# THE INTERNAL MODEL PRINCIPLE OF CONTROL THEORY

W. M. Wonham, 2018.06.17

## I. Introduction

The Internal Model Principle (IMP) of control theory states (informally) that “a good controller incorporates a model of the dynamics that generate the signals which the control system is intended to track.” Briefly, the controller contains a model of its “exosystem”, or ”outside world.”
While more formal statements have appeared in the control literature starting in the 1970s, these have been developed only within rather specialized frameworks such as linear multivariable systems or certain nonlinear systems defined on smooth manifolds.
Our aim in this note is to develop a version of the IMP in as elementary a setting as possible, namely just that of plain sets and functions. In particular we discuss the necessity of feedback, and how from feedback structure the controller's internal model structure can be naturally derived. While several questions remain open, this setting has the appeal of being universal and readily specialized.

## II. GENERAL BACKGROUND

In psychology and human experience generally the IMP is by no means new. According to Kenneth Craik [1],“[O]nly [an] internal model of reality- this working model [in our minds]- enables us to predict events which have not yet occurred in the physical world, a process which saves time, expense, and even life. [In other words] the nervous system is viewed as a calculating machine capable of modelling or paralleling external events, and this process of paralleling is the basic feature of thought and of explanation.” Mark Twain, apprenticed as a teenager to a Mississippi river pilot, later [2] reports his mentor as saying, “You only learn *the* shape of the river; and you learn it with such absolute certainty that you can always steer by the shape that's *in your head*, and never mind the one that's before your eyes” (italics in original). We shall later indicate what that “shape of the river” might be. Our mimicry of the IMP with automata (or discrete dynamics) is nothing new either; as Thomas Hobbes [3] declared, “For seeing life is but a motion of limbs ... why may we not say, that all automata (engines that move themselves by springs and wheels as doth a watch) have an artificial life?”

## III. Control Background

By the 1930s, thanks to research at the Bell Telephone Laboratories, the mathematical foundation of “classical” linear feedback control based in the frequency domain was soundly established. Its implications were the following. 1. Error feedback (i.e. output feedback followed by the precise differencing of output and input reference signals to form the tracking error) by itself can reduce parameter sensitivity, and final tracking error, but only at the expense of high loop gain. 2. Error feedback, plus an internal model of the reference signal generator, together reduce final tracking error to exactly zero (i.e. ensure perfect tracking), regardless of (reasonable) parameter perturbations, while requiring only moderate average loop gain. A familiar example is the “integrator” component of PID control used to track (specifically) step reference inputs. In general the price to be paid for perfect tracking was extra control complexity, including a stabilizing compensator, specific to the reference signals to be tracked. Later, Otto Smith [4] incorporated an internal model in his scheme of predictive control; while from the 1970s the study of parameter-insensitive perfect asymptotic tracking led to the recognition of both error feedback and the internal model as necessary and sufficient structural features of “robust” linear multivariable systems; for references see Wonham [5]. In this way the step reference generator of PID control was generalized to an arbitrary linear “exosystem” as well as outputs of arbitrary vector dimension.

## IV. The Internal Model Principle

Since the achievement of “robustness” or structural stability of perfect regulation with respect to parameter variations is largely a matter of technology, we shall consider only the converse questions:

Is error feedback a necessary condition for “good” regulation (i.e. perfect tracking)?

Is an internal model a necessary condition for “good” regulation?

If “Yes”, shouldn't these statements hold for a very wide class of regulator systems, linear or nonlinear?

Thus we shall assert as the Internal Model Principle (IMP):

For a very general class of systems:

Assertion 1. Error feedback + Perfect regulation $\Rightarrow$ Internal Model

Assertion 2. Structurally stable (or “robust”) perfect regulation $\Rightarrow$ Error feedback + Internal Model

Author is with the Systems Control Group of the Electrical and Computer Engineering Department, University of Toronto. (e-mail: wonham@ece.utoronto.ca)

