LUND UNIVERSITY

A PID Tuner Based on Phase Margin Specification

Hägglund, Tore

1981

Document Version:
Publisher's PDF, also known as Version of record

Link to publication

Citation for published version (APA):
Hägglund, T. (1981). *A PID Tuner Based on Phase Margin Specification*. (Technical Reports TFRT-7224). Department of Automatic Control, Lund Institute of Technology (LTH).

Total number of authors:
1

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

CODEN: LUTFD2/(TFRT-7224)/1-020/(1981)

A PID TUNER BASED ON PHASE MARGIN SPECIFICATION

TORE HÄGGLUND

DEPARTMENT OF AUTOMATIC CONTROL
LUND INSTITUTE OF TECHNOLOGY
SEPTEMBER 1981

LUND INSTITUTE OF TECHNOLOGY
DEPARTMENT OF AUTOMATIC CONTROL
Box 725
S 220 07 Lund 7 Sweden

Document name
Internal report

Date of issue
September 1981

Document number
LUTFD2/(TFRT-7224)/0-020/(1981)

Author(s)
Tore Hägglund

Supervisor
Karl Johan Åström

Sponsoring organization
Swedish Board of Technical Development
Contract 78-3763

Title and subtitle

A PID tuner based on phase margin specification

Abstract

The report describes a method to tune a PID-controller so that the controlled system gets a desired phase margin. By introducing a relay with hysteresis in the control loop, it is possible to estimate the characteristics of the process needed for the tuning procedure. The estimation is based on the describing function technique. The tuning procedure can be performed either manually or automatically by a computer.

Key words

Classification system and/or index terms (if any)

Supplementary bibliographical information

ISSN and key title

ISBN

Language
English

Number of pages
20

Recipient's notes

Security classification

Distribution: The report may be ordered from the Department of Automatic Control or borrowed through the University Library 2, Box 1010, S-221 03 Lund, Sweden, Telex: 33248 lubbis lund.

A PID TUNER BASED ON

PHASE MARGIN SPECIFICATION

Tore Hägglund

Department of Automatic Control
Lund Institute of Technology
September 1981

Table of contents

1. INTRODUCTION

Today, most controllers in the industry are analog PID-controllers. Engineers and process operators are familiar with them, and have often a good feeling for the effects of their different parameters.

Due to the rapid development of computers, the analog PID-controllers are more and more substituted by small computers. It has therefore been possible to introduce more advanced control algorithms.

When the operating conditions in a process change, the regulators should also be retuned. This is often both a difficult and time consuming work, and it is therefore ignored in many cases. Self-tuning controllers can consequently save time for operators and improve the behavior of the process.

There are many examples where ordinary self-tuning controllers have improved the process control. The controllers are mainly based on minimum variance or pole placement design. It is a rather big difference in complexity between a PID-algorithm and these self-tuning controllers. For some computers, the self-tuning controllers are still too complecated. The complexity of a self-tuning controller is dependent on the order of the system model, but not on how hard the system is to control.

From what is said above, it can be concluded that there is a need for simple self-tuning controllers. These shall be able to handle systems of high order. It is also favorable if the PID structure of the controller can be retained.

In Wittenmark et al(1980), a self-tuning PID-controller is presented. It is an ordinary self-tuning controller based on pole placement design. The model is however always of second order, independently of the process order, and the controller parameters are translated to the ordinary PID parameters. Experience is that the controller works well, if the process does not contain any time delay and if the behavior of the process can be well approximated by a second order model.

A fundamentally different PID-tuner is given in Åström(1981). The tuning procedure is not running all the time, as in ordinary self-tuning controllers, but only when it is initiated by an operator. This is of course a drawback if something suddenly happens to the process. On the other hand, many operators like to have it arranged in this way, they want to decide themselves when the controller is to be tuned. When the tuning procedure is to be performed, a relay is introduced in the control loop. It makes the system oscillate. By measuring the amplitude and the frequency of the oscillation, it is possible to determine certain properties of the process. The tuner is based on the Ziegler-Nichols design, a method created for second order systems with a time delay. The behavior of the closed loop system is therefore highly dependent on whether that is a good approximation of the true system or not. The PID-tuner can be viewed as a helping aid for further manual tuning. It is simpler than the one described in Wittenmark et al(1980), and therefore suitable even for small computers.

