Should we forget the Smith Predictor?

Chriss Grimholt Sigurd Skogestad*

Abstract: The PI/PID controller is the most used controller in industry. However, for processes with large time delays, the common belief is that PI and PID controllers have sluggish performance, and that a Smith predictor or similar dead-time compensator can give much improved performance. We claim in this paper that this is a myth. For a given robustness level in terms of the peak sensitivity (M_{S}), we find that the performance improvement with the Smith predictor is small even for a pure time delay process. For other first-order processes a PID controller is generally better for a given robustness level. In addition, the Smith Predictor is much more sensitive to time delay errors than PI and PID controllers.

Keywords: PI, PID, Smith Predictor, Performance-Robustness trade-off

1. INTRODUCTION

We find time delays in most industrial processes $G(s)$, see Figure 1. Time delay is an important aspect to consider when applying feedback control because it imposes serious limitations on the performance (Skogestad and Postlethwaite, 2005). For the controller $K(s)$, we mostly use the proportional-integral (PI) controller, which is the workhorse of the process industry with more that 95% of all control application being of this type (Åström et al., 1995). For integrating processes with large time delays, PI control is somewhat sluggish, and to impove performance we can add derivative action, i.e. using PID control (Grimholt and Skogestad, 2013).

An alternative to PID control, is to make use of a Smith predictor (SP), also known as a dead time compensator, (Smith, 1957). The SP uses a model of the process without a time delay to predict the process output, and this new process is controlled by a conventional controller, for example, a PI controller. See Figure 2. The SP controller has good setpoint response because it removes the internal delays from the closed-loop transfer function. One drawback is that it has poor performance for input (load) disturbances for processes with slow dynamics because the original open-loop process poles remain unchanged. However, this can be rectified by alternative designs, e.g. see Normey-Rico and Camacho (2007).

(Kristiansson and Lennartson, 2001; Ingimundarson and Hägglund, 2002; Larsson and Hägglund, 2012) have compared the performance of PI, PID, and SP controllers. These papers investigate performance for load disturbances for a fixed robustness. Most of these papers conclude that PID has better performance than SP, and our work further confirms this.

In this paper, we examine the optimal performance-robustness trade-off for a SP and compare it to the optimal trade-off for PI and PID with the same robustness level. For performance we consider an average integrated absolute error (IAE) performance of step disturbances at the process input and output. The disturbance at the process output is a special case of set-point response. Robustness is quantified by the peak of the sensitivity and complimentary
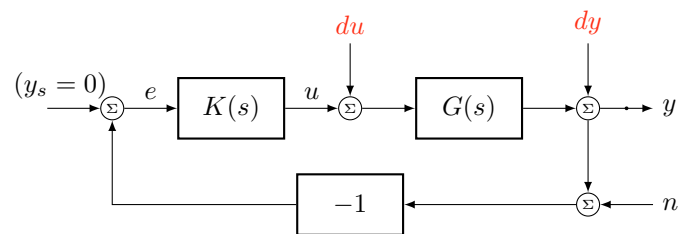

Fig. 1. Block diagram of the feedback control system. We may treat an output disturbance (d_y) as a special case of setpoint change (y_s)

sensitivity function (M_{S} and M_{T}). In addition, we consider time delay robustness which is not captured by M_{S} and M_{T}.

2. THE FEEDBACK SYSTEMS

We consider a range of first-order plus delay (FOPD) processes

$$G = \frac{ke^{-\theta s}}{\tau s + 1} = G_{\circ}e^{-\theta s}, \qquad (1)$$

where k, τ, and θ is the gain, time constant, and time delay of the process, respectively. The process without time delay is represented with $G_{\circ}$.

