Chapter 6

Early Abnormality Detection

Early detection means to observe an abnormality without knowing what the problem is going to be. It is the hardly noticeable vibration in a humming engine that you ignore until the meters are in the red or the engine fails. Having owned an older car, you know adjust yourself to it’s peculiarities, as you develop a gut-feeling for the combinations of vibrations, analog meters behaving binary and blinking lights that can be ignored ... and which not. It is not that different for operators of plants; there are only more vibrations, meters and blinking lights. There is no obvious theoretical knowledge from physics or logic supporting distinctions between the “acceptable” and “abnormal” behavior. Yet it seems to make sense to react to gut feelings to decide when the car should be taken to the garage. The health of complex distributed systems should not be trusted to gut feelings! The question arises how we can model for early detection? A model of the system and abnormalities from inside, the blue-print, is insufficient, too complicated, or not available to distinguish the acceptable from the abnormal. Looking from-the-outside-in we consider the behavior of the system as a whole, including it’s acceptable variations and abnormalities. The profound abnormalities are best detected from structure in parameters ... without confining a model to dimensions dictated by the system architecture, such as the “car engine”.

We reconsider the drivers to design for redundancy, and discuss model redundancy in relation to an inadequate but reigning modeling paradigm, reductionism. In section 6.2 we argue for redundancy inside the system model. Earliness is a key driver to enable fault prevention. We explain in section 6.3, that a single model is over-constrained if it is to detect profound change early as well as to reliably predict the severity of abnormality. In section 6.4 we address what we need to detect when abnormality is not a priori modeled, and why this requires monolithic modeling. In section 6.5 the requirements are considered once more from the perspective of system complexity, model redundancy and complexity and modeling accuracy, and we arrive at a surprising and challenging modeling requirement for early detection.

6.1 Motivation and preliminaries

6.1.1 A view on systems and abnormalities

System design versus system behavior

The design of a system starts with the concept of the desired function. An exact description of that function can serve as a system specification. Designing optimal system control, based on the system blueprint and expected deviations to compute control actions, is a deterministic procedure. The correct behavior of controllers is only guaranteed within a limited part of the state space. The specification of the controller is a control model that is based both on the system and on it’s environment. Exact modeling is valuable for controller design. However, even with
“optimal” control the manifest behavior is different from the behavior generated by an ideal model. Otherwise detection and accommodation are not required. Despite this difference, the desired function can be recognized in the behavior. But in practice the function of the system cannot be isolated from various unintended effects. In LADS, such intended functions result from a complex interaction of many processes, all introducing some uncertainties as none are perfect. The ideal “process” performing the function is an artifact, and not an actual entity. Abnormalities are sometimes taken for deviations of the ideal process. However, the detection is based on measurements of behavior which is in practice already not ideal to begin with.

**Definitions of types of behavior**

Behavior is what can be observed, and what is reflected in the observed data. In detection we consider the differences between what we expect from, desire of and observe in the behavior. Since these different perspectives on behavior are the key concepts in our discussion we will first clarify what we mean by them.

*Ideal behavior* results from a system and its environment in accordance to the logical or physical laws applied in the system design; the “idea” in “ideal” is key in this definition. We use the term ideal in the sense of archetypical, paradigmatic and conceptual. Hence in our interpretation, ideal means not necessarily perfect behavior, but rather behavior as far as can be explained and exploited from understanding. Ideal behavior is typically the behavior associated with a nominal model.

*Desired behavior* is optimal in terms of the desired function and qualities. Function and quality are conceptual a priori notions, whereas behavior is not. Functions or qualities may be isolated from behavior in theory but in reality they never exist in isolation. Stating that desired behavior in terms of a priori functions and qualities is fictitious may seem arbitrary, even obvious. But, consider that almost all the effort and energy spent on the design, control, monitoring and accommodation of complex systems is required for the pursuit of “optimal” behavior in a non-ideal reality.

*Actual behavior* is the behavior the system exhibits in reality. The measurement of actual behavior only provides an approximation of the dependencies generated by the system, since measurement has limitations (chapter 2). We will use the notion of actual behavior as the manifestation of behavior in the observed variables.

*Acceptable behavior* provides the desired function and qualities with just sufficient performance. Known deviations with unknown pathology are in practice often part of the acceptable behavior. Acceptable variations are deviations of the actual behavior from the ideal behavior and within limits of the desired behavior.

**Changes, abnormalities, severity and profoundness**

A disturbance is an error of the model in representing the system, which has intrinsic structure in it (subsection 2.3.6). However, it is not the disturbance we need to detect, but rather the abnormality that causes the disturbance, as the abnormality constitutes a change in the information source. The information source is the combination of the system and its environment, figure 2.7. Abnormalities are defined (section 4.1) as deviations in the manifest behavior of instances of the information source compared to the expected manifest behavior of that information source. We interpret this as differences causing the dependencies among the measured variables to change. An abnormality is not necessarily a fault or a failure. Changes as well as
faults are abnormalities; hence we need to distinguish between the cause of changes and the changes themselves. This we do through the concepts ‘severity’ and ‘profoundness’.

**Profoundness.** A profound change is a change in the actual system behavior relative to the ideal one, pertaining to the dependencies within the system. We discern three levels of profoundness. These levels correspond with the complexity of the system and abnormalities. Complexity is expressed by the d.o.f. required to model (figure 4.5). The profoundness of an abnormality increasing with each level:

1. superposed residuals;
2. state space aberrations;
3. change in the laws regulating the state transitions.

**Severity.** Severity is a measure for the degradation with respect to desired function and quality of a system. A severe disturbance structurally exceeds the boundaries of acceptable behavior.

**Complexity of system behavior and abnormalities**

In the design of complex systems, the divide-and-conquer strategy decrees the partitioning of the desired function into processes and the delegation of tasks. Each process performs a number of tasks to realize a specific function. However, the desired function is often more than the sum of all sub-functions, e.g. conditions of the system may play a role in the feasibility of the function. Moreover the cumulative behavior of all the processes and their interaction with the environment is more than the intended function. The divide-and-conquer strategy pursues a decomposition of functions up to a point where a single principle or technology exists to implement the function. In isolation processes that are understood can be monitored adequately for performing the right function. However, establishing the fact that all processes are performing their function adequately does not guarantee that the system is not headed toward failure. In particular, local control is often inclined to damp local variations in the performance, which obscures profound changes on the system level. This local damping effect introduces complex interactions between processes.

The complexity of processes is chosen such that they can be properly understood; this is the aim of the divide-and-conquer approach. Often an exact model of the local processes is possible in design, by following a strategy to simplifying and refining. The exact modeling of dynamic processes is made possible by simplifications, such as linearization, that are allowed as long as the equilibrium is sustained. An active controller is part of the design to take care of just that. A single optimal controller for the whole system, considering the whole state-space and the non-linear dependencies, is almost practically impossible. The state-space is partitioned and even within the parts the controller deals only with a limited fraction of the local state space. A distributed hierarchical control mechanism takes care of the translation of the global objectives into local set-points.

In reality behavior will manifest with a much higher complexity than can be expected from a composed nominal system model with it’s hierarchical control. This complexity results from deviations of the state space equilibrium. Outside the equilibrium the dependencies are not approximately linear, moreover inter-process dependencies can no longer be ignored. In system design, the independence of errors between parallel and sequential processing steps is an important condition. It allows for the separation of processing steps that is necessary to implement a complex distributed system.
Consider a system that is undergoing changes in its internal dependencies. It is likely that this results from aspects in the system that have neither been foreseen nor controlled. So new signal components and dependencies are introduced corresponding to modifications in dependencies beyond those already modeled in a control-oriented nominal process model. This implies that there is increased complexity. System changes that are not isolated in the local autonomous processes using the composed nominal system model, cause an unknown, possibly non-zero increase of the complexity of system behavior (and hence of the complexities of the disturbances).

Another complicating effect is the existence of instances. The desired function is similar for each instance, but there is much diversity in the behavior of the instances due to different operating modes and different operating conditions.

The challenge of detection when systems and abnormalities are intertwined
Abnormalities, that can lead, causally, to severe disturbances, are profound. Abnormalities that occur as a consequence of non-local interactions, are not isolated from the system but stay within the system. The system itself is not ideal, and the ideal process does not exist anymore when profound abnormalities occur. Profound abnormalities will manifest within the boundaries of acceptable behavior. The challenge is to detect these abnormalities, that are a priori unknown, knowing that systems and abnormalities are intertwined.

