alpha = \frac{R_{lower}}{R_{lower} + R_{upper}} \tag{3.185}$$

In the previous example, the op amp undergoes a local feedback via the element $Z_f$. The presence of this element brings a well-known characteristic called a virtual ground. It means that considering an infinite or very large op amp open-loop gain $A_{OL}$, the voltage $\varepsilon$ between the plus and minus input is null, or extremely small: both input levels must be equal at steady state. If it is not the case, the op amp fights the difference along its dynamic range. If it cannot ensure this equality, it is stuck in its upper or lower stop. The presence of this virtual ground has an impact on the ac analysis of this block. In ac, as we consider a voltage regulator, we are interested in the converter's ability to reject the incoming perturbations such as the input voltage or the input current. During these tests, the reference voltage does not change: its ac modulation is 0. As both op amp inputs have equal levels because of the virtual ground, there is no ac voltage modulation across the lower-side resistor $R_{lower}$, and it simply disappears from the ac analysis. We can quickly go through a few equations to show that.

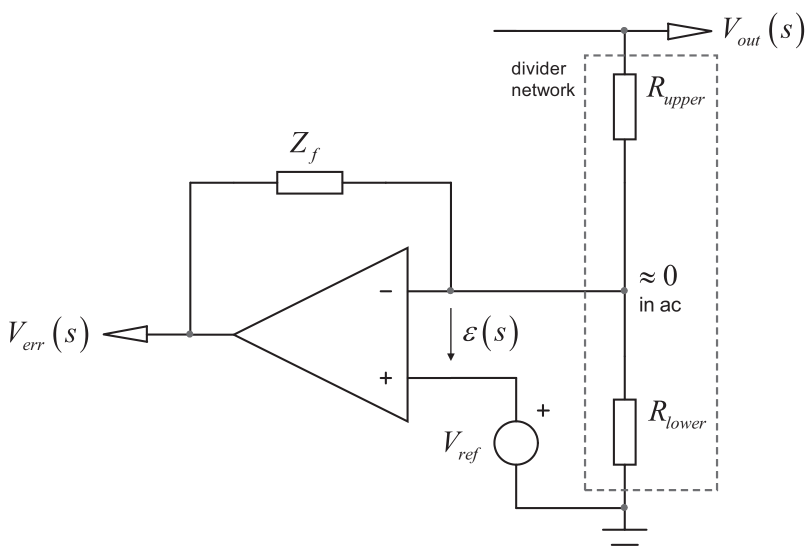



**Figure 3.58** In power converters, a divider is inserted between the observed output voltage and the reference voltage.

The output voltage of Figure 3.58 configuration can be obtained by applying the superposition theorem:

$$V_{err}(s)\Big|_{V_{out}=0} = V_{ref}(s)\left(\frac{Z_f}{R_{upper} \parallel R_{lower}} + 1\right) \tag{3.186}$$

$$V_{err}(s)\Big|_{V_{ref}=0} = -V_{out}(s)\frac{Z_f}{R_{upper}} \tag{3.187}$$

Adding both expressions leads to the error voltage equation:

$$V_{err}(s) = V_{ref}(s)\left(\frac{Z_f}{R_{upper} \parallel R_{lower}} + 1\right) - V_{out}(s)\frac{Z_f}{R_{upper}} \tag{3.188}$$

If we now derive the previous expression with respect to $V_{out}$ and keeping $V_{ref}$ constant, we obtain the ac small-signal response in which the lower-side resistor plays no role:

$$V_{err}(s) = -\frac{Z_f}{R_{upper}} \tag{3.189}$$

Please note that the resistor plays a role in dc as it fixes, together with $R_{upper}$ and $V_{ref}$, the operating point in (3.188). But owing to the virtual ground, only $R_{upper}$ plays a role in poles/zeros calculations.