2.1 PID control

There are serveral different parametrizations of the PI and PID controller. For optimization purposes, we use the following linear parametrization

$$K_{\mathrm{PI}} = k_p + k_i/s \quad \text{and} \quad K_{\mathrm{PID}} = k_p + k_i/s + k_d s, \qquad (2)$$

We can transform the controller (2) to the standard parallel ("ideal") PID controller,

$$K_{\mathrm{PID}}^{\mathrm{parallel}} = k_c \left(1 + \frac{1}{\tau_i s} + \tau_d s\right) \qquad (3)$$

by the following transformation

$$k_c = k_p, \quad \tau_i = {}^{k_p}/k_i, \quad \text{and} \quad \tau_d = {}^{k_d}/k_p. \qquad (4)$$

2.2 The Smith Predictor

In this paper we consider the "original" Smith predictor controller

$$K_{\mathrm{SP}} = \frac{K_\circ(s)}{1 + K_\circ(s)G_\circ(s)\left(1 - e^{-\theta s}\right)} \qquad (5)$$

where $K_\circ$ in the primary conventional controller, which in this paper is a PI controllers, and $G_\circ$ and G are the internal delay-free and delayed models, respectively. A block diagram of the SP for the case where $K_\circ$ is a PI controller is shown in Figure 2.

The main advantage of the SP is the potential excellent septpoint response. However, because it is impossible to eliminate the open-loop poles in the input disturbance (load disturbance) transfer function, the "original" SP has slow settling time for for input disturbances (load disturbances). It is possible to improve this by using a modified SP as discussed later.

3. QUANTIFYING THE OPTIMAL CONTROLLER

3.1 Performance

In this paper we choose to quantify performance in terms of the IAE,

$$\mathrm{IAE} = \int_0^\infty \left| y(t) - y_s(t) \right| \mathrm{d}t. \qquad (6)$$

To balance the servo/regulatory trade-off, we choose as the performance index a weighted average of IAE for a step input disturbance d_u and step output d_y,

$$J(p) = 0.5 \left(\frac{\mathrm{IAE}_{dy}(p)}{\mathrm{IAE}_{dy}^\circ} + \frac{\mathrm{IAE}_{du}(p)}{\mathrm{IAE}_{du}^\circ} \right) \qquad (7)$$

where IAE_{dy}° and IAE_{du}° are weighting factors, and p is the controller parameters. In this paper, we select the two weighting factors as the optimal IAE values when using PI control, for input and output disturbances, respectively (as recommended by Boyd and Barratt (1991)). To ensure robust reference PI controllers, they are required to have $M_{\mathrm{S}} = 1.59$ [1], and the resulting weighting factors are given for four processes in Table 1.

It may be argued that a two-degree of freedom controller with a setpoint filter can be used to enhance setpoint performance, and thus we only need to consider input disturbances. But note that although a change on the

[1] For those that are curious about the origin of this specific value $M_{\mathrm{S}} = 1.59$, it is the resulting M_{S} value for a Simple Internal Model Control (SIMC) tuned PI controller with $\tau_c = \theta$ on first-order plus time delay (FOPTD) process with $\tau \leq 8\theta$.
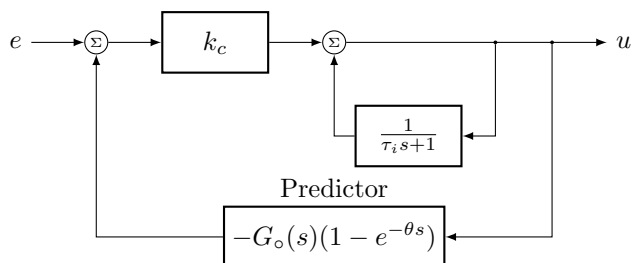

Fig. 2. Block digram of the SP controller, K_{SP}, when $K_\circ$ is a PI controller.

output d_y, is equivalent to a setpoint change y_s for the system in Figure 1, it is not affected by a setpoint filter. Thus, we consider disturbance rejection which can only be handled by the feedback controller $K(s)$ (Figure 1).