In Schuck(1959), an adaptiv controller which maintains a constant amplitude margin in the system is presented. An essential part of that controller is a relay, as in Åström(1981). The PID-tuner described in this report is inspired by Schuck(1959) and Åström(1981). It has many properties in common with the one in Åström(1981). The tuning procedure is only performed when the operator wants it. At these instants, a relay with hysteresis is introduced in the loop, see figure 1.1. It causes a self oscillation in the system. The relay with hysteresis makes it possible to identify other characteristics of the process than those which can be identified by an ordinary relay. It will be shown that it is possible to tune the PID-controller so that a desired phase margin of the system is obtained. The phase margin is a useful design parameter, mostly chosen between 30° and 60°.

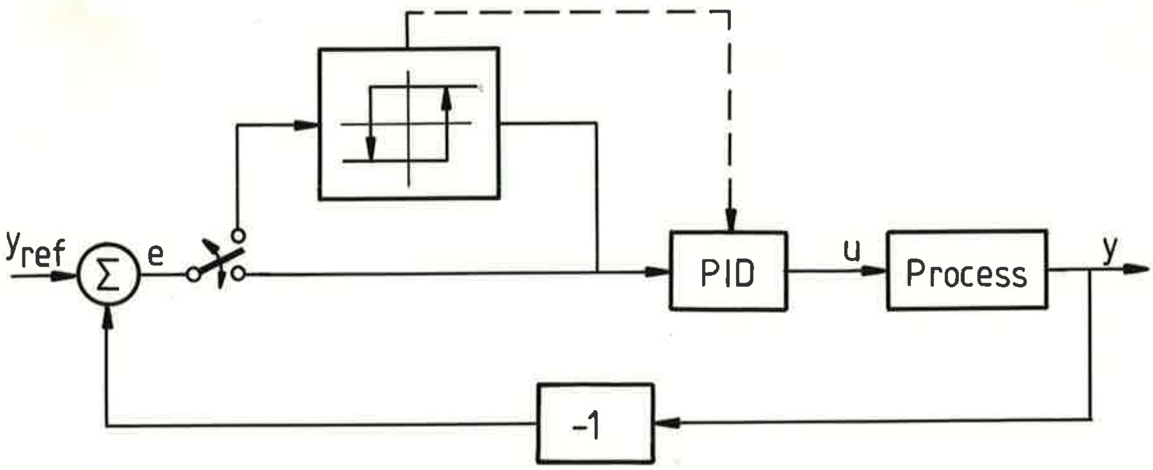

Figure 1.1. The principle for the PID-tuner.

2. THE ESTIMATION OF G(iω)

If a relay with hysteresis is introduced in a closed-loop system, as in fig. 2.1, the signals in the system will oscillate with constant amplitude and frequency, provided that the linear part, G(s), fulfills certain requirements. By measuring the amplitude and the frequency of y, it is possible to determine the position of the frequency curve of G(s) for that certain frequency. This knowledge will be the basis for the tuning rule described in the next chapter.

In section 2.1, the describing function technique is briefly reviewed for a relay with hysteresis. It is also shown how G(s) can be determined, by the knowledge of the amplitude of the oscillation and the characteristics of the relay with hysteresis, at the oscillation frequency. There are several ways to estimate the amplitude and the frequency of a signal. Some of these are discussed in section 2.2.

2.1 The describing function technique for a relay with hysteresis

The describing function for a nonlinear element is defined as the complex valued ratio of the fundamental component of the output to the input, when the input is a sine wave, see e.g. Graham and McRuer(1961). Suppose, that the relay with hysteresis has the characteristics shown in figure 2.2, and that the input to the relay with hysteresis is a sine wave with amplitude A and angular frequency ω. Then the

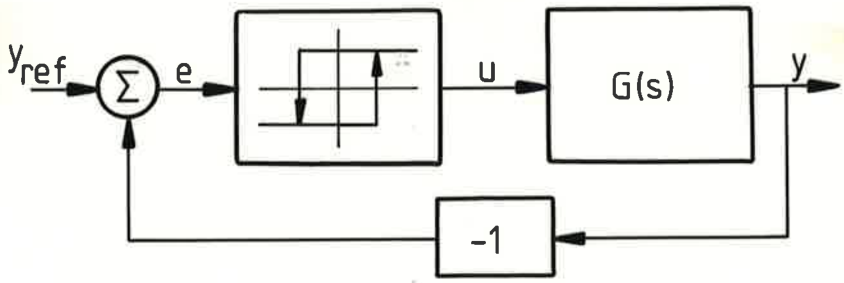

Figure 2.1. A closed loop system with a relay with hysteresis.