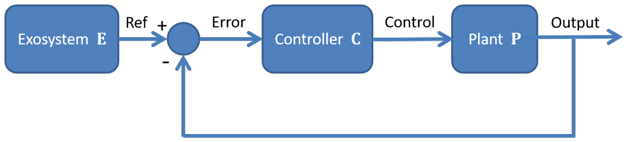



Fig. 1: Total System **S**

Our goal is to establish the IMP in a general but rudimentary discrete-time framework, using just ordinary sets and functions, without any sophisticated technical or geometric machinery. We begin with the easier Assertion 1. Referring to Fig 1, we consider the total system $\mathbf{S}$ = Exosystem $\times$ Controller $\times$ Plant $= \mathbf{E} \times \mathbf{C} \times \mathbf{P}$, with state space the set $X = X_{\mathbf{E}} \times X_{\mathbf{C}} \times X_{\mathbf{P}}$, say. While there's no need to distinguish sharply between Controller and Plant, we do so here for the sake of intuition and control tradition. Bring in the total one-step transition function $\alpha : X \to X$. We consider $X_{\mathbf{C}} \times X_{\mathbf{P}}$ to be $\alpha$-invariant, and $X_{\mathbf{E}}$ the corresponding factor with induced map $\alpha_{\mathbf{E}} : X_{\mathbf{E}} \to X_{\mathbf{E}}$. Thus $(X_{\mathbf{E}}, \alpha_{\mathbf{E}})$ is the dynamic model of the exosystem or "outside world", providing the reference signal for tracking by $\mathbf{C} \times \mathbf{P}$. Think of $\mathbf{E}$ as "driving" $\mathbf{C} \times \mathbf{P}$.

For the total system $\mathbf{S}$ we need to define internal stability, error feedback, and exosystem detectability. For internal stability we assume that $X$ is a finite set and that $(X_{\mathbf{E}}, \alpha_{\mathbf{E}})$ induces an $\alpha$-invariant subset of $X$ via an injection $i_{\mathbf{E}} : X_{\mathbf{E}} \to X$ as shown in the (commutative) diagram of Fig. 2. Thus $\alpha \circ i_{\mathbf{E}} = i_{\mathbf{E}} \circ \alpha_{\mathbf{E}}$ . Write $\widetilde{X}_{\mathbf{E}} := i_{\mathbf{E}}(X_{\mathbf{E}})$ and assume (crucially) that $\widetilde{X}_E$ is a global attractor, namely that, for every initial state $x_o$ in $X$, there is an integer $N$ with $\alpha^n(x_o) \in \widetilde{X}_{\mathbf{E}}$ for all $n \geq N$. [For simplicity we omit the technical details in case $X$ is infinite.] Think of $\mathbf{E}$ as an orchestra and $\mathbf{C} \times \mathbf{P}$ as an attentive but passive audience.

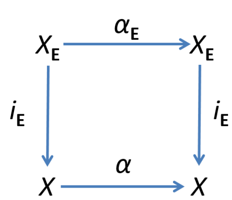



Fig. 2: Internal stability commutative diagram

To define error feedback, first let $K \subseteq X$ be the target subset for regulation: in the standard case $K$ is exactly the subset where tracking error is zero. Also let $\gamma : X \to X_{\mathbf{C}}$ be the natural projection defining state of the controller. Error feedback is then the property that the controller $\mathbf{C}$ is externally driven only when state of $\mathbf{S}$ deviates from the target $K$, namely the dynamics of $\mathbf{C}$ are autonomous as long as $x \in K$, or "tracking remains perfect". Suppose $x \in K$, so the controller state is $x_{\mathbf{C}} = \gamma(x)$. By feedback the next controller state $x'_{\mathbf{C}} = \gamma(\alpha(x))$ depends only on $x_{\mathbf{C}} = \gamma(x)$, i.e. for $x \in K$, $\gamma \circ \alpha(x)$ is computable from $\gamma(x)$. Formally $ker(\gamma|K) \leq ker(\gamma \circ \alpha|K)$, where $ker(.)$ denotes equivalence kernel of the functional argument and $\leq$ means "is a refinement of". Note that $K$ itself need not be $\alpha$-invariant and usually isn't.