Is this always the case? No, when there is no virtual ground, as with an operational transconductance amplifier (OTA), for instance, the virtual ground is lost and both $R_{upper}$ and $R_{lower}$ play a role in the ac analysis. We will see it in the chapter dedicated to compensating with an OTA. This remark also applies if $Z_f$ is not a resistance but an impedance made of a series resistor-capacitor combination. In this case, we have well a virtual ground in ac (the capacitor offers a certain impedance) and $R_{lower}$ is off the picture for the calculation. However, in dc, at *s* equals 0, if you try to evaluate the output static error, the capacitor impedance is infinite and the virtual ground is lost: the divider comes back in the calculation and you must account for its presence together with the op amp open-loop gain. Fortunately, in the vast majority of cases, because of the origin pole we insert via the series combination of a resistor and a capacitor for $Z_f$, and also because we perform an ac sweep where the op amp virtual ground is effective, we simply ignore $R_{lower}$.

If we place ourselves in this latter case, we can transform Figure 3.56 representation to now include the divider ratio, solely affecting the reference voltage and not the rest of the chain. The new drawing appears in Figure 3.59.

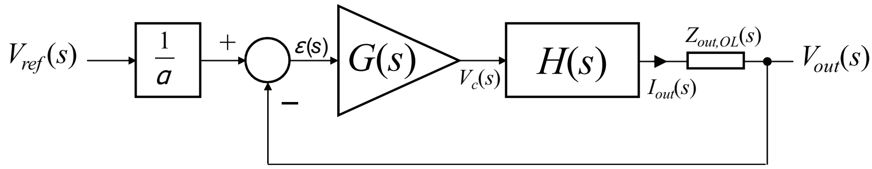



**Figure 3.59** The resistive divider does not play a role in the output impedance expression; it only affects $V_{ref}$.

We are still in a unity feedback system; all the equations we have derived so far still apply. If you rederive the output voltage expression, you should be able to show that the updated expression obeys

$$V_{out}(s) = \frac{V_{ref}}{\alpha}\frac{G(s)H(s)}{1+G(s)H(s)} - \frac{Z_{out,OL}}{1+G(s)H(s)}I_{out}(s) \tag{3.190}$$

In this equation, the first term, again, is the theoretical dc output value you would expect from the converter if we neglect its output impedance contribution:

$$V_{out} \approx \frac{V_{ref}}{\alpha}\frac{G_{(0)}H_{(0)}}{1+G_{(0)}H_{(0)}} \tag{3.191}$$

In presence of a large dc open-loop gain $G_{(0)}H_{(0)}$, this equation simplifies to

$$V_{out} \approx \frac{V_{ref}}{\alpha} \tag{3.192}$$

From this expression, we can see that a divider value $\alpha$ allows us to scale the output voltage to any combination of the reference voltage independently from the other loop parameters. This is what feedback is all about: we want the output to match the input setpoint by a factor made independent from the forward path of the control system. With a ratio of 0.5 and a 2.5-V reference voltage, we expect the output voltage to equal 5 V. However, this is a theoretical value. Equation (3.191) shows that a correction factor must be added to account for the open-loop gain. This open-loop gain, as large as it can be, always introduces a small deviation between the expected target and the final measurement. In presence of extremely large open-loop gains (90 dB or more), as when an origin pole exists in the transfer function $G(s)$, the error becomes extremely small. Neglecting the closed-loop impedance effect, the static output error between the theoretical value given by (3.192) and that obtained via (3.191) becomes

$$\varepsilon_0 \approx \frac{V_{ref}}{\alpha} - \frac{V_{ref}}{\alpha}\frac{G_{(0)}H_{(0)}}{1+G_{(0)}H_{(0)}} = V_{ref}\left(\frac{1}{\alpha} - \frac{1}{\alpha\left(1+\dfrac{1}{T_{(0)}}\right)}\right) \tag{3.193}$$

From this equation, we can see that despite having an extremely precise reference voltage, when embedded into a low gain loop, the output will always deviate from the expected target defined by (3.192).

To exercise these new equations and results, we have built a buck converter delivering 5 V from a 10 V dc source.

Its linear representation is given in Figure 3.60, using an averaged circuit, the PWM switch model, which is extremely useful in the ac analysis of switching converters. To understand how this model works, please check [1] at the end of this chapter. As confirmed by the bias points in Figure 3.60, the switching regulator delivers the right output when loaded by a 2-Ω resistor (2.5 A).