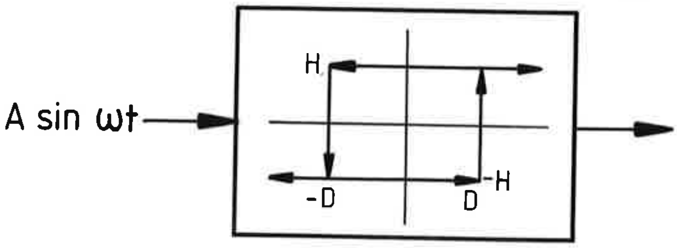

Figure 2.2: The relay with hysteresis.

describing function N(A) is

$$N(A) = \frac{4H}{\pi A} e^{-i\phi} ; \qquad \phi = \arcsin(\frac{D}{A}).$$

It is assumed that $A \geq D$.

The input u to G(s) in figure 2.1 is a square wave. In the describing function technique it is assumed that the input to the relay with hysteresis is a sine wave. If G(s) is a low-pass filter, this will almost be the case when the loop is closed, as in figure 2.1. From now on, it is therefore assumed that G(s) has a low-pass filter action, so that the amplitudes of the high frequencies in y are small compared to the amplitude of the fundamental frequency. This is not a restrictive assumption, since almost all practical processes are of low-pass type.

The frequency curve of G(s) is usually plotted in the complex plane together with the negative inverse describing function. The negative inverse of N(A) is given by

$$-\frac{1}{N(A)} = -\frac{\pi A}{4H} e^{i\phi} = -\frac{\pi A}{4H} (\cos\phi + i\sin\phi) =$$

$$= -\frac{\pi}{4H}\sqrt{A^2 - D^2} - i\frac{\pi D}{4H} ; \qquad A \geq D. \qquad (2.1)$$

Obviously, the imaginary part of $-1/N(A)$ is constant, i.e. the curve will be a straight line, parallel to the real axis. The real part of $-1/N(A)$ will decrease as A increases.

Oscillations in the closed loop system will occur if there are intersections of the $G(i\omega)$ and $-1/N(A)$ curves. The values of the amplitude and frequency parameters at an intersection give the amplitude and the frequency of the oscillation.

When the nonlinearity is a relay with hysteresis, the oscillation at an intersection is stable if and only if the function $\arg[G(i\omega)]$ is decreasing at this point. E.g. in the case shown in figure 2.3, it is possible to get a stable oscillation with amplitude and frequency corresponding to either intersection a or c, but not any corresponding to b.

Since the amplitude and the frequency of the oscillation are given by the parameters at the intersection of the curves, it is possible to determine $G(i\omega)$ at the oscillation frequency. From equation (2.1)

$$G(i\omega) = -\frac{\pi A}{4H} e^{i\phi} = \frac{\pi A}{4H} e^{i(-\pi+\phi)} \quad ; \quad \phi = \arcsin(\frac{D}{A}).$$

2.2 Estimation of amplitude and frequency

An essential part of the estimation of $G(i\omega)$, is the determination of the frequency and the amplitude of the output y. This is a familiar problem, with several well-known solutions. Some of them will be stated and shortly commented here.

The choice of identification method is a weighting between at one side, the demand for a simple algorithm which

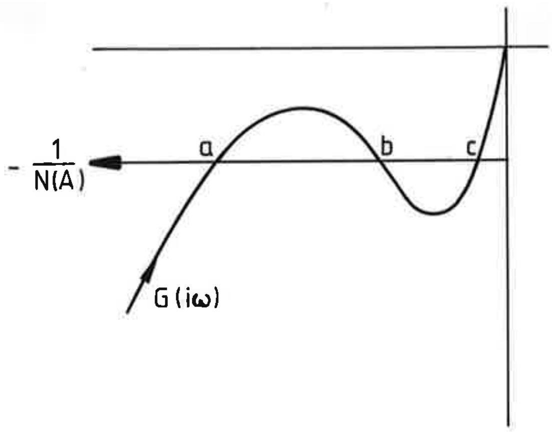

Figure 2.3: An example of a frequency curve $G(i\omega)$ and the negative inverse describing function $-1/N(A)$.

requires little space in the computer, and at the other side the demand for a fast and precise algorithm.

Estimation of the amplitude

The simplest way to estimate the amplitude is probably to compare all the measured values and let the one with the greatest magnitude determine the amplitude. This method is sensitive to disturbances. It is also ineffective, since it does not use the information given by all the measurements except the one with the greatest magnitude. If the algorithm is suggested for a computer with very small memory, it may however be a good choice.