6.1.2 The problem of modeling limitations in detection
Dealing with complexity of systems in design
A single design or control model is usually too complex and thus not feasible for large systems. The prevalent design strategy utilizes a divide-and-conquer approach. A system is decomposed hierarchically into subsystems, subsystems into sub subsystems, etc. down to a level of detail where function, form and behavior coincide. This is the level of logical or physical components where the model follows directly from known logical or physical principles.

Locally autonomous distributed systems (LADS).
In the design of LADS the desired function is decomposed into sub-functions of lesser complexity, which are realized by subsystems. It requires quite an effort to regulate the interaction of all the components of the system. A global direct control over all components is often neither possible nor necessary, as - ideally - subsystems perform their function autonomously. Local control processes and the hierarchical distribution of set-points allows for this.

Consider the models used in the design and control for LADS. The overall model is a composition of submodels. Each model consists of some equations describing the desired or acceptable traversing through the systems state space. Differential equations and finite state machines are suitable paradigms to model these dynamics. Through these models the state changes can be related to changes in the input-output behavior and (partly) vice versa. Limitations are due to limited observability and controllability of the processes in the system. The composition of models describing the desired traversing is the nominal system model.

Detection problems for which conventional methods are adequate
Deviations that can be a priori specified accurately are best detected based on their model, using the methods described in chapter 4. In conventional approaches the disturbances are
described in relation to the desired dynamics as defined in the design and control model. Dis-
turbances can be detected adequately with conventional approaches, if

1. they can be described accurately and independently of the desired system behavior
2. the system behavior is explained by an invariant system model

The complexity of a system can be characterized in terms of the order or capability of the model required to describe the system's behavior (as measured in data). The capability of the model is formally expressed by the degrees of freedom in the model. In conventional model-based approaches, the system model and the abnormality model are a priori assumed -- or at least the model architecture is). The degrees of freedom for both system and abnormality model are conventionally thus fixed, or rather, constrained.

Severity and profoundness are related to the complexity of the abnormality. Since the abnormality and the system are intertwined, the vertical and horizontal axes in figure 4.5 actually coincide in the case of profound non-local abnormalities. Consider the diagram in figure 4.5. The upper-left quadrant (finite d.o.f.) excludes the unknown abnormalities and unknown system models. For these types of system changes and abnormalities the methods described in chapter 4 are adequate for LADS as long as the nominal system model is valid. Even if a re-estimation of the parameters of the nominal system model is required for fitting data of a disturbed system, conventional detection methods are adequate for LADS as long as abnormalities and system are finite and not intertwined.

**The system models from the design table are inadequate**

A change in the system dynamics can be detected very well if the dynamics are adequately modeled, such as in robust control and adaptive filtering (chapter 4). These approaches, however, depend on a good coverage of the possible dynamics through the state-space. In other words, for conventional approaches to be adequate the system model needs to be well parameterized and very complete (=powerful). This composed nominal system model, which has a limited set of parameterized differential/difference equations, is valid only near the equilibria. Abnormalities may cause disturbances revealing dependencies that are not explained by the nominal model. The problem for detection of abnormalities in LADS from measurements becomes clear: it is due to the partitioning of the state space and truncation of the state space equations. But how else can one handle the complexity?

**Blind identification and projection methods are too inaccurate**

In cases where abnormalities coincide with invalidation of the nominal system model, the only remedy provided by conventional methods is blind detection. Blind detection ignores the reference model of the systems behavior completely, i.e. detection thresholds are defined on model-free projections of the measurements resulting in a very coarse separation into “signal” and “null” space. The alternative novelty detection methods from the domain of computational intelligence are so-called resource allocation networks (RAN). RANs don’t have a priori bounds on the degrees of freedom. These blind detection strategies abandon the concept of relating behavior to internal states or internal state transitions. As a result of the absence of a parameterized model of the dynamics, the acceptable variations appear as a high level of “noise” on top of the (obscure) desired system behavior, and subsequently prevent choosing a tight threshold for sensitive detection of potentially harmful abnormalities.
The challenge to overcome the complexity explosion is a call for a new view

The gap between blind detection methods and model-driven approaches is too big. Detection based on exact system models fails because of an explosion in the complexity. Blind detection methods fail because they are too inaccurate to provide sensitivity. A new approach for detection is needed, but this requires a new perspective, different from conventional FDI and novelty detection methods. In our view we should assume:

- **Non-composability.** The behavior of LADS cannot be fully explained by composing a system model from models of locally autonomous processes. Hierarchical set-point control derived from a presumed assumed exact model will cause systematic deviations (key remark 5.4). A model to observe disturbances should therefore not be restricted to the state-space of local processes. The laws governing the local processing are not sufficient to explain global behavior.

- **Non-superpositional disturbances.** The disturbances and faults, occurring in the system behavior, are often inherent system changes. This implies that the nominal system model can be invalid. Then the faults and disturbances cannot be seen modeled as signals on top of the core system behavior as modeled by a nominal system model.

- **Unexpected.** The faults, disturbances and their causality are not known a priori.

**A parameter-based approach should be pursued**

We will pursue a parameter-based detection from a model. One argument is that blind projection methods ignore existing, but possibly unknown, invariant coherence, which can only be reflected in a model. Consequently they do not allow for tight decision boundaries, and cannot separate acceptable variations from profound and severe changes. A model that has the capability to describe the common underlying structure has more distinctive power to discern such variations than general statistical measures. Moreover a model is a specific projection of data that can be designed for certain objectives, whereas generic projection methods are not customizable. Finally the different levels of profoundness discussed above can only be distinguished using a parameterized dynamic model of the system.

### 6.1.3 Causes and consequences of bias

There are several contributors to residual structure that can be mistaken for change in the information source. Structure in the residual and in parameters can be a consequence of bias. Since it is the objective of detection to distinguish desired behavior from abnormalities we will have to pay attention to the matter of bias. Here we start with a brief introduction to the causes of bias and their manifestations in the residual and parameters of a model.

**Causes of bias**

The causes of bias arise from modeling limitations from 2.4.6. They are:

1. The bias-variance issue comes into play with the sampling of data. The information source is considered to be stochastic in nature, while the model parameters are estimated from limited data (subsection 2.4.2);
2. The choice of model architecture and the associated parameterization;
3. The configuration and choice of learning process.
There is the statistical claim that over-parameterized models suffer from memorization: they are more biased and fail to generalize. Structure and variations that result from the model rather than from the process (bias-variance but even worse than that memorization: patterns that are not generally present) are captured in the model. The structure due to modeling artifacts, such as bias due to memorization and finite quantization levels as far as uniform distributed data is concerned, will be reduced by averaging over larger amounts of data, either using more data for a single model or averaging the estimate of several models trained on fractions of the entire data set, i.e. boosting and bootstrapping. The average model will be “whiter” i.e. less biased.

**Bias appearing as structure in parameter adaptations**

There are two artifacts in the structure of parameter adaptations: 1) the bias of the model which will cause a non-zero average drift; 2) the parameter dependencies due to either the architecture of the model or the interaction between model components for a particular choice of parameters (we call this self-structure of a model). Note that such a drift on the part of the domain of the modeled function does not mean that the model is not among the “best” models minimizing the mean square error for the whole domain. The self-structure emerges as a result of estimating or learning a model from data, but does not correspond to structure in the data.

We can summarize the harmful consequences of bias by:

- A lack of parameters associated with dependencies underlying the observed behavior;
- Model artifacts, i.e. structure in residuals not caused by abnormalities;
- The estimation being ill-posedness preventing a sufficient detection from parameters.
6.1.4 Purpose and organization of this chapter

There are some design principles and assumptions in conventional detection approaches that cause limitations in the detection of profound abnormalities: (a) the idea of composability of a system model; (b) the superposition assumption; and (c) the assumption of a priori knowledge of abnormalities. The challenging category of abnormalities consists of changes in the information source that are in fact interactions that cannot be explained from conventional compositional models, as discussed in the chapter 5. The purpose of this chapter is to explore an alternative view on abnormalities and to analyze the consequences of this view for the requirements we impose on modeling. In detection the perception of what to detect is directly related to the modeling approach (how to model). Hence we focus our discussion on modeling for detection.

This chapter is organized around four topics. Each topic addresses the issue of what to model and how to model from a different perspective. Here we motivate the choice of these topics.

The first topic is redundancy. Redundancy is a key aspect in modeling detection, since it is required to observe behavior that we did not expect nor designed intentionally. We reconsider the drivers to design for redundancy, and discuss the redundancy of a model in relation to a seeming conflicting yet reigning modeling paradigm: reductionism.

The second topic is the conflict between earliness and accuracy. Earliness is a key driver to enable fault prevention. When pursuing early detection of a priori unknown, yet possibly profound, changes. Reliability calls for an estimate of the severity. We consider the requirements and possible trade-offs for modeling to see whether earliness and severity estimation can be conducted using a single model.