Lack of space prohibits formally defining exosystem detectability (see [6] , [7]); just recall that detectability is "local observability" on an invariant subspace, defined here as the property that the global observer congruence for the pair $(\gamma, \alpha)$ reduces to "full observation" (bottom element) on $\widetilde{X}_{\mathbf{E}}$. Intuitively this means that the controller is effectively coupled (via error feedback) to the exosystem, namely the latter is observable by the controller as long as regulation is perfect. This requirement could be dropped by replacing the exosystem by its "observable factor".

With $X_{\mathbf{E}}$ as defined above, write $\widetilde{\alpha}_{\mathbf{E}} := \alpha|\widetilde{X}_{\mathbf{E}}$, $\widetilde{\gamma}_{\mathbf{E}} := \gamma|\widetilde{X}_{\mathbf{E}}$. Now we can prove

**Theorem 1.** Internal Model Principle: Assertion 1 above

Assume that $\mathbf{S}$ satisfies internal stability, perfect regulation, error feedback, and exosystem detectability. Then

1) There exists a unique mapping $\alpha_{\mathbf{C}} : X_{\mathbf{C}} \to X_{\mathbf{C}}$ determined by $\alpha_{\mathbf{C}} \circ \gamma|K = \gamma \circ \alpha|K$
2) $\alpha_{\mathbf{C}} \circ \widetilde{\gamma}_{\mathbf{E}} = \widetilde{\gamma}_{\mathbf{E}} \circ \widetilde{\alpha}_{\mathbf{E}}$
3) $\widetilde{\gamma}_{\mathbf{E}}$ is injective

Statement 1 defines the controller's dynamics, as autonomous under the condition of regulation. Statement 2 identifies these controller dynamics as a copy of the dynamics of E on the global attractor (i.e. exosystem dynamics). Statement 3 asserts that this copy is faithful, namely incorporates fully the exosystem dynamics. The result is shown in the commutative diagram Fig. 3. The proof (omitted) amounts to building up the commutative diagram Fig. 4.

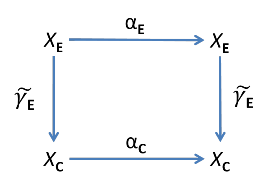



Fig. 3: Commutative diagram for Assertion 1

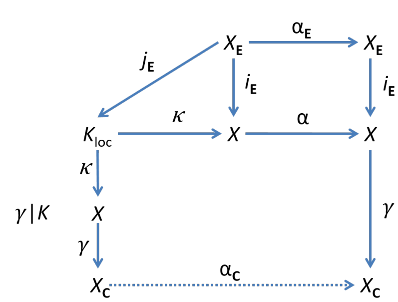
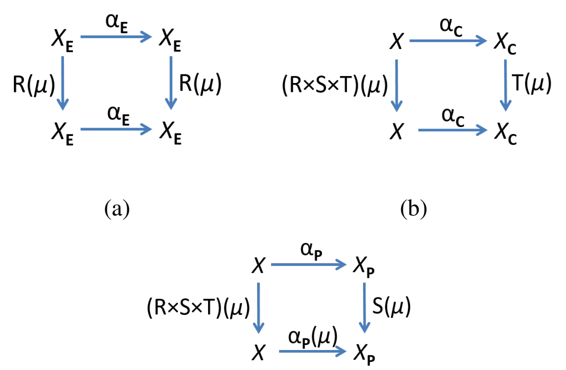



Fig. 4: Commutative diagram for Theorem 1

To formalize Assertion 2 above we enlarge the state structure by forming products with a parameter set M; and require that internal stability, perfect regulation, and exosystem detectability hold for every element $\mu$ in $M$. For realism and to avoid overkill we specialize $M = M_{\mathbf{E}} \times M_{\mathbf{C}} \times M_{\mathbf{P}}, \mu = (\mu_{\mathbf{E}}, \mu_{\mathbf{C}}, \mu_{\mathbf{P}})$, resulting in the commutative diagrams of Fig. 5. The resulting perturbation model (Fig. 6) leads in turn to the equation

$$\alpha_C[R(\mu_{\mathbf{E}})(x_{\mathbf{E}}), S(\mu_{\mathbf{P}}) \circ i_{\mathbf{P}}(x_{\mathbf{E}}), T(\mu_{\mathbf{C}}) \circ i_{\mathbf{C}}(x_{\mathbf{E}})] = T(\mu_{\mathbf{C}}) \circ i_{\mathbf{C}} \circ \alpha_{\mathbf{E}}(x_{\mathbf{E}})$$