A more complicated method is the recursive least squares identification. Here the function

$$\sum (y(t) - A_s \sin\omega t - A_c \cos\omega t)^2$$

is minimized with respect to A_s and A_c. The amplitude is then given by

$$A = \sqrt{A_s^2 + A_c^2} \, .$$

The procedure requires an estimate of the frequency ω. The frequency estimation must therefore run for a while before the initiation of the amplitude estimation. The estimate A will oscillate with the frequency ω. It is only supposed to have a correct value when $\omega t = n\pi$, $n = 1,2,\ldots$, see Åström(1975). Since all the measurements are used to estimate the amplitude, the method is less sensitive to disturbances than the previous one. On the other hand, it requires more space in the computer.

Another difference between these two methods is, that the first one estimates the true amplitude of the output, while the second one estimates the amplitude of the fundamental component. If the process G(s) was an ideal low pass filter, there would not be any difference. Since in practice the higher frequencies form a disturbance on the amplitude, there is a difference. However, it is not possible to conclude which choice of amplitude that gives the best tuner in the end.

A third way to estimate the amplitude, is to use a Kalman filter. The output is here modelled as a second order system, an oscillator. The system is

$$\begin{bmatrix} x_1(t+h) \\ x_2(t+h) \end{bmatrix} = \begin{bmatrix} \cos\omega h & \sin\omega h \\ -\sin\omega h & \cos\omega h \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} (y(t)-x_1(t)-x_2(t))$$

where h is the sampling period, and k_1 and k_2 are the constant gains in the filter. The amplitude A, is given by

$$A = \sqrt{2(x_1^2 + x_2^2)} \; .$$

<u>Estimation of the frequency</u>

In Lindgren(1974) some methods for frequency estimation are compared.

The simplest method is the zero crossing method. The time between the zero crossings of the output gives an estimate of the oscillation period, and thus of the frequency. The method is simple and therefore favourable in small computers. It can be improved by an additional use of nonzero levels.

It is also possible to estimate the frequency by recursive least squares identification. Here the function

$$\sum (y(t) - 2\cos(\hat{\omega}h)y(t-h) + y(t-2h))^2$$

is minimized with respect to $\hat{\omega}$, where h is the sampling period and $\hat{\omega}$ is the estimate of the frequency ω. This method is superior to the one above, if the identification time is short. As time increases, it is depending on the type of noise which method is the best, Lindgren(1974).

3. THE TUNING RULE

In the previous chapter, a method for estimating $G(i\omega)$ for some frequencies was described. This possibility allows a procedure to tune the parameters in the controller so that the system gets a desired phase margin. The PID-controller in figure 1.1 has the wellknown structure

$$u = K \left(e + \frac{1}{T_I} \int e \, dt + T_D \frac{de}{dt}\right).$$

During the tuning phase, only the proportional part is present. The integral time is afterwards set to

$$T_I = \frac{T}{\nu \cdot 2\pi} \tag{3.1}$$

where T is the final oscillation period. ν is a constant, usually chosen between 0.1 and 0.2. The derivative part is not used in this design.

Let φ_m denote the desired phase margin. When the integral part is introduced in the controller, the phase margin will decrease with the amount φ_I. From (3.1) it is concluded that

$$\varphi_I = 90° - \arctan\left(\frac{2\pi T_I}{T}\right) = \arctan(\nu).$$

The tuning rule is therefore to adjust the gain K so that the phase margin for $KG(i\omega)$ becomes $\varphi_m + \varphi_I$, and then to introduce the integral part. If K is chosen in this way, and T_I given by (3.1) is introduced afterwards, the system will get the phase margin φ_m.

As was said before, the system will oscillate during the tuning phase. It is depending on the actual process, how high amplitudes that can be tolerated. It is possible to influence the amplitudes, by specifying an amplitude A^*, which is the amplitude corresponding to the correct gain K. If noise is present, a large A^* will improve the tuning procedure, since the signal to noise ratio then becomes large.

The adjustment of the gain K will be deduced from the amplitude estimates.

The characteristics of the relay with hysteresis

The gain K is to be adjusted so that the KG(iω) curve goes through the point p in figure 3.1.a. To make it possible to decide when this point is reached, the $-1/N(A)$ curve must also go through this point, as in figure 3.1.b.