The third topic is the object of detection and the type of model required to detect. Profoundness is related to the amount of structure changed. We arrive at the thesis that a monolithic model is required to overcome the limitations of conventional composite models: an unconventional thesis for which we provide the arguments.

The fourth topic addresses the conflict between redundancy and accuracy. This conflict becomes relevant once we have arrived at the conclusion that a model for detection needs to be both monolithic as well as redundant. Optimization of a detection model to separate acceptable from abnormal behavior requires a trade-off between redundancy and accuracy. We discuss the limitations of exact and risk-optimized models to allow for such a trade-off.

We conclude this chapter with the requirements on model design to provide for detection of profound abnormalities that are beyond the capabilities of conventional methods.
6.2 Why redundancy inside the model?

6.2.1 The driver of observability

What is observability?

In a system theoretical sense, observability is the possibility to derive the internal state of an information source from its manifest behavior (see section 2.4.4). State space observers [Olsder, 1994] and Kalman filters (section 4.3.4) are specific methods to find states from measurements. We can see that the mapping of measurements to states is actually a form of estimation. Hence unobservability can be understood from the limitations of estimation as discussed in section 2.4.6. There are two main causes (section 2.4.6) of unobservability:

1. Ill-posedness of the estimation problem is a result of the model architecture or a result of the system architecture and the measurement set-up;
2. Under- and over-determination. Both are an unsuitable ratio of availability or number of measurements and the degrees of freedom in the model.

In practice we are always dealing with a finite number of measurements. There is a difference between fundamental limitations due to ill-posedness and limited quality of the approximation due to the availability of data. Observability limitations appear as a result of model architecture and of the estimation procedure, hence these limitations do not relate to availability of data.

Why observability should be a key driver?

How do we judge the quality of a model for its intended purpose (detection)? The model must allow for a comparison between actual and desired behavior. The model quality is verifiable in cases where the ideal and the abnormal behavior are known. This verification is not possible when we assume that the system and abnormality are intertwined and that the abnormalities are not a priori known. We only know from our analysis on the causes and consequences of bias, that composed and idealized models cause limitations, specifically limitations in the observability of non-local changes in the information source. It is through a falsification, through analysis of the architecture or empirically, that we can determine observability limitations. On this basis we can reject certain modeling approaches for abnormality detection. Observability is a necessary alternative for conventional design criteria since those cannot be applied due to a lack of a priori knowledge.

6.2.2 Channel analogy

An analogy between abnormality detection and communication is not far-fetched, considering that much of the detection theory has been developed in signal detection and communication. The analogy is based on a similarity in objectives: detecting and isolating a “structure” from measurements. The analogy serves to understand the requirements on the modeling for parameter-based detection, pursuing observability.

The Channel

In communication there are several steps between the posting and receiving of a message. We consider all these steps to be a part of the channel. We are specifically interested in the effects of the steps on the quality of the communication: the preservation of information. The channel has the following steps: Coding/Mixing, Transmitter, Communication Medium, Receiver/filter, Decoding.
The analogy is as follows: the effect of an a system change (the message) on the system is mixing/coding, while the system is the transmitter. The system, the environment and the sensors form the medium, the receiver/filter is the nominal model and the decoding correspond to the signature computation. A key concept in communication is anti-symmetry: all the transformations of the message in the coding, transmission and communication are reversed in the receiver and decoder. There is an implicit mutual knowledge enabling this anti-symmetry.

Table 6.1: A comparison between communication (outer cells) and detection (inner italic cells)

<table>
<thead>
<tr>
<th>Coding Mixing</th>
<th>Effect on system</th>
<th>Signatures</th>
<th>Decoding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transmitting</td>
<td>System</td>
<td>Data</td>
<td>Model</td>
</tr>
<tr>
<td>Physical Emitter</td>
<td>Interaction</td>
<td>Physical Medium Environment</td>
<td>Sensing</td>
</tr>
</tbody>
</table>

The message, noise and interference.

In communication the message coding is known by the receiver. This is not the case in abnormality detection, since abnormality (the “message”) is not created (“sent”) on purpose. The message itself, in case of a profound abnormality, is a change in the channel (the system and/or its environment). Analogous to the message, we can consider the acceptable variations as interference and the incidental randomness as noise.

Receivers are divided into: a) noise suppression filters; b) interference cancellation; c) signal specific filtering/projection. The analogy can be refined for the receiver. The noise suppression filters are the common pre-processing steps in detection modeling. The interference cancellation corresponds to the residual fault-sensitive filters. The filtering is the parameter-based detection, i.e. the mapping to physical properties.

Parameters, design criteria and mechanisms

The key design parameter in communication is bandwidth; it is the primary design parameter to influence the time it takes to communicate a message. Bandwidth has to be increased by each step of the coding and transmission to preserve information. Similarly, in the receiver the bandwidth is reduced as the signal is cleaned of noise and interference, and the common coding and transmission keys are used to filter and demodulate. The key driver in this sequence is the preservation of information. Knowing that noise and interference exist, the message has to be communicated in a redundant way to allow for a perfect reconstruction. In our context, that relates to the preservation of information and perfect reconstruction of the transmitted signals:

- Spacing: In communication the coding and transmission is chosen to ease separation of the different symbols being communicated. The simplest method is to distribute the symbols equally according to their Hamming distance (the number of bits that differ from symbol to symbol). There are elaborate methods of combining coding and modulation using phase, frequency and amplitude, yet all pursue orthogonality between the signals that correspond to the symbols.

- Parity and redundancy: Knowing the distribution of bit errors that occur in the signals, extra information can be added to the message prior to coding to ease the detection and possible correction of the communication errors. Examples are parity bits, checksums.
Conventional modeling for detection heavily exploits the similarity between the system and its ideal model, i.e. the symmetry between sender and receiver in the communication analogy. The orthogonality property, which is based on the assumption of independence between system and abnormalities, is likewise exploited. Considering the profound abnormalities, which are changes in the system itself, we can see that the symmetry is no longer valid. It is easy to understand that a false presumption of the shared knowledge such as on coding keys or modulation schemes introduces a bias. Abnormalities that do not invalidate the nominal model can be filtered and cleaned up neatly into orthogonal signals, whereas the more profound abnormalities are deformed by a false and biased model of the system.

Bias of the model reduces the observability; consequently less bias means better observability. Conventional parameter-based detection models aim to reduce the “bandwidth”. A compact representation of abnormality is pursued by estimating parameters from measurements. Ideally the conventional detection framework captures the abnormality in some parameters, with a physical or logical meaning; then boundaries can be derived algebraically. Since we assume that systems and abnormalities are intertwined, we have to reconsider how and if a bandwidth reduction in the detection model (analogous to the receiver/decoder) can be realized in the estimation of parameters from measurements without unacceptable loss of information.

### 6.2.3 Observability versus reductionism

**Reductionism pursuing exactness and statistical optimality**

In chapter 2 we start out with two modeling paradigms: the deterministic and the stochastic belief. The deterministic belief is expressed strongly in exact modeling and the theory of physics and logic; we perceive it as a modeling “from within”. The stochastic belief is expressed strongly in statistics and in probability theory centered on observation, we perceive it as a modeling “outside in”. Reductionism is the reigning paradigm underlying both beliefs.
An exact model of a system is a model that describes the laws that govern the behavior of the system. Moreover a model is exact only if it is derived from first principles, i.e. the laws of physics and logic that are commonly believed to be true, and only if it is expressed in unambiguous mathematics. Exact models describe the governing laws, and not necessarily the behavior itself. Occam’s Razor (Numquam ponenda est pluritas sine necessitate) is the central paradigm in the sense that the simplest model is considered the best model: the model that expresses no more than is strictly necessary. Exact modeling is important in design, where a useful application of the laws of nature is pursued. It is common sense to limit the expression of the desired function and quality in the simplest possible way in a design model to avoid misinterpretation.

In statistical and probabilistic modeling, which is applied especially in signal detection (chapter 4), there is also a clear pursuit of the simplest models. Statistical simplicity entails a reduction to a few stochastic variables with a known distribution - typically a normal distribution - and a minimal use of degrees of freedom (see subsection 2.4.5). There is an obvious pursuit of reduction to the simplest set of variables (simple distributions) and to the simplest set of relations between those variables (preferable linear relationships, and mutually independent variables). If we consider the modeling approaches in chapter 4, we conclude that conventional approaches depend on the assumptions that the variables are mutually independent and that dynamic and spatial relationships are approximately linear.