3.2 Robustness

In this paper, we quantify robustness in terms of M_{ST}, defined as the largest value of M_{S} and M_{T} (Garpinger and Hägglund, 2008),

$$M_{\mathrm{ST}} = \max\{M_{\mathrm{S}}, M_{\mathrm{T}}\}. \qquad (8)$$

where M_{S} and M_{T} are the largest peaks of the sensitivity $S(s)$ and complimentary sensitivity $T(s)$ functions, respectively. Mathematically,

$$M_{\mathrm{S}} = \max_\omega \left| S(j\omega) \right| = \left\| S(j\omega) \right\|_\infty,$$
$$M_{\mathrm{T}} = \max_\omega \left| T(j\omega) \right| = \left\| T(j\omega) \right\|_\infty,$$

where $\|\cdot\|_\infty$ is the H_∞ norm (maximum peak as a function of frequency), and the sensitivity transfer functions are defined as

$$S(s) = {1}/{(1+G(s)\,K(s))} \quad \text{and} \quad T(s) = 1 - S(s). \qquad (9)$$

For most stable processes, $M_{\mathrm{S}} \geq M_{\mathrm{T}}$. For a given M_{S} we are guaranteed the following gain margin (GM) and phase margin (PM),

$$\mathrm{GM} \geq \frac{M_{\mathrm{S}}}{M_{\mathrm{S}} - 1} \quad \text{and} \quad \mathrm{PM} \geq 2\arcsin\left(\frac{1}{2M_{\mathrm{T}}}\right) \geq \frac{1}{M_{\mathrm{T}}}. \qquad (10)$$

For example, $M_{\mathrm{S}} = 1.6$ guarantees GM ≥ 2.7 and PM $\geq 36.4° = 0.64$ rad. Another important robustness measure is the delay margin (DM), (Åström and Hägglund, 2006),

$$\mathrm{DM} = \frac{\mathrm{PM}}{\omega_c} \qquad (11)$$

where ω_c is the crossover frequency. Note that the units for PM [rad] and ω_c [rad/s] must be consistent. The delay margin is the smallest change in time-delay that will cause the closed-loop to becomes unstable. We will see for the SP, that robustness in term of M_{ST} does not guarantee robustness in term of DM.

4. OPTIMAL CONTROLLER

For a given first-order plus delay process, the IAE-optimal PI, PID or SP controller are found for a specified robustness level by solving the following optimization problem:

$$\min_p \quad J(p) = 0.5 \left(\frac{\mathrm{IAE}_{dy}(p)}{\mathrm{IAE}_{dy}^\circ} + \frac{\mathrm{IAE}_{du}(p)}{\mathrm{IAE}_{du}^\circ} \right) \qquad (12)$$

$$\text{subject to:} \quad M_{\mathrm{S}}(p) \leq M^{ub} \qquad (13)$$
$$M_{\mathrm{T}}(p) \leq M^{ub} \qquad (14)$$

where the two or three parameters in p are for a PI or PID controller. Our Smith predictor controller always uses a PI controller (two parameters in p, see Figure 2) For more details on how to solve the optimization problem, see Grimholt and Skogestad (2018). The problem is solved repeatedly for different values of M^{ub}. One of the bounds in (13) or (14) will be active if there is a trade-off between robustness and performance. This is the case for values of M^{ub} less than about 2 to 3.

Table 1. Cost function weights and optimal controllers with $M_{\mathrm{ST}} = 1.59$.

Process	Weights in J			Optimal PI					Optimal SP				
	IAE°_{dy}	IAE°_{du}	M_{ST}	k_c	τ_i	IAE_{dy}	IAE_{du}	J	k_c	τ_i	IAE_{dy}	IAE_{du}	J
e^{-s}	1.61	1.61	1.59	0.20	0.32	1.61	1.61	1.00	0.73	0.32	1.45	1.45	0.90
$e^{-s}/(s+1)$	2.07	2.02	1.59	0.54	1.10	2.08	2.04	1.01	1.37	0.93	1.69	1.68	0.83
$e^{-s}/(8s+1)$	2.17	1.13	1.59	3.47	4.04	3.10	1.16	1.23	9.94	3.35	2.04	1.38	1.08
$e^{-s}/(20s+1)$	2.17	0.60	1.59	8.42	5.16	3.67	0.61	1.36	22.7	3.31	2.34	1.12	1.47

IAE_{dy} and IAE_{du} are for a unit step disturbance on output (y) and input (u), respectively.