Suppose, that the desired amplitude of the output signal at the point p is A^*. Equation (2.1) then gives

$$|-1/N(A^*)| = \frac{\pi A^*}{4H} = 1 \quad \Rightarrow \quad H = \frac{\pi A^*}{4}. \tag{3.2}$$

D is determined from

$$\varphi_m + \varphi_I = \arcsin\left(\frac{D}{A^*}\right) \quad \Rightarrow \quad D = A^* \sin(\varphi_m + \varphi_I). \tag{3.3}$$

The characteristics of the relay with hysteresis are thus given by the desired phase margin, the relation between T_I and T and by the desired amplitude of the output signal.

The position of the relay with hysteresis

The output from the relay with hysteresis is a square wave. The describing function technique requires that no zero-order harmonic is present, i.e. the times when the output is +H are as long as the times when it is −H. Let

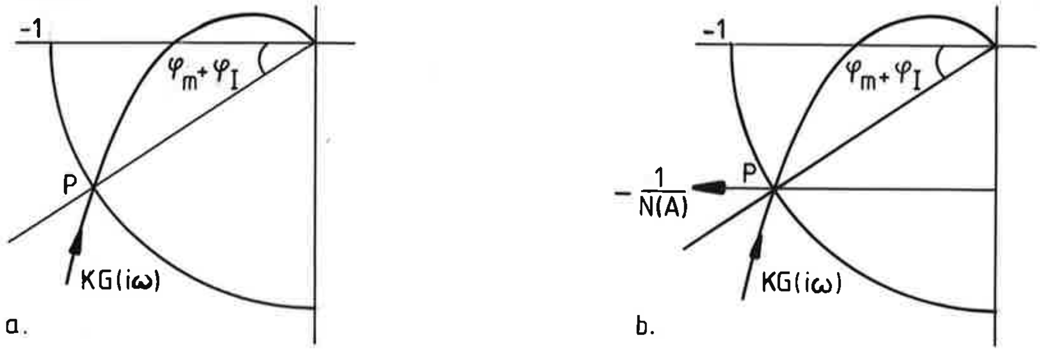

Figure 3.1. a. The desired location of the KG(iω) curve.
b. The correct location of the −1/N(A) curve.

u_{ref} denote the input to the process G(s) which gives the desired output y_{ref}. Then the output will be free from the zero-order harmonic only if the relay with hysteresis is chosen as in figure 3.2. Before the adjustment of the gain K starts, the reference control signal, u_{ref}, must therefore be estimated. This can be done e.g. by measuring the times T_+ and T_-, defined in figure 3.3, of the output from the relay with hysteresis.

Introduce P(t) as

$$P(t) = \frac{T_-(t)}{T_+(t)} .$$

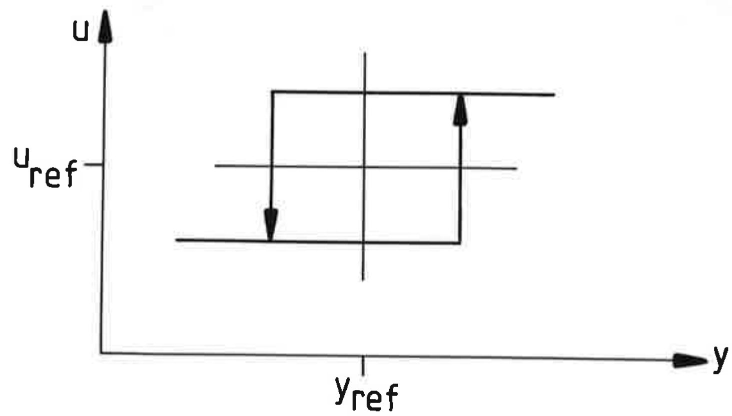

Figure 3.2. The correct position of the relay with hysteresis.

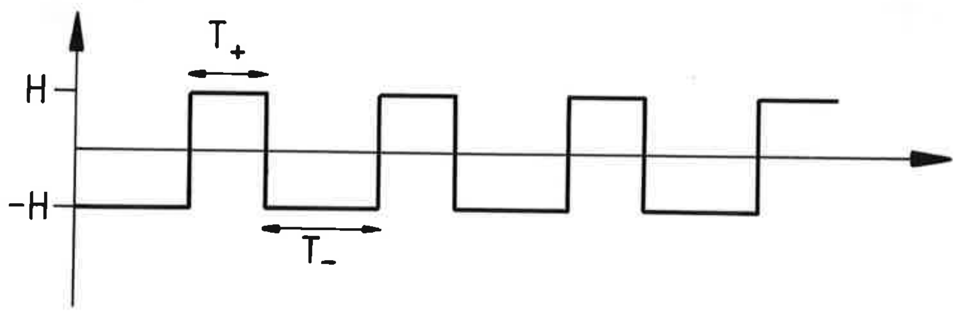

Figure 3.3. The output from the relay with hysteresis.