**Reductionism and LADS**

We have illustrated the reductionism in the modeling of LADS in chapter 5. The key issue in modeling for LADS is that the design complexity necessitates a divide-and-conquer approach. The divide-and-conquer approach induces a composite model, in which local models can be exact or statistically optimized for the processes in the system as long as control provides the conditions to allow for linear approximation and independence of the distributed processes.

**Conventional detection approaches are based on nominal models**

Isermann [Isermann, 1984] states the following on the role of process models in detection: “Process models should express as closely as possible the physical laws which govern the process behavior. Therefore... requires theoretical modeling”. The model from the design table, the so-called blue print, is the design target and expresses the desired system behavior. In conventional detection blue print is the reference model to explain the behavior, and abnormalities are considered a surplus “on top of” and not “part of” the nominal model. This conventional use of modeling for detection was discussed in section 5.1. The key issue is that the models of disturbances (whether residual or parameter-based) are separated from the nominal model.

Why do the conventional detection approaches stick to a reductionistic approach? The first reason is the strong belief that exact and statistically optimal models provide accurate models of the systems behavior. The second reason is that such a reductionistic model is readily available from the drawing board, and it reflects the principles, which are assumed to correspond with the desired function and quality. Abnormalities or disturbances are considered deviations from the model expressing the desired behavior. The third reason is the belief that an exact model of the system facilitates sufficient interpretation and understanding of the behavior. A fourth reason is that an exact method reduces the number of parameters, and also degrees of freedom, which in turn reduces the amount of data required to fit a model, which improves promptness.
Reductionistic modeling does not provide the best observability

So reductionistic modeling is the paradigm in conventional detection for providing accuracy. Is it right to pursue accuracy of a model, and is it right to equate exactness with accuracy? We have to return to the purpose of modeling for detection, to evaluate the effect of reductionism on the models quality for detection. The ideal detection provides a response if and only if (<= ) there are abnormalities. In this respect there are two requirements on modeling:

1. Presence of abnormalities implies deviation from the model (=>)
2. Deviations from the model imply abnormalities. (<=)

Can we deduce the existence of abnormalities from the residual of exact models? We can but only “if and only if”:

1. the exact models describe the desired behavior;
2. the “normal” behavior is the desired behavior;
3. the abnormal behavior is not intersecting the normal behavior in the residual.

We know these constraints are not met in practice. First, the “normal” behavior of a system in it’s environment includes acceptable variations. The abnormality has to be separated from these acceptable variations. Second, the exact models describe an ideal law underlying the behavior under specific conditions, which may fail in case of abnormalities. Third, modeling artifacts arise as a consequence of bias, as discussed in section 6.1.3, so deviations from the exact model do not always imply presence of abnormalities.

Since residuals are not a sufficient measure of abnormality, the objective of modeling for detection is not to provide an exact and accurate representation of the desired behavior. A model for detection is required to allow for a comparison between desired or accepted behavior on the one hand, and measured behavior on the other. The model is also required to observe abnormalities through parameter comparison. Is a parameterization, such as the one resulting from reductionistic modeling for LADS, providing the best observability?

When we assume that systems and abnormalities are intertwined, the model will under normal circumstances not have a sufficient capacity to describe the more complex behavior in case of abnormalities. Moreover the abnormalities are not a priori known. Nor are the acceptable variations, since the acceptable variations are, by definition, not explained by the nominal model. Consequently an exact model based on a priori design or a reductionistic statistical model based on normal behavior cannot be optimal for observing the differences between desired behavior with acceptable variations and abnormalities.

Sensitivity to the unknown is required. We quote [Venkatasubramanian, 2003], as discussed in section 4.1.4: “novelty detection demands sensitivity to the unknown, the novel, the malfunction. One has access to a good dynamic model but it is possible that much of the abnormal operations region may not have been modeled adequately”. In other words, a model that is exact and accurate for normal behavior is not necessarily the right model for abnormal behavior. Exact composite models, as well as a statistical reductionistic models, lack parameters associated with global dynamics, which become relevant in case of abnormalities. In view of the reductions required to compose an exact or statistically optimal model for LADS, we conclude that an exact (nominal) model of the normal behavior is not optimal for fitting the behavior of changing systems.
### 6.2.4 Reasons to avoid assumptions on system and abnormalities

#### Typical assumptions on systems and abnormalities
The assumptions on systems and abnormalities for conventional detection approaches can now be summarized:

1. The belief that the laws governing the behavior of a system can be known and understood a priori, yielding a nominal model in the form of mathematical equations which explain or generate the behavior (the white-box paradigm).
2. The belief that a combination of proper design and hierarchical control, where an optimal control sustains an equilibrium in local processes, allows for a linear approximation of the local process behavior independent of the global dynamics.
3. The belief that a superposition of abnormalities leaves the nominal model invariant, as expressed by the independence of nominal models from the residual or parameter-based fault models and fault filters.

These assumptions are reflected in the diagram in figure 4.5. The a priori knowledge of systems and abnormalities is the basis of model-based detection with finite models: the models of systems and abnormalities are independent and are either fixed or have finite parameters.

#### Drawbacks of assumptions on systems and abnormalities
The arguments against the assumptions have already been provided. We summarize the result of our analysis so far.

1. Global disturbances occur in LADS that cannot be explained from a composed reductionistic model of the system. We therefore have to assume system and abnormalities are intertwined and they are not independent (6.1.1);
2. Bias results from false assumptions and causes modeling artifacts (6.1.5);
3. Reductionistic models, arising from assumptions on first principles or pursuing statistical optimally for the “normal behavior”, do not provide the optimal observability. The interaction between abnormalities and the system results in more complex dependencies than present under normal behavior. A model, based on assumptions only valid under normal conditions, does not offer a suitable parameterization to fit abnormal behavior. It’s architecture obstructs the fitting;
4. Not all abnormalities can be assumed to be known a priori;
5. Acceptable behavior is more complex than the ideal or desired behavior.

Consequently we arrive at the proposition of universal modeling:

#### Proposition 6.1:
To detect abnormalities beyond the capacity of conventional detection, one should not make assumptions on the system, the abnormalities or their interaction, when designing a model for detection.

### 6.2.5 Arguments for redundant modeling

#### What is redundancy?

Does a car need four wheels to take you from A to B? No, it does not! In this sense it is redundant. But there are other aspects to transportation, such as comfort and safety, which are easier to implement with four than with two or three wheels. The laws of economy in the design of systems (parsimony) and comprehensibility in modeling pursue a minimal resource usage and
Why redundancy inside the model?

Chapter 6

EARLY ABNORMALITY DETECTION

Redundancy inside the model?

a simplicity respectively. Redundancy in this perspective can be interpreted as obsolete or unnecessary. Is 80 percent of our brain cells redundant in this sense? Maybe we could do without them, but clearly most of us would be reluctant to part from 80% of the grey matter. In this thesis redundancy does not mean obsolete, rather we define it as non-parsimonious, and it is only meaningful in relation to an objective, i.e. for going from A to B at least two of the four wheels are redundant.

Redundancy in conventional detection

When designing for quality and for FDIA, redundancy relative to the reductionistic system model is necessary to describe the functionality of the system. Some forms of redundancy in the context of conventional detection (see chapter 4) are:

- Hardware redundancy: extra resources for monitoring;
- Analytical or algebraic redundancies: parity relations describing the desired dependencies in various ways, so they become more observable. They are called algebraic because they can be deducted from the equations of the a priori first-principles model;
- Functional redundancies: additional models next to the nominal model to describe specific abnormalities, which are known a priori.

In conventional detection the “redundant” models typically come in two trivial forms:

1. As specific fault models and matched filters for specific abnormalities if the abnormalities are assumed a priori;
2. As projection methods (signature computation) which pursues a maximal separation (ideally orthogonal) between the desired and abnormal behavior.

Typical examples are null spaces and parity spaces. Note that such an orthogonalization, being an algebraic manipulation, requires an exact model.

Redundancy in the context of detection means to have more degrees of freedom than necessary to express the desired behavior. Depending on the approach the desired behavior can be the function and quality of the ideal nominal model.

Why is a non-trivial redundancy required?

Any detection approach pursues the redundancy for optimal sensitivity to abnormalities. If the objective is to identify any deviation from the ideal, then redundancy boils down to no more than a trivial boundary test on the residual. The objective, however, is not to detect deviations from the ideal or desired behavior. Unavoidable noise and other unstructured deviations as well as acceptable variations in the behavior are simply not explained by the exact or statistically optimal model. Therefore the abnormal behavior needs to be separated from the acceptable behavior. Both the deterministic (fault models, matched filters) as well as blind detection depends on additional tests to compare structure in the behavior. This is why non-trivial redundancy is required with respect to the reductionistic model.