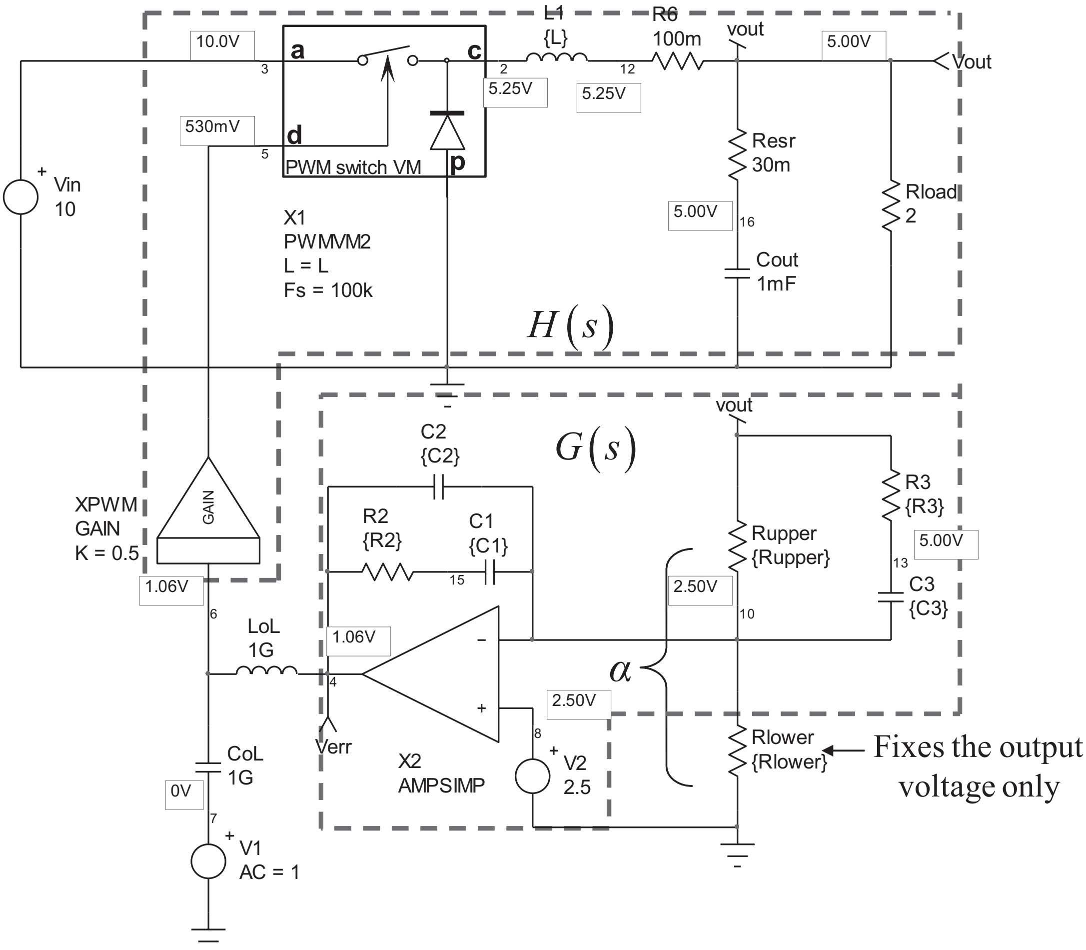



**Figure 3.60** A simple linear buck averaged model can be used as an example to check the validity of our approach.

Suppose we have the following values in our system:

$V_{ref} = 2.5$ V, the reference voltage
$H_0 = 4.7$, the power stage dc-gain
$G_0 = 2000$, the error amplifier dc-gain

In that case, (3.193) would give the following output error:

$$\varepsilon_0 \approx 2.5 \times \left( \frac{1}{0.5} - \frac{1}{\left( \frac{1}{4.71 \times 2000} + 1 \right) \times 0.5} \right) = 530 \ \mu\text{V} \tag{3.194}$$

Leading to an output voltage of $5 - 0.53m = 4.9995$ V. Figure 3.60 output file (.OUT) gives us a level of 4.9993 V, very close to this number.

