Nonparametric–Validated Computer–Simulation Surrogates: A Pareto Formulation

by

Miltos E. Kambourides

B.S., Mathematics, Massachusetts Institute of Technology (1996)

Submitted to the Department of Mechanical Engineering
in partial fulfillment of the requirements for the degrees of

Master of Science in Mechanical Engineering

and

Bachelor of Science in Mechanical Engineering

at the

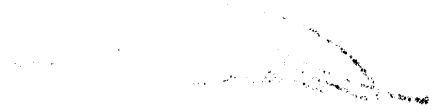
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 1997

Author .....
Department of Mechanical Engineering
May 9, 1997

Certified by .....
Anthony T. Patera
Professor of Mechanical Engineering
Thesis Supervisor

Accepted by .....
Ain A. Sonin
Chairman, Departmental Committee on Graduate Students



Nonparametric–Validated Computer–Simulation Surrogates: A Pareto Formulation

by

Miltos E. Kambourides

Submitted to the Department of Mechanical Engineering
on May 9, 1997, in partial fulfillment of the
requirements for the degrees of
Master of Science in Mechanical Engineering
and
Bachelor of Science in Mechanical Engineering

Abstract

In numerical optimization routines, computer–simulations are typically executed repeatedly in an *on–line* search to minimize some objective. Many computer–simulations, however, are too resource–intensive to be incorporated directly. Instead, they are evoked only to construct and validate *off–line* a simplified model that serves as a simulation *surrogate* in subsequent optimization studies. Surrogates are, by construction, computationally inexpensive and therefore create a highly interactive and flexible optimization tool, particularly powerful in multipoint design studies.

In this thesis we present "basic" and Pareto surrogate formulations for optimization, through an illustrative application from fluid dynamics. The critical ingredient of both formulations is a nonparametric statistical validation and error estimation procedure which, based on verifiable hypotheses, precisely quantifies the effect of surrogate–for–simulation substitution on system predictability, stability, and optimality.

The Pareto formulation extends the basic approach to multicriteria optimization problems in which we wish to minimize multiple objectives. Any solution for which *all* objectives cannot be further minimized is Pareto optimal. Geometrically, in the space defined by the outputs as coordinates, the Pareto optimal solution set lies on the boundary of the output achievable set which is closer to the origin; it is the image of a low–dimensional manifold in the design space. The surrogate Pareto formulation inexpensively identifies this manifold by appealing to the surrogate. As a result, in the presence of many inputs and few outputs, it considerably reduces the dimensionality of the optimization problem, and correspondingly improves the surrogate error estimates.

Thesis Supervisor: Anthony T. Patera
Title: Professor of Mechanical Engineering

Acknowledgments

First, and foremost, I would like to express my graditude to my advisor, Professor Anthony T. Patera, for his impeccable guidance, ample support, and tremendous encouragement he has provided me since my undergraduate years. Working with him has been an incomparable experience.

I am also grateful to Dr. Serhat Yeşilyurt for supplying the simulation models for this work. His continuous assistance and attention to all difficulties I was facing at the early times has been invaluable.

I thank Marius Paraschevoiu, Nikolas Hadjiconstantinou, John Otto, and Jeremy Teichman for their enlightening discussions, intelligent advice, and most important for their friendship. Special thanks to Marius for his patience to all my technical questions.

I would like to thank all the members of the Fluid Mechanics Laboratory for creating a very friendly and supportive atmosphere. I also extend my appreciation to all my friends in MIT who I shared plenty of exciting experiences with, in all my five years at this great school.

Lastly, but most emphatically, I thank my parents Georgia and Elias, and my sister Kallia, whose love and trust I never lacked.

This work was supported by DARPA Grant N00014-91-J-1889, AFOSR Grant F49620-94-1-0121, and NASA Grant NAG 1-1613. We thank Nektonics, Inc., for their assistance with the NEKTON code.

To my parents

Georgia and Elias

Στους γονεις μου

Γεωργεια και Ηλια

Contents

List of Figures

Chapter 1

Introduction

1.1 Motivation

For most engineering design and optimization problems, the objective is to improve the performance of a physical system. The performance is characterized by the system outputs which are a function of the design variables. Their relationship is not usually available in an explicit form and, in many cases, numerical optimization routines rely on a large–scale accurate numerical model of the physical system in the form of a computer simulation.

In most current simulation–based optimization studies, a mathematical programming procedure is called directly as many times as necessary for the solution of the optimization problem. This approach, known as "direct insertion," offers predictive accuracy and, for local searches, effective treatment of high–dimensional design spaces. However, many simulations remain too resource–intensive, despite the advent of fast and inexpensive parallel computers. Direct insertion may then be prohibitively costly, inefficient in multipoint or "evolving" design frameworks, and incapable for real–time interaction. It might even be terminated prematurely, since the number of simulation calls cannot be determined *a priori*; and even if a solution is reached, it might not be relevant for future studies with different design variables or objective

function.