Redundancy “inside” the model is required

What does it mean to have redundancy inside the model? To understand this, consider the conventional FDI as in figure 5.1. and conclude that the redundancy is outside the nominal system model, i.e. the redundancy constitutes either a fault model or a fault filter for specific faults or it is a projection. Either way the redundancy is in the procedure of signature computation. The
key here is the use of a nominal system model or, more generally speaking, a reductionistic model of the normal behavior. A model has redundancy “inside”, if the system model, or rather the model architecture, is also capable of describing the behavior when a changing or changed system brings more complex dependencies in the data, without the need to add new adaptive parameters.

One example is the blind identification discussed in 4.3.5, where the model architecture is capable of capturing various types of behavior as long as it fits in the linear system's model architecture parameterized by the A, B and C matrices. The key feature, however, is that the parameters (A, B, C) in the example represent the behavior and are estimated rather than chosen a priori, and hence are redundant. The essential difference between reductionistic modeling and redundant modeling is the use of a static nominal system model versus an adaptive behavioral system model. To detect such abnormalities from the parameters of the model, the model needs to be capable of describing the normal as well as the abnormal behavior, and therefore has to be redundant with respect to the reductionistic model of normal behavior.

We have two main arguments to propose redundancy “inside” the model for the purpose described above:

- Consider the limitations occurring as a result of reductionism. There are dependencies manifesting as disturbances but they are not explained from the first-principles model, as we have concluded in chapter 5. The conventional detection approach with a reductionistic model causes gaps in the modeled dynamics, which are required to be described the acceptable variations as well as the abnormalities. Global disturbances that result if these acceptable and abnormal deviations exist, as illustrated in chapter 5. Consequently the control-oriented and reductionistic composite system model is not providing a suitable parameterization. The gaps in the global dynamics appear since the redundancy is outside the model, and because it is based on the assumption that the structure associated with normal behavior remains invariant, even when abnormalities occur. We, however, have assumed the system and the abnormality are intertwined, and abnormality are not known a priori.

- Recall the channel analogy. The system model should be considered part of the channel because we require parameter-based detection. The model should therefore not limit the observability by assuming a structure that introduces bias. A sufficient redundancy can prevent unobservability.

### 6.3 Separate long term analysis from early detection

#### 6.3.1 Earliness

**What is earliness?**

Earliness is the ability to detect abnormalities from errors before possible underlying changes surface as severe and sustained performance degradation. Early detection aims at a merely sufficient detector response in case of abnormalities using a fixed, but limited or a minimal amount of measurements, rather than being driven by a fixed confidence bound.

**Why earliness?**
As soon as an abnormality surfaces as a severe disturbance it is already (by definition) beyond acceptable boundaries. After detection it takes time to isolate, and diagnose the abnormality and to determine the proper action to prevent escalation. Therefore abnormalities that are potentially severe should be detected early. The most unexpected abnormalities are often the most disruptive to the system operation, so these need to be tackled prior to having a impact.

Isolation and diagnosis should not be applied before an abnormality is detected. This is first of all because the isolation and diagnosis procedures will often raise false alarms and false identifications due to ad-hoc variations in the operation of the system. Secondly, early detection can coarsely indicate the start of an abnormality, allowing for a search on a much smaller number of measurements.

Early detection starts from the absence of a priori knowledge of abnormality and it’s occurrence in time. Diagnosis takes possible symptoms and causes as a perspective, and relies on clearly surfacing symptoms. Furthermore, diagnosis is also incomplete, since it can only cover known relationships between symptoms and their causes. Hence early detection is a required complementary guard.

No confidence bounds

Early detection is not driven by a fixed confidence bound. Constraining a detector design driven with fixed confidence bounds implies that the types of manifestations in behavior are well categorized and have known a priori probabilities, such as the receiver operating characteristics in communication. In the practice of LADS there are many different but individually rare events. The system behavior, an interaction between system and environment, can evolve in many different unforeseeable ways. The number of possible types of behavior is unbounded. Hence a sensitive detection is not possible through a classification with fixed confidence bound on generalized parametric data models.

6.3.2 Array processing inspiration

First part of the analogy

This concept illustrates the key objective of detection: isolating an object from noise and interferences. A key measure of the quality of detection is the Signal-to-Noise ratio. In interference mitigation the quality of the detection and mitigation procedure is expressed as the ratio of the Interference-to-Noise Ratio (INR) prior to and after cleaning. This is typically for radio astronomy, because as the interferer is the “unwanted” signal and the “noise” is the desired signal. In early detection there is no reference signal of the source to be detected, as the source of disturbances is not assumed to be known and is expected to hide below the noise.

How to reduce noise and interferers?

Noise is reduced by averaging over measurements. Random noise reduces as it is not a coherent signal while the structure in the signal is amplified with each additional input signal or measurement. Common averaging takes place either in time, frequency or space, or all three domains [Boonstra, 2005].

Sensitivity of an array

Design parameters to increase sensitivity of array processing systems are: effective reception area \((A_{\text{eff}})\); system noise temperature \((T_{\text{sys}})\); the integrated bandwidth \((B)\); and, the integration
time \((T)\), i.e. the number of observations. The sensitivity of an array is given in equation 6.1. The key idea is that incidental structure or unstructured noise will disappear when averaging, while coherent structure remains.

\[
S = \frac{A_{\text{eff}}}{T_{\text{sys}}} \sqrt{B \cdot \bar{T}} 
\]  

(6.1)

**How this inspires design for early detection**

The strategy for improvement is to boost earliness and reduce bias using spatial rather than temporal redundancy. Our driver is to minimize the number of samples \((T)\) for early detection. This can be done in various ways:

- \(A_{\text{eff}}\): increase number of baselines/(diversity in orientations), increase the d.o.f. in i/o’s.
- \(T_{\text{sys}}\): Use a model to separate random noise from structure
- \(B\): increase range of variables (diversity in input); also, use on different time-scales

Earliness is improved by spatially redundant representation (multi-view, extra d.o.f.). The required time for averaging can be shortened by averaging over the distinct observations of the structure of possible sources. The number of observations for the detection of system changes decreases (under invariant confidence) with the system behavior being modeled (under invariant risk) with increasingly more redundancy, provided the redundant degrees of freedom remain relevant to the representation of acceptable and desired behavior.

**6.3.3 Blind identification versus earliness**

**Earliness and analysis of severity**

Detection serves to guard the functions and qualities of the system. It is applied to find severe potential degradation of those qualities in the long term. Early detection thus requires modeling to provide:

1. Earliness: the ability to detect the presence of an abnormality as soon as possible;
2. The ability to estimate the propensity of a system to evolve towards severe faults.
The detection of abnormalities has to be blind because assumptions on the system or on abnormalities should not be made. Hence the latter requirements calls for blind identification. This is in conflict with earliness: solving both by a single model will result in over-constraining the modeling approach, as we shall explain hereafter.

**Optimal number of samples is different for earliness and severity.**

The drivers for modeling are different. Earliness requires that the presence of an abnormality has to be estimated within a given allowable number of samples, while the confidence of it’s identification is not important.

The dynamics of the abnormality have to be identified and extrapolated to meet requirement 2, given a fixed confidence level while independent of the number of required samples. In addition the model has to be good enough to make extrapolations, so good it can accurately identify the abnormality. To identify the abnormality the number of samples has to be larger than the number of samples needed to detect the presence of an abnormality.

**Optimal accuracy of the model is different for earliness and severity**

Earliness requires observation of abnormalities as differences between “normal” behavior and behavior observed in new measurements. Earliness does not require that the model’s accuracy is optimized by relating the measurements to qualities and functions of the system, as long as differences in behavior become observable with a limited number of measurements.

The relation between measurements and system qualities is of key importance to determine the severity. An accurate model with respect to the qualities of the system is more important than a model describing the behavior. In the absence of abnormalities the accurate model with respect to the qualities of the system may not be that different from an exact model of the behavior, but the gap will be larger when a system behaves different from the idealized first-principles model. The model for an early detection of profoundness of abnormalities is parameterized differently from a model for accurate estimation of the severity of abnormalities.

**The ideal use of d.o.f. in a model is different for earliness and severity**

We have argued for redundancy inside the model for detection. Given the two requirements of earliness and severity analysis, what is the optimal redundancy to be used? If, in the design of a model, we had a fixed d.o.f, how would the two requirements utilize them? First note that the analysis of severity is only meaningful after the presence of an abnormality is determined. In order to detect early we use the d.o.f. to look in all directions with a limited accuracy for each direction. In case we already known that an abnormality is present we would use the d.o.f. to accurately capture it’s direction and extrapolate along that direction to assess it’s impact. Two short analogies will clarify this difference.