(a) (b)

(c)

Fig. 5: Admissible transformations

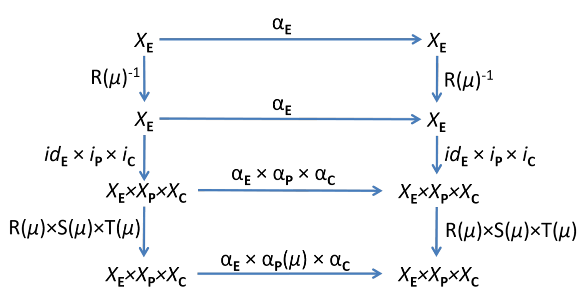



Fig. 6: Perturbation model commutative diagram

We must make the final crucial assumption of Rich Parameter Perturbation:

For each fixed $x_{\mathbf{E}}$, as $\mu_{\mathbf{E}}$ varies through $M_{\mathbf{E}}$ and $\mu_{\mathbf{P}}$ varies through $M_{\mathbf{P}}, R(\mu_{\mathbf{E}})(x_{\mathbf{E}})$ varies through $X_{\mathbf{E}}$ and $S(\mu_{\mathbf{P}}) \circ i_{\mathbf{P}}(x_{\mathbf{E}})$ varies through $X_{\mathbf{P}}$. It follows that

$$\alpha_{\mathbf{C}}[R(\mu_{\mathbf{E}})(x_{\mathbf{E}}), S(\mu_{\mathbf{P}}) \circ i_{\mathbf{P}}(x_{\mathbf{E}}), T(\mu_{\mathbf{C}}) \circ i_{\mathbf{C}}(x_{\mathbf{E}})]$$

depends only on $T(\mu_{\mathbf{C}}) \circ i_{\mathbf{C}}(x_{\mathbf{E}})$. In other words, for each fixed parameter value $\mu$, the system $\mathbf{S}$ has feedback structure on the attractor $\widetilde{X}_{\mathbf{E}}(\mu)$, namely for every $\mu$ the controller $\mathbf{C}$ is autonomous when regulation is perfect. As before, we deduce that $\mathbf{C}$ contains an internal model of $\mathbf{E}$, establishing Assertion 2 above.

So what is *the shape* of Mark Twain's river (which wanders about under perturbation)? Note that in Fig. 5, $\alpha_{\mathbf{E}} \circ R(\mu) = R(\mu) \circ \alpha_{\mathbf{E}}$, where $R(\mu)$ is in fact an automorphism. Thus a "small" perturbation $\mu$ merely shifts the current state of $\mathbf{E}$ to one on a "neighboring" trajectory of the same dynamics $(X_{\mathbf{E}}, \alpha_{\mathbf{E}})$. For *the shape* take any "nominal" trajectory you like!

## V. Conclusions

The scheme above may provide a basis for versions of the IMP in a variety of more structured technical settings, for example bang-bang or sliding mode. Not to mention refinements topological, metric, differentiable ..., where (essentially) the same commutative diagrams ought to work.

As stated here, the IMP crudely represents only a primitive "intelligence"; issues of adaptation, learning, computing power, and "real" problem-solving intelligence are open for investigation.

## References


[1] K. Craik, *The Nature of Explanation*. Cambridge U. P., Cambridge UK, 1943.

[2] M. Twain, *Life on the Mississippi*. Osgood, Boston, 1883.

[3] T. Hobbes, *Leviathan*. London, 1651.

[4] O. J. M. Smith, *Feedback Control Theory*. McGraw-Hill, 1958.

[5] W. M. Wonham, *Linear Multivariable Control: A Geometric Approach*. Third ed., Springer, 1985.

[6] ——, "Towards an abstract internal model principle," *IEEE Transactions on Systems, Man, and Cybernetics, 6(11)*, pp. 735–740, 1976.

[7] W. M. Wonham and K. Cai, *Supervisory Control of Discrete-Event Systems*, Monograph Series Communications and Control Engineering, Springer, 2018, in press.