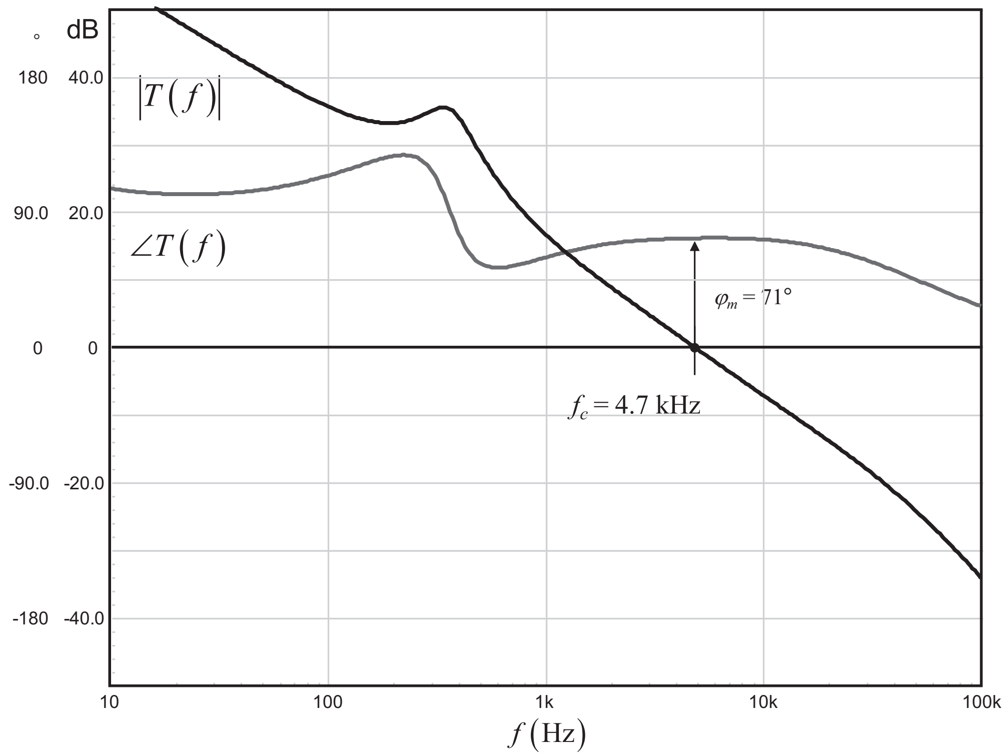



**Figure 3.61** The compensated regulator exhibits a 4.7-kHz bandwidth together with a comfortable phase margin of 71°.

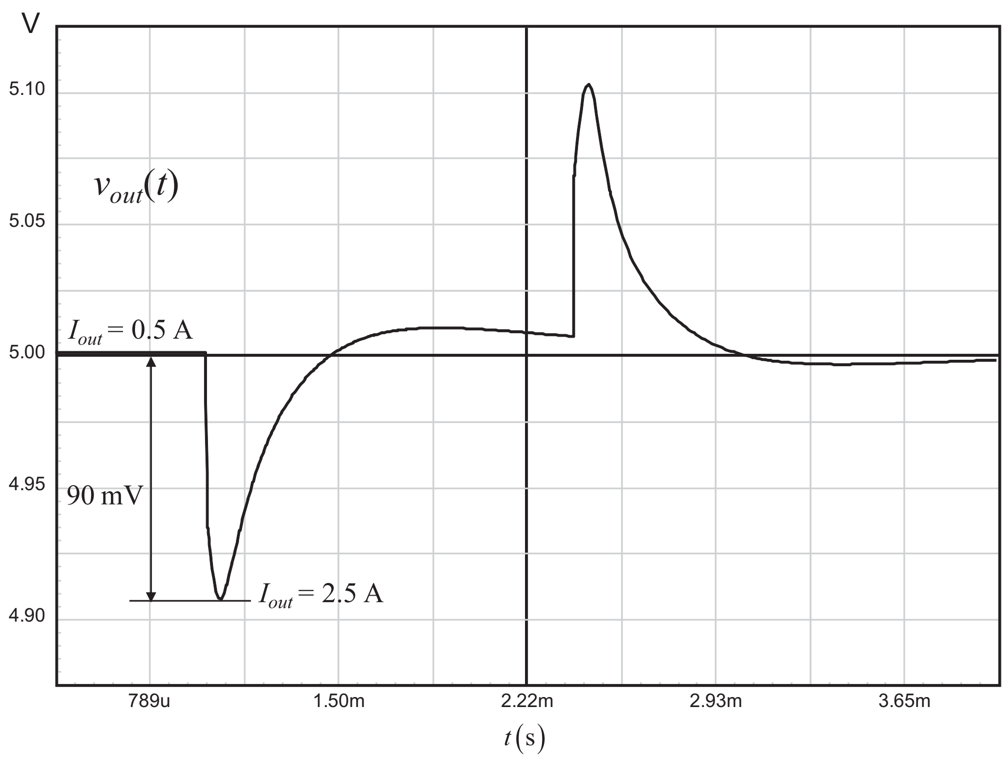



**Figure 3.62** The step-response shows a signal whose undershoot is close to the target.

Without disclosing the operating details yet, we have compensated the regulator to offer a 4.7-kHz bandwidth together with a 71° phase margin, following the recommendations of (3.184). The resulting Bode plot appears in Figure 3.61.

The load resistor $R_{load}$ is now been replaced by a current source stepping the output from 0.5 A to 2.5 A with a slope of 1 A per microsecond. When we observe the output, the voltage deviation reaches 90 mV as shown in Figure 3.62. This is very close to the values we have targeted in the design example. Of course, a real experiment has to be carried out with the selected capacitor to check if the breadboard results match the calculations. This is a general practice that any serious designer must follow: always verify on a bench prototype if the theoretical assumptions were correct and lead to an acceptable practical result.

### 3.4.5 Increasing the Crossover Frequency Further

In the given example, we have selected the crossover frequency to reduce the capacitive contribution and make the total drop fit the specifications. This method works well for slow converters such as ac-dc adapters for notebooks or netbooks, where the crossover frequency often lies between 1 and 5 kHz. The output capacitor is selected for its rms current capabilities and the acceptable output ripple. Given the available