Consider Heisenberg’s uncertainty principle: \( \Delta x \Delta p \geq \frac{h}{2} \). The accuracy of localization times the accuracy of impulse \( (p=m\cdot v) \) is constant. This means that in a test we have to trade-off the accuracy with which position is determined for the accuracy of speed at which a known particle is traveling. Either the 'where' is clear and the 'what' (impulse) is less clear or vice versa.

Another analogy is that of the Fourier theory. A fundamental limitation of the Discrete Fourier Transform is \( \Delta \omega \Delta \tau \geq 2\pi \). The accuracy (or resolution/stepsize) in normalized frequencies times the time-resolution is fixed by the unit-circle. When taking N samples, the N samples can be divided up in several ways to make a spectrogram (matrix with frequencies against time). In
one extreme we have one frequency (\(f=0\), equals the mean), and \(N\) small time steps, in the other extreme we would have \(N\) frequencies but no time-information (an \(N\) point DFT). Early detection requires many inspections in time with a coarse granularity in the frequency domain, whereas an accurate analysis of the severity requires an average over time to get a clearer estimate of the abnormality and use only a few time-steps to get a coarse idea of the trend.

### 6.3.4 Separate long term analysis from early detection

The requirements on a model for long-term analysis are different from the requirements on the model to enable early detection. A reliable detector responds if and only if there are profound abnormalities that result in severe disturbances. Accurate predictions and simulation are necessary to conclude that an abnormality can evolve towards a severe disturbance, fault or failure. In particular, it will be crucial to have an accurate approximation of the change. This requires a longer series of measurements than for early detection.

Where, as inspired by the array-processing analogy, the presence of an abnormality can be concluded from changes in spatial dependencies of a dynamic model, the dynamics only have to cover a time-window of the desired behavior, which is often much smaller than the number of measurements over time to make accurate predictions. However, without early detection, we can forget about long-term analysis, diagnosis and prevention. So, we conclude that modeling for early detection must be resolved without over-constraining the model with a requirement of confidence on the actual severity of the impact of an abnormality.

### 6.4 What to detect, and why monolithic modeling?

#### 6.4.1 Focus on amount of structure in drift

**The profoundness of abnormalities is not the amplitude of disturbances.**

As soon as an abnormality surfaces as a severe disturbance it is already (by definition) beyond acceptable boundaries. How do systems respond to profound changes prior to surfacing? Well, the local controllers are designed to mitigate the effect of these changes on their local performance. If the abnormalities are strictly local, they are also effectively mitigated either by local control or local FDIA. If the abnormalities are not local but a part of a global system change, local control will not prevent the evolution of the change but attempt to mitigate it’s symptoms locally. The consequence is that the net effect of the profound changes is initially reduced until local control can no longer suppress a local disturbance resulting from a global interaction. Consequently, although an abnormality may be profound and potentially severe, this is not necessary reflected in the amplitude of corresponding disturbance, as it first emerges on the outside.

**What can we detect?**

The amplitude of the disturbances is not a measure for the profoundness and potential severity of the abnormality. Moreover assumptions on the system, abnormalities and their mutual relationship should be prevented. So, what then can we detect?

**What is information?**

The analogy of detection design with a communication channel has yielded the understanding that we need to preserve information along the path from the changes in the system to the sig-
natures reflecting the changes. Apparent information is the object to be detected. Hence the question arises: how can we preserve information?

Information is structure. Since we only have measurements, the structure constitutes dependencies between measured variables in time. Preservation of information does not mean preservation of the exact “message” in an absolute sense, since we do not pursue identification but merely detection. However, early detection does require preservation of the amount of structure.

**Structure compared to what?**

Abnormalities, that manifest themselves as new structure in the data, can be detected by blind methods (chapter 4). They can be revealed as well by some data analysis measures (chapter 2). However, blind detection ignores the possibility to use of a model of acceptable or desired behavior. The usage of a model and detection from parameters is advantageous as we have argued in section 6.1. A model by itself describes structure as it relates variables to one another; this structure is the reference for comparison.

So, can we compare measurements to a model of the ideal behavior? No, we cannot, because the desired ideal is a specification of a function and it’s qualities; we observe through measurements only the behavior of a real system in a real environment, which is an attempt to implement the desired function. We can describe the behavior through a model of observed variables. In fact, when acceptable variations relative to the ideal behavior occur, the “ideal” behavior can only be guessed from many instances that are all acceptable variations. Ideal behavior is a conceptual notion that is obscurely present somewhere within the boundaries of the observed acceptable behavior. The challenge is to separate abnormalities from acceptable behavior, not to isolate abnormalities from ideal behavior.

What is an appropriate model of acceptable behavior in the case of LADS? To improve on blind detection directly from data, we depend on modeling of acceptable or desired behavior, and parameter-based detection. A model of the system behavior is required, because a change in a system is measured as the drift of it’s model, which is equivalent to a change in the structure, i.e. the modeled dependencies. The amount of structure within the model drift is a measure for profoundness of change in the system.

**Separate acceptable variations from abnormalities**

Acceptable variations from a virtual and obscured ideal behavior are likely to have structure rather than to be random variations (white noise). This structure must be observable if we require abnormalities to be observable. Otherwise, we have to know a priori that the abnormalities compared to the ideal behavior are different from the acceptable variations compared to the ideal-behavior, and that in turn implies that we can isolate and describe the ideal behavior from measurements. In practice, we cannot make these distinctions between abnormality, ideal system behavior and acceptable variations a priori.

The earlier discussion on the conflict between reductionism and observability also indicates that a separation of ideal behavior (from an presumed exact a priori model) is harmful to observability; hence such ideal behavior must not be a priori isolated from acceptable variations. Similar to our statement that abnormalities and the system are a priori inseparable, we have argued that the acceptable variations are a priori inseparable from abnormalities. What are the implications of this seemingly trivial statement? It means that the difference between
acceptable variations (structure) and abnormalities (new structure) must be measurable from a model of the systems behavior (structure). Hence, by the design of the reference model, the acceptable variations must be measurable to separate them from abnormalities.

### 6.4.2 Monolithic modeling

#### What is monolithic modeling?

If we consider the modeling of a system behavior, we mean with the term monolithic that a single, non-composite architecture for all temporal and spatial structure. In other words, in the model architecture we do not distinguish a priori particular specialized functions. Having a single model for temporal structure implies an invariant single architecture across multiple instances. Therefore the monolithic model must generalize to identify common structure from multiple instances. The lack of bias to structure implies that there is no modularity and all structure depends on the choice of parameters. A module is an a priori (hard-coded) specialization or a distinct function; hence it refers to structure in the architecture rather than a structure in input-output behavior that is parameter determined.

**Conventional detection models for LADS are not monolithic**

A straightforward application of the detection modeling of chapter 4 to the LADS of chapter 5 is not possible due to an explosion in complexity. The modularity appears naturally in the modeling of LADS. First of all, the starting point of a detection model is the nominal process model expressing the logical or physical idealized dependencies underlying the system behavior. There is a fundamental belief, expressed in particularly by Isermann (chapter 4), that a sensitive detection should come from an exact model based on physical principles. The abnormalities and disturbances are modeled on-top-of the nominal model (figure 4.1), which implicitly means that the underlying idealized dependencies are to be assumed invariant.

Chapter 4 provides many fault-specific detection models and filters. These are common approaches applied largely in complex systems, and a clear illustration of modularity in the conventional detection model. The nominal model for LADS itself is a composed model, assuming that the processes are independent while all local processes are controlled hierarchically to sustain locally optimal equilibria to allow for linearization.

Finally, the reality of modeling for LADS often yields a specific model per instance, which varies from the one extreme, in which a model is completely dedicated to an operational mode (conditions/configuration), to the other extreme, in which the solution is based on a core model patched with model extensions specific to operational modes.

**Assumptions used to reduce complexity do not apply to detection modeling**

The desired function and quality of the system are in fact not equal to the sum of function and quality of the subsystems, even though the design-and-control philosophy relies on the compositionality assumption to adopt a divide-and-conquer approach, which in turn yields a certain partitioning of the system. The need for detection itself shows that a model based on this partitioning of the system is inadequate to explain the behavior of the system within it’s environment (chapter 5).

Hence we propose (proposition 1) not to assume a nominal system model, neither to assume a priori the types of abnormalities or their interaction with the system. Clearly, a priori partitioning requires a priori knowledge. Yet, especially this kind of knowledge can become invalid in
case of abnormalities. The modularity in the system is allowed under the assumptions of (1) ideal control; (2) decoupling of the function/qualities of the autonomous processes; and (3) assumption that local optimality implies global optimality.