5. OPTIMAL TRADE-OFF

In this section, we present the optimal IAE-performance for PI, PID and SP controllers as a function of the robustness level M_{ST}. We have considered four FOPD process models, defined in (1),

Pure time delay: $\tau/\theta = 0$
Balanced dynamics: $\tau/\theta = 1$
Lag dominant: $\tau/\theta = 8$
Close to integrating: $\tau/\theta = 20$

We have not included results for an integrating process model because SP does not work for integrating processes.

5.1 Performance

In Figure 3, we show the Pareto optimal IAE-performance (J) as a function of robustness (M_{ST}) for the four processes.

For the pure time delay process (top left), there is only a small difference between the optimal PI/PID and SP controllers. For very robust controllers with $M_{\mathrm{ST}} = 1.1$, the improvement with SP is only 2%. For robust controllers with $M_{\mathrm{ST}} = 1.69$, the improvement is 11%. This small improvement is surprising, because SP was expected to substancially improve performance for delay-dominant proceeses.

Note that we write optimal "PI/PID" controller. This is because for a pure time delay process there is no advantage in using derivative action, so the optimal PID is a PI controller. Note that the maximum value of M_{ST} is 1.92 for PI and PID control. Performance (J) actually gets worse if M_{ST} is increases beyond this value, and this region should be avoided.

For balanced and lag dominant dynamics (top right and bottom left in Figure 3), SP has somewhat better performance than PI. For a robust controller with $M_{\mathrm{ST}} = 1.69$, the SP has 23% and 17% improved performance, respectively. However, the PID controller is even better, with 27% and 33% improvement relative to the PI controller.

For a close to integrating process (bottom right Figure 3), SP has worse performance than the PI controler. This is because the "original" SP retains the large time constant from the open-loop process model (Hägglund, 1996). For a robust controller with $M_{\mathrm{ST}} = 1.69$, the SP controller has -8% worse performance than PI. On the other side, the PID controller has an improved performance of 35%.

In summary, if we consider performance for a given robustness level in terms of M_{ST}, the PID controller is better than the SP with PI, except for a pure time delay process where

the potential improved performance is marginal, especially for the interesting cases with M_s less than 1.6.

5.2 Delay margin (DM)

In the above discussion the trade-off between IAE performance and M_{ST} robustness, we did not consider the delay error which for SP is not captured by the M_{ST} value and which is actually the main disadvantage of the Smith predictor. The delay margins corresponding to Figure 3 are shown in Figure 4. For PI and PID control the delay margin follows our M_s robustness measure quite smoothly. This is not the case for the SP, which has a sudden and large drop in DM for higher M_{ST} values. For the four process models in this paper, the drop in DM occurs at M_{ST} equal to 1.63, 1.80, 1.82, and 1.84, respectively.

In pratice, this means that the the SP controller is unusable for higher M_{ST} values. Thus, the SP controller is actually worse than what is shown in Figure 3.

6. DISCUSSION

6.1 Robustness

Palmor (1980) showed that a SP with very high gains will have good GM and PM, but have arbitrary small DM. This agrees with Figure 4. Thus, when designing a Smith predictor using only the classical robustness margins (GM and PM) you easily end up with an aggressive controller with very small DM. Adam et al. (2000) showed that these robustness toward error in time delay can be arbitrary small and unsymmetrical, and this is further discussed below.

Figure 5 shows for the FOPD process

$$G(s) = e^{-s}/(s+1) \qquad (15)$$

the loop function L for a SP synthesized with the IMC method $(\tau_i = \tau)$ with $M_\mathrm{s} = 1.81$. The resulting L has three crossover frequencies, and two regions with magnitude larger than one. The closed-loop becomes unstable if the Bode stability criteria is violated,

$$\left| L(\omega) \right| < 1 \quad \text{for } \omega_{180}, \omega_{540}, \omega_{900}, \dots, \qquad (16)$$

where ω_{180}, ω_{540}, and ω_{900} are the frequencies where the phase is $-180°$, $-540°$, and $-900°$ respectively. Note that the first "sudden drops" in Figure 4 occurs when we get $|L| \geq 1$ at ω between ω_{180} and ω_{540}, which makes ω_c jump to a higher frequency. Further jumps occur as as ω_c jumps to even higher frequencies (Gudin and Mirkin, 2007).