An alternative, complementary approach to direct insertion is response surfaces [1], or "surrogates" [2, 3], in which the large–scale simulation is evoked only to construct and validate — off–line — a simplified input–output model; the model then serves as a simulation *surrogate* in subsequent engineering optimization studies. The major advantage of this approach is that, by construction, surrogates are computationally very inexpensive, and thus easily incorporated into optimization procedures. Additionally, their low cost creates a very flexible environment for global and multipoint design studies, with significant interactive and real–time potential. The number of appeals to the simulation is predetermined, assuring a complete solution with a fixed computational cost. Data from previous runs or other resources may also be readily expoited by the surrogate approach.

The primary drawback of surrogate–based optimization is the introduction of a new source of error. A surrogate validation strategy and associated error norms must be developed to quantify the discrepancy between the surrogate and the accurate simulation, and to estimate the effect on system predictability and optimality. The former provides a bound on the behavior of the actual objective function in the neighborhood of the surrogate–predicted minimizer, while the latter provides an ascertainment of the additional computations required to improve upon the surrogate result. Furthermore, the surrogate is a stationary substitute for a specific simulation; adaptive screening techniques (alteration of the design variables for instance) cannot be pursued rigorously.

In previous work [4, 5, 6, 7], a nonparametric statistical validation and error estimation procedure was introduced, which, based on verifiable hypotheses, precisely — albeit probabilistically — quantifies the effect of surrogate–for–simulation substitution on system predictability, stability, and optimality. The fundamental limitation of this *"basic" formulation* is the rapid decrease in design localization and predictability as the number of inputs, M, that is, the number of design variables, is increased —

the well–known curse of dimensionality.

1.2 Objective

One approach to reducing dimensionality is to more tightly couple the optimization procedure to surrogate construction and validation [8]. A second approach, the one we pursue in this thesis, is to identify a low–dimensional manifold in the design space $\Omega \subset I\!R^M$ (M design variables) as a region of "special interest." In particular, we develop a surrogate *Pareto formulation* in which we exploit the concept of Pareto optimality [9] to identify such a manifold; we then show that the minimizer of our objective function *must* lie on this manifold, and that the image of this manifold in the ouput space is the efficient frontier $\partial\mathcal{A}$ of the output achievable set $\mathcal{A}$. Due to the fact that any point not on $\partial\mathcal{A}$ is either not optimal or not achievable — which is the underlying property of the efficient frontier — the identified manifold is certainly the relevant restriction of Ω for a large number of engineering optimization problems. As a collateral benefit, we extend our earlier surrogate notions to problems with objective functions that are subject to constrains. This class of problems, treated in multiobjective or multicriteria optimization, is typically a much more realistic model for engineering design than the corresponding scalar problem, and much more cumbersome to solve without significantly restricting *a priori* the design space.

We shall see that there is a synergy between the surrogate and the Pareto notions. In particular, the dimension reduction of the latter greatly sharpens the error estimates which characterize the former, while the simplification associated with the former renders the min–max scalarization subproblems required by the latter [10] computationally tractable. Although the advantages of simplified models in Pareto analysis have certainly been recognized [11], the benefits of Pareto analysis in modeling efforts have not, to our knowledge, been previously recognized or exploited.

1.3 Outline

In this thesis we demonstrate both the basic and the Pareto formulation through an illustrative application that involves fluid flow in a channel interrupted by a regular periodic array of cylinders [12]. In Chapter 2, we define the optimization problem by describing the physical system, the simulation models, and the design objectives. In Chapter 3, we present the basic formulation as a foundation for subsequent development of the Pareto formulation, which is presented in Chapter 4. Each formulation comprises the same steps: surrogate construction; surrogate validation; surrogate–based design; and, finally, error analysis. Lastly, in Chapter 5, we summarize the main steps through an abstract formulation, and discuss outstanding problems and potential palliatives. All the figures are found at the end of each chapter.

Chapter 2

Problem Definition

To effectively demonstrate our surrogate procedure, we choose an application from computational fluid dynamics. The application is "real" in the sense that the associated simulations are sufficiently complex that surrogates are, in fact, required for the solution of the optimization problem.