u_{ref} can now be estimated by the Regula falsi method

$$u_{ref}(t+1) = u_{ref}(t) - (P(t) - 1) \frac{u_{ref}(t) - u_{ref}(t-1)}{P(t) - P(t-1)}. \qquad (3.4)$$

Remark: When the gain K is introduced in the loop, u_{ref} must be changed to u_{ref}/K.

Adjustment of the gain K

When both the characteristics and the position of the relay with hysteresis are determined, it is time for the tuning procedure. As was said before, the goal is to adjust the gain K so that the KG(iω) curve goes through the point p as in figure 3.1. One useful method is to use the two latest measurements of KG(iω), and approximate the KG(iω) curve by a straight line going through these points. The new value of K can then be chosen by the Regula falsi method, i.e.

$$K_{n+1} = K_n - (A_n - A^*) \frac{K_n - K_{n-1}}{A_n - A_{n-1}} \qquad (3.5)$$

Here A_n and A_{n-1} are the two latest amplitudes of the output signal, y.

The complete tuning rule can be summarized by the following scheme.

1. Calculate H and D from (3.2) and (3.3).
2. Insert the relay with hysteresis in the loop. Remove the integral- and derivative part of the controller.
3. Let (3.4) be performed until an acceptable estimate of u_{ref} is obtained.
4. Measure the output during some oscillation periods, and estimate the frequency and the amplitude.
5. If A_n is close to A^*, go to 7.
6. Adjust the gain K according to (3.5). Change the reference control signal to u_{ref}/K, and go to 4.
7. Switch on the integral part of the controller according to (3.1).
8. Remove the relay with hysteresis.

4. AN EXAMPLE

The identification method and the tuning rule are here illustrated with an example. The process is given by the transfer function

$$G(s) = \frac{1}{s(s+1)(s+2)} .$$

For simplicity, both y_{ref} and u_{ref}, defined in chapter 3, are equal to zero. Let the desired phase margin be 30°, and ν in equation (3.1) be 0.1.

The frequency of the oscillation is estimated by the zero crossing method, and the amplitude is estimated by the least squares method.

The result of a simulation is shown in figure 4.1. Before each change of the gain K, the amplitude and frequency estimations are interrupted. After one oscillation period, the frequency estimation starts. After one additional period, the amplitude estimation starts.

Since the two latest measurements are used in the updating of K, see equation (3.5), the second change in K is supposed to be rather efficient. It is the case in this simulation, and the result is also verified by other simulations.

The estimated parameters in the method do not converge exactly to the true values. The final gain and frequency in the example are 1.77 and 0.68 rad/sec respectively, while the correct values are 1.81 and 0.70 rad/sec. This is not surprising, since the method is an approximative one. In this case, the phase margin becomes 31.3° instead of 30.0°. In a practical system, such a small difference can be considered to be completely negligible in comparison to other approximations.

In figure 4.2, the $KG(i\omega)$ curves for the different K:s are presented. The efficiency of the second correction of K is again obvious.