The two phase crossover frequencies ω_{180} and ω_{540} are marked in Figure 5. From the Bode criterion, the system becomes unstable if these frequencies move into the shaded
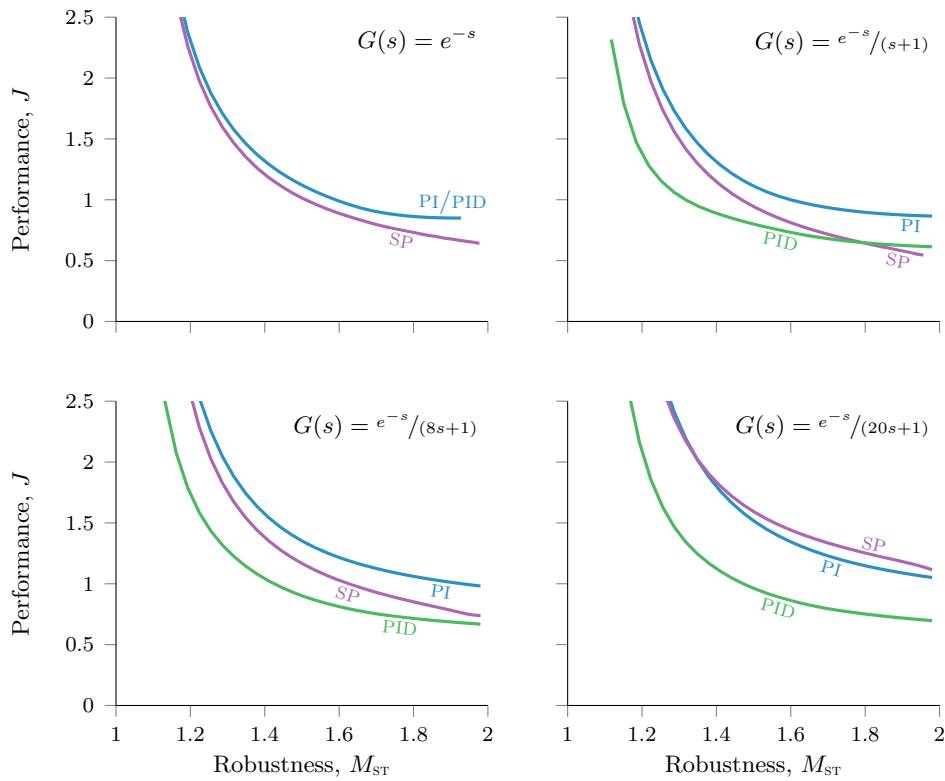
Fig. 3. Pareto-optimal tradeoff between IAE-performance and M_{ST}-robustness for PI, PID and SP control.
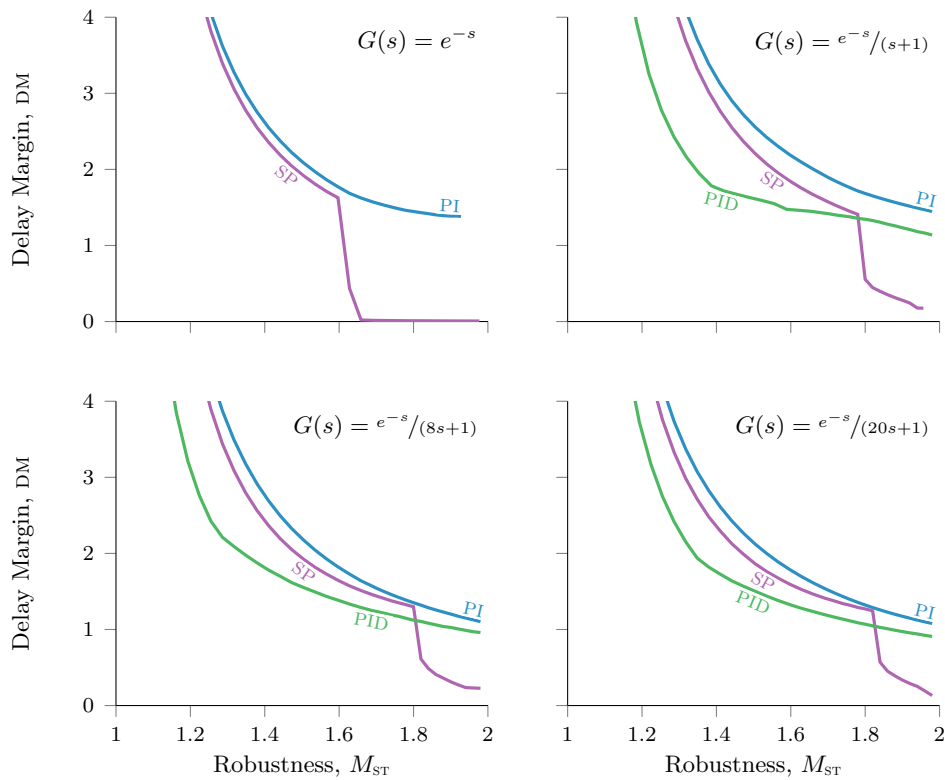