ESR, the designer can apply this strategy to reduce the capacitive drop. In high-speed converters such as dc-dc switchers for the telecom market, the drop must really be reduced to a minimum value that is the ESR contribution. The only possibility is then to increase the crossover frequency until the capacitive drop vanishes to a negligible value, the ESR term remaining alone. This is what we did in Figure 3.63 where the converter crossover frequency was purposely increased while the open-loop phase margin was purposely kept to 60°.

As the crossover frequency increases, the undershoot brought by the capacitor is reduced as the converter reacts faster. A possible compensation strategy, in that case, would be to select the crossover frequency so that the capacitive contribution becomes negligible, leaving the ESR drop alone. It would allow us to push the crossover frequency just enough to reduce the undershoot brought by the capacitor and not beyond what is really necessary. This naturally reduces the risks of picking up parasitic noises and making the loop unstable. As an arbitrary value, let's assume we want to reduce the capacitive contribution to a maximum of 20 percent of $r_C$. The output impedance equation could thus be updated so that its value gives $1.2r_C$ at the crossover frequency value we seek:

$$\sqrt{\left(\frac{1}{2\pi f_c C_{out}}\right)^2 + r_C^2} = 1.2r_C \tag{3.195}$$

Solving for fc gives us

$$f_c \approx \frac{0.24}{C_{out} r_C} \tag{3.196}$$

Figure 3.63 curves were obtained with a 1000-µF capacitor featuring a 20-mΩ ESR. If we apply the previous formula to reduce the undershoot to almost the ESR