A detection model cannot make these assumptions simply because it has to be susceptible to dependencies across processes, that can cause global disturbances. Detection modeling cannot assume a hierarchical structure of the control nor can it assume the modularity of the system to simplify the complexity of the modeling.

**A priori specialization prevents the preservation of structure in abnormalities**

Avoiding assumptions on the internal structure means any underlying structure must be identified from the observed behavior. This means that there can be no a priori specialization in the model inspired by the design of the system or from known physical principles. The structure of the direction of the accommodation necessary to describe the change in behavior is unknown, but it is the subject of detection as we argued in the previous section. Moreover the redundancy must be inside the model.

Consequently it is necessary that the accommodations of the model to abnormalities are not independent of the acceptable behavior model. All the structure in the data that is interfering with the structure as captured in the model of the acceptable behavior should have an impact on the model parameters. Because of the necessary observability a detection model must fit both 1) acceptable behavior; and 2) emerging abnormal behavior. Typically, to complement the local FDIA, the dependencies across distributed processes, which are ignored by the design and control model, become relevant.

- **Lack of parameters.** In case the detection model is a modular model, some changes in dependencies are not possible a priori, because it will lack independent parameters.

- **Controllability of parameters.** A mathematical analogy will help to understand this. Consider the whole space of potential behavior spanned by a basis. Each module, in the model, can be thought of as a vector and the parameters are scalars representing the “presence” of this vector in the behavior. If a model is truly modular, the vectors form independent subspaces. The possible mixtures of modules determines all possible behavior. However the parameters of the modules and the parameters to combine the modules cannot be determined independently from measurements. The scalars to determine the presence of the basis vectors cannot be chosen independently. The structure in measurements will hence only be partially observable in the adaptation of the parameter.

When the disturbance (structure) caused by an abnormality is projected onto the parameter space, the amount of structure must be preserved. The projection of measurements to parameter space, which is a combined effect of model and estimation procedure, should not be a priori limited. A specialization of the model towards any instance must be reflected in the parameters. Hence in the model architecture there should not be a priori specialized components or components biased towards certain specialization. Hence it must be a monolithic model.

**Specialization per instance prevents preservation of structure in variations**

Detection demands a robust model, i.e. the detection of abnormalities must be insensitive to acceptable variations. This insensitivity cannot be achieved by designing the model such that acceptable variations are in the null-space, because these cannot be a priori separable from the ideal behavior. Consequently robustness implies with respect to acceptable variations the need
for their observability. Therefore the modeling needs to have the capacity to describe and distinguish these variations, not from the ideal but at least from the abnormalities that may occur. This implies three requirements on the detection model:

1. The model should potentially describe the “common” structure across different instances, i.e. the model should be good on average;
2. The model should be able to accommodate the variations and all acceptable variations should be modeled sufficiently accurate; an equal amount of structure in variations implies an equal level of adaptation in the model;
3. A comparison should be possible between the different effects that the variations have on the model.

**Remark 1.** This means that no specific models per instance can be allowed, since the instance-specific structure will not be comparable to an average over all instances. It also means it is necessary to capture common underlying behavior of components that are not known a priori.

**Remark 2.** The degrees of freedom in the model must be finite, i.e. a dependency between variables in the model is affected when this dependency in system changes. Hence the model is redundant only in the sense that the temporal or spatial complexity of the relationships can vary; and memorization of individual instances is not allowed.

### 6.5 Redundancy, complexity and risk

#### 6.5.1 Redundancy versus minimal-risk

**Risk or Loss**

Recall the bias-variance problem discussed in subsection 2.4.2: the actual error of a model for a whole data space cannot be known but only approximated, as there is only a limited set of samples. The expected quadratic loss is given in equation 6.2.

\[
E_{y \in D, \hat{y} \in M}[(y - \hat{y})^2] = E_{\hat{y} \in M}[(\hat{y} - E_{\hat{y} \in M}[\hat{y}])^2] + (E_{y \in D}[y] - E_{\hat{y} \in M}[\hat{y}])^2 + E_{y \in D}[(E_{y \in D}[y] - y)^2] 
\]

In the actual risk or loss or mean square prediction, there are three components:

- The variance \( V \) over the different possible models: \( V = E_{\hat{y} \in M}[\hat{y} - E_{\hat{y} \in M}[\hat{y}]] \);
- The squared bias: \( B^2 = (E_{y \in D}[y] - E_{\hat{y} \in M}[\hat{y}])^2 \);
- The noise or variance in the actual behavior \( \sigma^2_y = E_{y \in D}[(E_{y \in D}[y] - y)^2] \).

Averaging over different models is required for \( \hat{y} \in M \). The different models result from splitting the data differently, i.e. different model estimates arise from minimizing \( V + B^2 \), using different samples from the database, and also the different possible optima give a redundant model, i.e. there may not be a single unique optimum \( w* \) for \( M(w*,x) \). We have discussed various ways to approximate the actual quality of a model in section 2.4.3, such as bootstrapping and cross-validation. The risk is minimized by \( \hat{y} = E_D[y] \).
Complexity mismatch and risk

As an example consider polynomials to describe functional dependencies, or similarly Fourier components. Each component describes a unique part of the functional behavior that is independent from the rest of the components. The components of a model, called basis vectors, kernels, or wavelets, are conventionally chosen to be independent, and even orthogonal such as with Fourier components, polynomial factors, or wavelets.

The number of independent components determines the complexity of the model. Truly redundant components have a zero-value parameter on average and will introduce an unnecessary bias when estimated from a finite amount of data. Moreover we run into estimation problems in case of linear models for risk-invariant redundant modeling: the complexity of the model must match the complexity of the system generating the data; otherwise the system is over-determined or under-determined and a solution will not be found.

In such models the parameters have a unique optimum with respect to the given data. The risk of the estimator (the model) depends on the selection of the proper dimensions. There is only one optimal model complexity, and the complexity mismatch determines the risk. The complexity of conventional models is fixed, i.e. the polynomial order, order of the dynamics for all (state) variables and their interaction is assumed to be known prior to parameter optimization. A model of independent or orthogonal components or linear mappings cannot be redundant. Recall the memorization-generalization issue (chapter 2): statistical redundancy in a model implies under-determination. It is claimed this will inflict harmful over-fitting of the data, resulting in bias. Such a model should fail to generalize.

| model: $\hat{y} = b_0 + b_1 x + b_2 x^2 + \ldots + b_m x^m$ |
| system: $y = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n$ |

b is fitted from data.

$m=O(\text{model}); n=O(\text{system})$

If $m \neq n$ we have a problem

Figure 6.4 : Example of complexity mismatch for polynomial regression
The complexity mismatch is related to this last issue. Roughly speaking the capacity or complexity of the model must be equal to the complexity of the system that is being modeled. Expressions and estimation of spatial and temporal complexity have been discussed in 2.4.5. In theory the risk grows with the complexity mismatch. There are different theories of how risk and complexity mismatch are related. Hence we write

\[ \text{Risk} - N = V + B^2 \sim |O(\text{model}) - O(\text{system})| = \text{complexity mismatch}. \quad (6.3) \]

The problem of redundancy versus risk (R-R) is that a model with more degrees of freedom than necessary has a greater loss or statistical risk. Hence when the model \( M \) has a d.o.f. \( O(M) \) suitable to describe the complexity \( O(S') \) of system \( S' \), and even though the complexity \( O(S) \) of the ideal system \( S \), requires less d.o.f: \( O(S) < O(S') \), then the risk \( L(M) \) of the model \( M \) is greater than \( L(M^*) \) of the optimal model \( M^* = E[M|S] \) for normal behavior \( L(M) > L(M^*) \). \( L \sim |O(M) - O(S)| \).

### 6.5.2 Risk-invariant redundancy

There is a conflict between redundancy and risk. We have argued that redundancy is required, even redundancy “inside” a monolithic model, but is it necessary to minimize the risk? We have discussed the occurrence of model artifacts. One of the causes of model artifacts is overparameterization, e.g. the incorrect estimation of an \( a_n x^n \) component. These artifacts can also be the consequence of an unstable or a non-converged learning process (subsection 6.1.4). If we do not optimize for risk (subsection 6.1.5):

- Structure in the residual will result from the model artifacts, which can easily be mistaken for variations or even abnormalities in the system.
- A bias in the model may cause structural deviations in the projection of errors to parameter adaptations \( \Delta w \); consequently actual disturbances caused by abnormalities are obscured.
Consider the subtle difference between model artifacts and bias on one hand, and the different possible optima given the acceptable variations in the system behavior on the other. In section 6.4 we argued that these different acceptable variations must be observable. Hence a customization of a model M* optimized for all variations leads to an improvement for the specific instance. We conclude that the model risk has to be optimized, so it should not affect the risk when redundancy is introduced in the model. The reason is not the desired accuracy of the model for acceptable behavior, but the minimization of model artifacts and bias that reduce the observability in the projection of errors to parameters. This is consistent with the earlier discussion on observability and the channel analogy.