Fig. 4. Delay Margins (DM) for the Pareto-optimal controllers in Figure 3.

regions where the magnitude is larger than one, and these frequencies shift with time delay error, $\Delta\theta$.

Since we can have $|L| > 1$ also for $\omega > \omega_{180}$ with SP, a SP with high gain can become unstable both for positive and *negative* time delay errors. This is not the case for PID which becomes unstable only for positive time delay errors.

For example, for the system in Figure 5 which has a nominal delay of 1, the system with SP becomes unstable for time delays errors $\Delta\theta$ in the intervals,

$$[-0.66, -0.45], \quad [0.48, 0.85], \quad \text{and} \quad [1.38, \infty). \tag{17}$$

Closed-loop responses for a SP and a PI controller tuned for $M_{\text{ST}} = 1.81$ for three values of $\Delta\theta$ are shown in Figure 6. We see that the Smith predictor may start to oscillate both for negative ($\Delta\theta = -0.45$) and positive time delay errors ($\Delta\theta = 0.40$).

One solution to deal with the problem of multiple instability regions, is to limit the loop gain such that it is strictly smaller than 1 at higher frequencies (Ingimundarson and Hägglund, 2002), that is

$$\left|L(\omega)\right| < 1 \quad \text{for } \omega > \omega_c. \tag{18}$$

where ω_c is the first crossover frequency. This will ensure that the system only becomes unstable if the delay error is $\Delta\theta$ is positive, as is the case with PID. However, for system with high loop gain where one of the peaks of $L(\omega)$ is almost one, a small under-estimation of the process gain can bring back the peaks.

Concerned about the conservatism of multiplicative bound for time delay error Larsson and Hagglund (2009) proposed the following robustness bounds,

$$M_{\text{S}_\Delta} = \max_{\Delta\theta \in [\Delta\theta_{\min}, \ \Delta\theta_{\max}]} \left\|S(s, \ \Delta\theta)\right\|_\infty \leq M_{\text{S}_\Delta}^{ub} \tag{19}$$

$$M_{\text{T}_\Delta} = \max_{\Delta\theta \in [\Delta\theta_{\min}, \ \Delta\theta_{\max}]} \left\|T(s, \ \Delta\theta)\right\|_\infty \leq M_{\text{T}_\Delta}^{ub} \tag{20}$$

which ensures a minimum GM and PM even for maximum delay error.

Form this we can conclude that by ensuring sufficiently small M_{S} and M_{T} peaks for SP, we get good robustness against delay error. But this also means that SP will not be able to achieve the lowest IAE values in Figure 3.