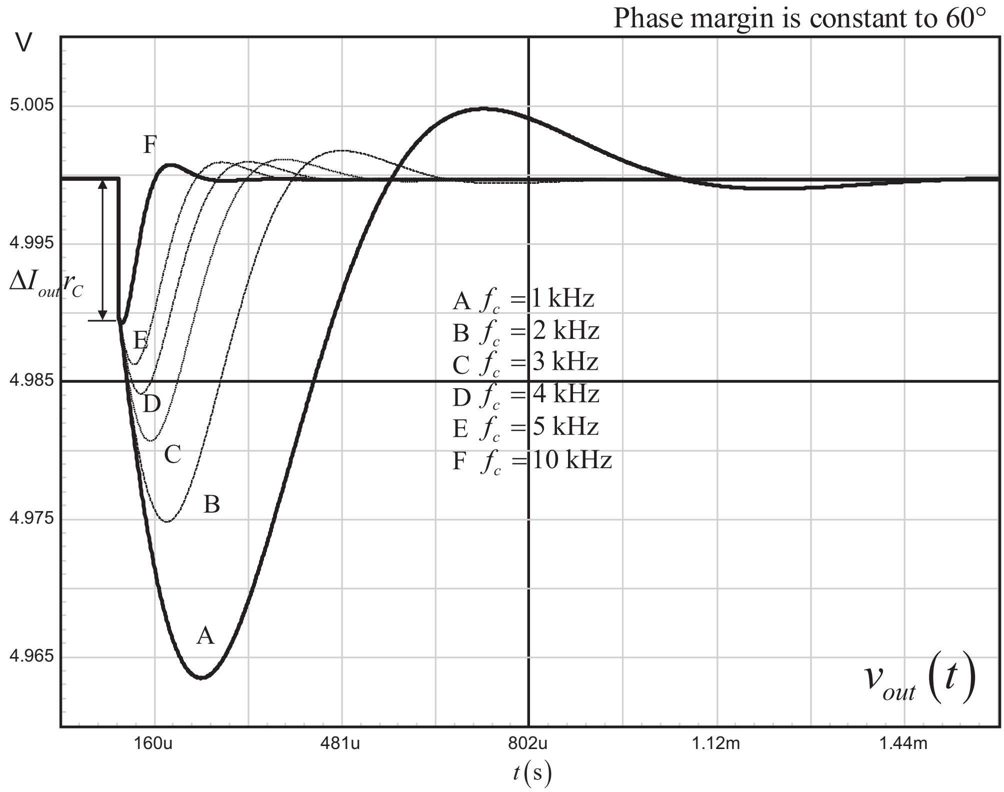



**Figure 3.63** When the crossover frequency increases, the capacitive undershoot reduces until the drop becomes dictated by the ESR only.

contribution alone, we obtain a crossover frequency of 12 kHz, in line with what curve F brings. We will later use this formula in our LDO compensation example described in Chapter 9.

## 3.5 Conclusion

In this chapter, we have tried to show that phase margin and crossover frequency could be analytically selected to match the project specifications. The simple formula linking the open-loop phase margin to the closed-loop quality factor gives the designer a relationship between a design criteria and a final response. Rather than pulling the minimum phase margin criteria out of thin air, as it often the case in textbooks, the proposed method explores an analytical path showing its relationship to the transient behavior you will get in closed loop. The designer can thus tweak his system, depending on the needed performance in terms of response speed or overshoot. Finally, the output impedance derivation for a switching converter highlights the link between the crossover frequency and the voltage drop in response to a current step. You will no longer select your crossover frequency based on a rule of thumb with the risk of pushing it too far: check the voltage drop specifications, find the output capacitor you have, and adjust the crossover to the needed value.

Among the design criteria, if the phase and gain margins are the most popular variables, we have shown that they are not a sufficient condition for system robustness. Modulus and delay margins should be used instead, especially if hidden resonances exists in the plant transfer function and if a high-bandwidth system is desirable.

It is also important to keep in mind that the analytic descriptions we went through, phase margin and transient response followed by the output impedance prediction, are approximations. They have to be used as rules of thumb during the design phase. For instance, should you want to obtain the exact transient signature of a system, you must calculate its response to a step excitation in Laplace and then convert the result to its time-domain equivalent as shown in [2]. Needless to say, most of the designers never undertake such a complex work when several poles and zeros are at stake. Furthermore, this is a small-signal result and we know that nonlinear elements are at play in this mode. The resulting signal is then an approximation as well.

## References

[1] Basso, C., *Switch Mode Power Supplies: SPICE Simulations and Practical Designs*, New York: McGraw-Hill, 2008.

[2] Peretz, M. M., and S. Ben-Yaakov, “Revisiting the Closed Loop Response of PWM Converters Controlled by Voltage Feedback,” APEC 2008, Austin, TX.

[3] Adar, D., and S. Ben-Yaakov, “Generic Average Modeling and Simulations of Discrete Controller,” APEC 2001, Anaheim, CA.

[4] Özbay, H., *Feedback Control Theory*, Boca Raton, FL: CRC Press, 2000.