![Figure 6.6: Complexity related to model quality: classical models (left), and desired model (right)](image)

Ideal risk-invariant redundancy means is that the complexity of the model can be chosen independent of the risk of the estimator. The ideal (figure 6.6 on the right) is not achieved from the illustrated complexity mismatch issue (figure 6.6 on the left). The redundant d.o.f. in M is $O(M) - O(M^*)$, and risk-invariance means $L(M) = L(M^*)$. $M^*$ being the minimal model for $S$. The desirable relation would be:

$$O(M^*) - O(S) \sim L(M^*) = L(M) - O(S)$$  \hspace{1cm} (6.4)

The complexity mismatch $O(M) - O(S)$ can be written $O(M) - O(M^*)$ since $O(M^*) \sim O(S)$, if $M^*$ is the optimal model for $S$. This is convenient as $O(S)$ is not known, but $O(M^*)$ can be estimated from the behavior (section 2.4.5). Hence we can restate the risk-redundancy conflict, and see that the desirable relation is not evident:

$$L(M) - L(M^*) \sim [O(M) - O(S)] - [O(M^*) - O(S)]$$  \hspace{1cm} (6.5)

$$\Rightarrow L(M) - L(M^*) \sim O(M) - O(M^*) > 0$$  \hspace{1cm} (6.6)

The redundant d.o.f. must show a potential capability of the model that is not used for optimizing the normal behavior. The redundancy provides degrees of freedom that increase the observability and that are “reserved” for the increment in the complexity of the system behavior in case of abnormalities. An obvious conclusion is that an adaptive model is required for detection, but there is more to it, as we shall discuss hereafter.

### 6.5.3 A soft-scaling complexity

If the redundancy provides a “potential” capacity to describe the behavior of a system with abnormalities, then we require a type of model that has an unused d.o.f. in case of $S$ that will be used in case of $S'$ (the abnormal system), or $O(M|S) \sim O(M^*)$ while $O(M|S') \sim O(S') > O(S) \Rightarrow$
O(M|S') > O(M|S). Let us consider the linear models with independent components (recall the examples of the Taylor series, Fourier series and wavelets). Is it possible to achieve \( O(M|S') > O(M|S) \), when \( M=P(\theta) \) is a linear model (with \( O(M)=\text{rank}(\theta) \) the order/complexity and \( \theta \) the parameters)? If we set \( n>O(M^*) \), then a presence of components in \( M|S' \) which are absent in \( M|S \) corresponds to a zero to non-zero transition in a few of the parameters \( \theta_i \). The question is whether inspecting the \( \theta_i \)'s after re-estimation can provide an estimate of profoundness.

**Why it isn't sufficient to monitor coefficients of a linear model**

The answer is yes; it is necessary to include these types of inspections. But no, it is not sufficient! First, it is not the most important thing to watch for the amplitude, it is the amount of structure that counts, because it is a fair estimate for profoundness. Secondly, look at the observation “acceptable variations are always present” and the assumption “the abnormalities are not superposed on but intertwined within the system”. A model optimized for the acceptable behavior will not have components to capture abnormalities (the coefficient of these components will be zero). When the abnormalities arise, previously unused components come into play independently of used components. Consequently nothing can be concluded on the profoundness of the abnormalities. Thirdly, the model of the complex system behavior is rarely a simple exact linear model, or even a combination of several simple linear models. The behavioral model is rarely exact. The abnormality does not surface strongly in one a priori known dimension, because the dimensions where it can manifest cannot be designed a priori. The dimensions of the behavioral model’s parameter space will not align with dimensions along which abnormalities emerge. Consequently abnormalities manifest across many parameters with weak amplitudes in most dimensions. The amount of structure in an abnormality is not observed in the individual parameter changes of the model. A strategy that merely inspects the coefficients of a linear model in isolation, will not provide the maximal sensitivity.

**Parameter-based abnormality detection requires a soft-scaling complexity**

Recall the requirement of a parameter-based detection (section 6.1), an even stronger requirement is derived (section 6.4): early detection depends on inspection of the parameters of a *monolithic model*. The independent components in a linear model for \( S' \) will not be necessary to describe the behavior of \( S \). However we need to detect abnormalities and separate them from acceptable variations by inspecting the parameters \( W \) of model architecture \( M \). First of all this comparison is impossible if \( M \) has the capacity to describe \( S' \), i.e. \( O(M)<O(S') \). Secondly, if the model \( M \) is capable of describing \( S' \), but the components of \( M \) are not really used in \( M|S \), then the measure of how far systems and abnormalities are intertwined is not reflected in the model. The abnormalities are then modeled independent of the acceptable system behavior, and the system behavior is indirectly considered invariant.

Consider that the dimensions of the model can be chosen such that the parameters allow for independence between abnormal behavior and acceptable or ideal behavior. This is not possible in case abnormalities are not known a priori. The model ought not to be based on an assumption of independence between systems and abnormalities, and the parameterization must maximize observability; hence the parameterization must be redundant (i.e. have potential d.o.f.). Moreover, the model should have a parameterization that is suitable for both acceptable as well as abnormal behavior. Hence the abnormalities will be scattered across the parameters of the model.
This analysis implies is that the increase in complexity should not be in the increased use of independent model components, if this constitutes a parameter transition from zero to non-zero. The complexity of the model is not determined by the number of parameters used, \text{rank}(W), but rather by the parameters themselves. The abnormalities in the system must cause a gradual transition in the constellation of parameter values. The complexity of the model is consequently a soft-scaling property rather than a hard integer property scaling with \text{rank}(W).

**Two-tier models cannot provide soft-scaling complexity**

Two-tier arrangements, like physical-principle models and parametric statistical regression, are model types in which there is a direct relation between the type of component in the model and the type of component in the system or signal. In physical-principle and white-box models, this correspondence couples model parameters to actual physical properties, i.e. there is a physical interpretation and a separation based on knowledge of the system. Similarly the parametric statistical regression presumes certain distributions of variables and independence between variables, which are optimally chosen or designed for a dataset. Both methods are reductionistic and pursue a decoupling between the parameters of the model, which corresponds with an independence of model components. Consequently the complexity (in the sense of dimensionality) is a function of the rank of the parameter space. The correspondence of model components with actual components or factors in the system results from the assumption of a countable and deterministic reality, it prevents a soft-scaling complexity in models.

**6.6 Conclusions**

Earliness in detection requires to target the profoundness rather than the severity of abnormalities. The latter is an after-the-fact-observation, while the former corresponds to inherently changing systems. Complexity is conventionally addressed with reductionism. Still, time-varying behavior cannot be exactly captured from physical principles in a modular model. Consequently abnormality can no longer be defined as orthogonal to the subspace covering the nominal model. In section 6.2 we argue that abnormalities are inside the system. Abnormalities and acceptable variations are not confined to a priori determined dimensions; therefore redundancy is essential inside a model to fit measurements of profoundly changing systems. In section 6.3 we argue that unfamiliarity with abnormalities requires to separate the detection from identification and impact analysis, as addressing these aspects with a single type of model over-constrains the design. In section 6.4 we argue that the unfamiliarity requires to detect from structure rather than by excess amount. Imposing a system architecture or modular construction to the model will confine parameter adaptations while abnormalities are not similarly confined. Hence we have proposed a monolithic model for early detection in section 6.4.

We reveal the implications of system and abnormality complexity on the requirements for model capability and accuracy. Having considered observability and earliness, we argue for redundancy inside a monolithic model, detecting profound change from dependencies in it's parameters. The capability of a model to reveal profound system change, beyond the complexity of the “normal” system, must not reduce this model's capability to achieve a minimal statistical risk in describing the normal system behavior. This calls for the *risk-invariant* redundancy (subsection 6.5.2). We state (subsection 6.5.3) that profound abnormality coincides with a changing system, implying that changes are not confined to the original system dimensions. Hence, if parameters have to reflect an abnormality, then soft-scaling complexity is required in
The model. This property is fundamentally not found in models that are linear compositions of linearized, possibly dynamic, kernels.

The key requirements in modeling are:

1. absence of a priori architecture through data-driven modeling;
2. a soft-scaling model complexity rather than a integer dimensionality;
3. risk-invariance, which can be taken as plasticity w.r.t. the required complexity

Early detection requires a model with the ability to fit and generalize from multiple instances while differences in instances are observable via the model parameters. Hence, a key requirement for a modeling approach is possibility of improving observability by increasing redundancy without increasing statistical risk.