6.2 Predictive PI

Another way of avoiding some of these limitations with SP is to use a modified SP. The PPI controller (Hägglund,
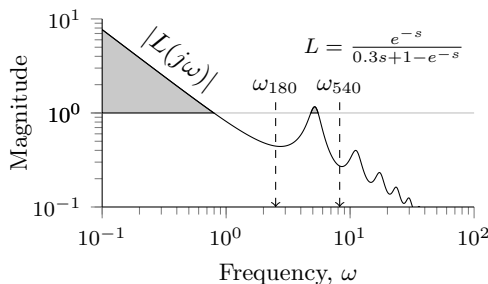
Fig. 5. Bode plot of the SP loop function $L = GK_{\text{SP}}$ for $G(s) = e^{-s}/(s+1)$ and $M_{\text{ST}} = 1.81$.
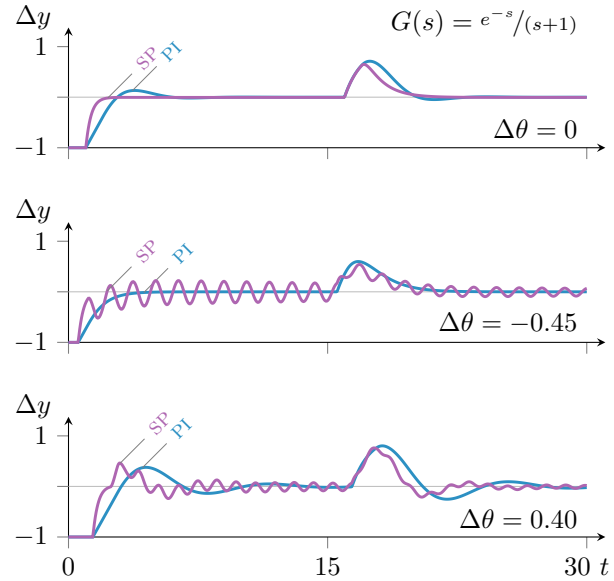
Fig. 6. Step response for PI and SP control with different time delay errors, $\Delta\theta$. That is, the controllers are tuned with $M_{\text{ST}} = 1.81$ for a model with $\theta_m = 1$, and tested on the real process with $\theta = \theta_m + \Delta\theta$.
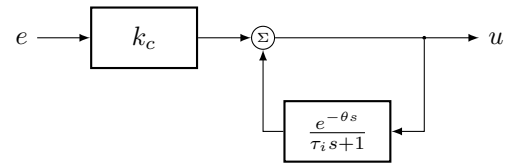
Fig. 7. Block digram of the Predictive PI (PPI) controller

1996) is a special case of the SP, where the parameters of the internal model, k and τ, is parametrized in terms of the controller parameters k_c and τ_i. This is done by assuming a FOPD and using the lambda tuning method to express the model parameters in term of the controller parameters. As a result, the PPI controller only needs 3 parameters to be specified, namely k_c, τ_i, and θ, which is same as the number of parameters needed for PID controller. The delay-free part of the model, $G_\circ$, is reduced and the PPI controller can be expressed as,

$$K_{\text{PPI}} = \frac{k_c \left(1 + \frac{1}{\tau_i s}\right)}{1 + \frac{1}{\tau_i s}(1 - e^{-k_\theta s})}, \tag{21}$$

where k_p, k_i, and k_θ are tuning parameters. The internal controller delay can be treated as a fixed parameter $k_\theta = \theta$ (normally done), or as a free parameter to additionally improve performance. Because of the modification, the PPI controller can achieve better disturbance rejection than the SP because it can avoid the zero-pole cancellation between the controller and process. Also, the PPI controller, unlike the SP, works for integrating processes because it retains integral action for these processes. However, otherwise performance compared to PID control is not improved (Larsson and Hägglund, 2012)

6.3 Tuning of PID controlles

We find in this paper than a PID controller generally outperformes a Smith predictor controller. This is in agree-

Mل