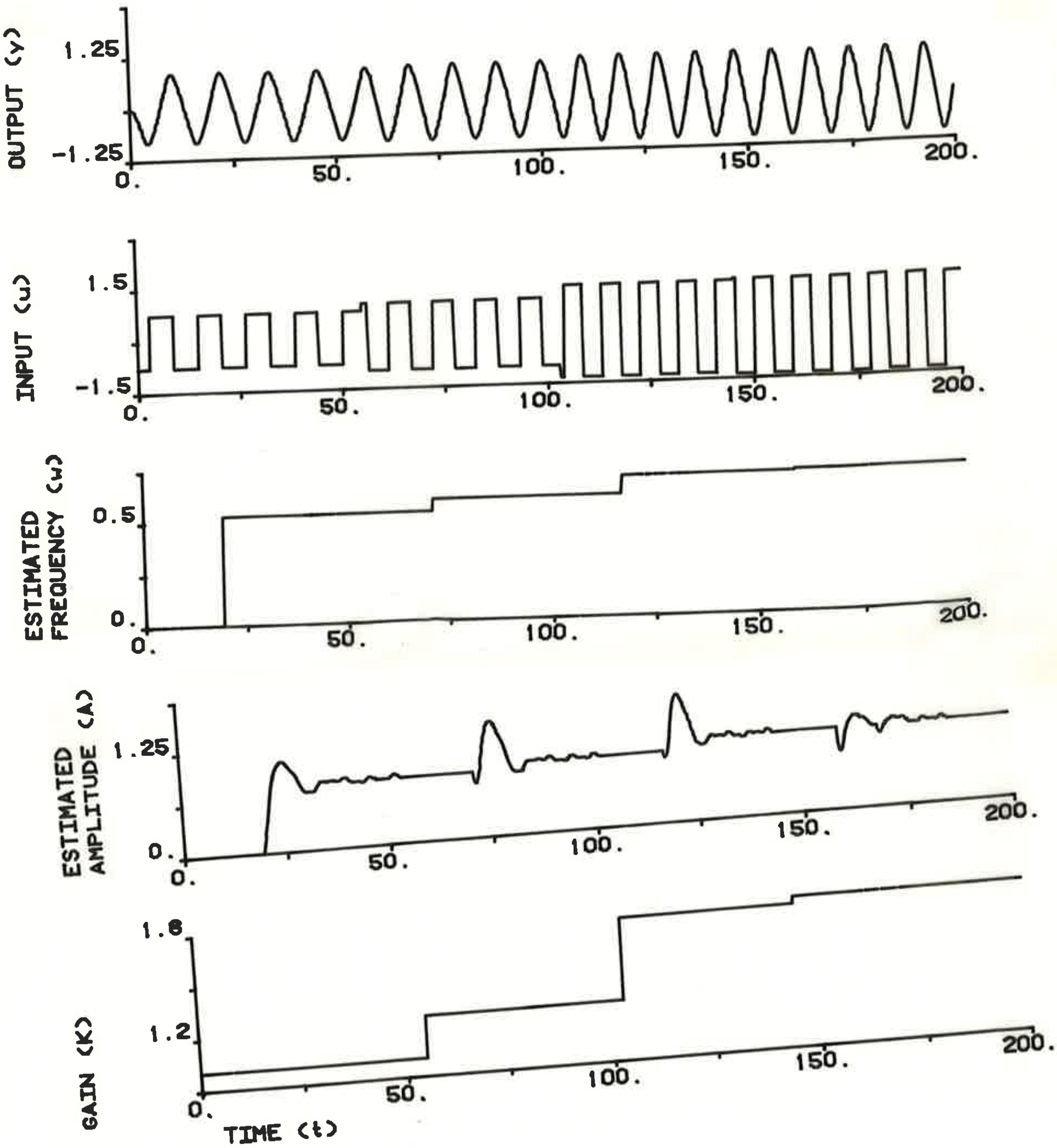

Figure 4.1. The result of the simulation.

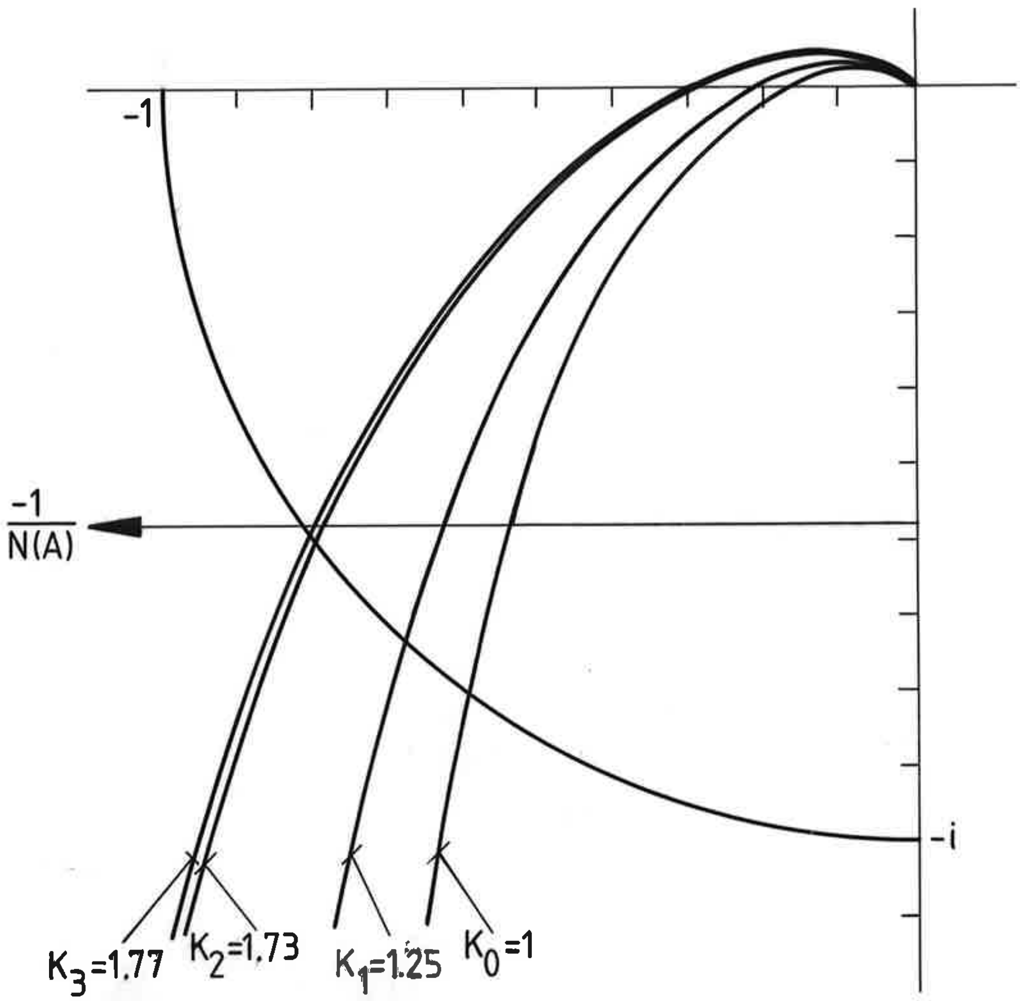

Figure 4.2. The KG(iω)-curves in the example.

5. PRACTICAL ASPECTS

The tuning procedure, described in the previous chapters, only uses the proportional and the integral parts of the PID-controller. These parts are sufficient to reach the goal for the design - to get a desired phase margin in the system. It is however possible to use the derivative part in the procedure too. It can be incorporated by a rule of thumb in the same way as the integral part.

In the estimation and tuning procedures, there are several parameters that must be chosen. It is a matter of taste which ones that shall be free for the operator to choose, and which ones that shall be fixed in the program. A natural choice is perhaps to let φ_m, A^* and ν be free to choose, and let the rest be fixed.

Instead of A^*, i.e. the final amplitude of the oscillation of the output, one would perhaps like to define a parameter A_{max}, which is the greatest permitted amplitude of the output during the tuning phase. This is possible to incorporate by some extra logic in the program.

As was said before, this PID-tuner is supposed to work even in very small computers. If the computer is large enough, it is possible to add extra facilities, like the A_{max} specification, to the program.

Finally, it should also be remarked that the relay with hysteresis can be used as a design help, without any automatic tuning. If the relay with hysteresis is introduced in the loop, and the signals are measured and presented e.g. on an oscilloscope, an operator can adjust the gain K by hand to get the desired phase margin.

6. ACKNOWLEDGEMENTS

I wish to thank Professor Karl Johan Åström, Per Hagander, Carl Fredrik Mannerfelt and Sven Erik Mattsson for many valuable advices and suggestions.

The work was partially supported by the Swedish Board of Technical Development (STU) under contract 78-3763.

7. REFERENCES

Åström K J (1975): Lectures on system identification - Frequency response analysis. Report TFRT-3093, Dept of Automatic Control, Lund Institute of Technology, Lund, Sweden.

Åström K J (1981): More STUPID. Report TFRT-7214, Dept of Automatic Control, Lund Institute of Technology, Lund, Sweden.

Graham D and McRuer D (1961): Analysis of nonlinear control systems. John Wiley & Sons, New York.

Lindgren G (1974): Spectral moment estimation by means of level crossings. Biometrica 61, 401-418.

Schuck O H (1959): Honeywell's history and philosophy in the adaptive control field. In Gregory (ed) Proc. Symposium on Adaptive Control. Wright Patterson Air Force Base 1959. Report WADC TR 59-49.

Wittenmark B, Hagander P and Gustavsson I (1980): STUPID - Implementation of a self-tuning PID-controller. Report TFRT-7201, Dept of Automatic Control, Lund Institute of Technology, Lund, Sweden.