2.1 Physical System

The system that serves as our illustrative example is the "eddy–promoter channel heat–exchanger" shown in Figure 2-1 (found at the end of the chapter). The system is described by laminar two–dimensional incompressible unsteady fluid flow and heat transfer in a channel interrupted by an infinite periodic array of cross–flow cylinders, the "eddy–promoters." For simplicity, we consider fully developed flow in a single periodic–cell of length $\widehat{L}$, where $\widehat{L}$ is the separation distance between successive cylinders; the flow is driven by an imposed pressure gradient, $-(\overline{d\widehat{p}/d\widehat{x}_1})$ (carats denote dimensional quantities). The two channel walls and the cylinder are maintained at uniform temperatures, $\widehat{T}_w$ and $\widehat{T}_c$, respectively. Finally, our system is controlled by a (non-feedback) oscillatory rotation of the eddy–promoter about the cylinder axis; more precisely, we assume that the tangential velocity of the eddy promoter is de-

scribed by $\widehat{A}\sin\widehat{\omega}\widehat{t}$, where $\widehat{t}$ is time.

We choose for our nondimensionalization length scale the channel half–width, $\widehat{h}$, for the velocity $\widehat{V} = (-\overline{d\widehat{p}/d\widehat{x}_1})\widehat{h}^2/2\widehat{\rho}\widehat{\nu}$, for time $\widehat{h}/\widehat{V}$, and for temperature $\widehat{\Delta T} = \widehat{T}_c - \widehat{T}_w$ (> 0, say), where $\widehat{\rho}$ and $\widehat{\nu}$ are the density and kinematic viscosity of the fluid, respectively. The governing equations are then the Navier–Stokes equations for the (nondimensional) velocity and pressure, $((u_1, u_2), p)$,

$$\frac{\partial u_i}{\partial t} + u_j\frac{\partial u_i}{\partial x_j} = -\frac{\partial p}{\partial x_i} + \frac{2}{Re}\delta_{i1} + \frac{1}{Re}\frac{\partial^2 u_i}{\partial x_j \partial x_j} \quad \text{in } \mathcal{D},\ i = 1,2\,, \tag{2.1}$$

$$\frac{\partial u_i}{\partial x_i} = 0 \quad \text{in } \mathcal{D}, \tag{2.2}$$

where $(x_1, x_2) \in \mathcal{D}$ is the single–cell periodic domain, $Re = \widehat{V}\widehat{h}/\widehat{\nu}$ is the Reynolds number, δ_{ij} is the Kronecker–delta symbol, and summation over repeated indices is assumed. The velocity and pressure are $L(= \widehat{L}/\widehat{h})$–periodic, and the velocity vanishes at all solid boundaries save the cylinder, where

$$u_i(x_1, x_2, t) = (A\sin\omega t)\widehat{\tau}_i. \tag{2.3}$$

Here $\omega = \widehat{\omega}\widehat{h}/\widehat{V}$, $A = \widehat{A}/\widehat{V}$, and $\widehat{\tau}_i$ is the unit tangent vector on the eddy promoter surface. The temperature $T = (\widehat{T}_c - \widehat{T}_w)/\widehat{\Delta T}$ satisfies a passive scalar equation

$$\frac{\partial T}{\partial t} + u_j\frac{\partial T}{\partial x_j} = \frac{1}{RePr}\frac{\partial^2 T}{\partial x_j \partial x_j} \quad \text{in } \mathcal{D}, \tag{2.4}$$

where $Pr = \widehat{\nu}/\widehat{\alpha}$ is the Prandtl number, and $\widehat{\alpha}$ is the thermal diffusivity of the fluid. The temperature vanishes on the channel walls and is unity on the cylinder. Note that we are interested only in the steady–periodic flows; it can be shown that the particular initial conditions chosen are not important except as regards computational efficiency.

Although not crucial to our discussion here, we briefly summarize the physics of the flow. For the fixed pressure–gradient Reynolds number studied here, $Re = 250$, the unperturbed ($A = 0$) flow is steady and effectively parallel. However, at a

slightly higher Reynolds number, Re_c, the flow undergoes a supercritical Hopf bifurcation caused by cylinder shear layer destabilization of the native channel Tollmien–Schlichting waves [12]; the shear layer dictates growth–rate, while the Tollmien–Schlichting waves govern frequency. For forcing frequencies ω near the natural frequencies of the unstable modes we expect a resonant response and hence a relatively well–mixed flow.

We consider $M = 2$ *inputs*, or design variables, with which we shall optimize the system: the amplitude, A, and frequency, ω, of the dimensionless tangential velocity of the eddy–promoter surface. All other quantities, including the channel geometry depicted in Figure 2-1, are held fixed for the purpose of our analysis. The Prandtl number is set equal to unity, and the Reynolds number is set to 250, which is below, but close to, Re_c of the primary supercritical Hopf bifurcation. In order to provide proportional resolution in the error estimation, the analysis will actually be carried out with respect to "log" inputs, $\mathbf{p} = (p_1, p_2) = (\ln A, \ln \omega)$.

The *input domain* Ω, the region in which the system inputs (p_1, p_2) must reside, is taken to be the box $\Omega \equiv (-2.30, 0) \times (-2.77, 0.63)$. This domain includes an extended range of amplitudes, $A_\ell = 0.1 \leq A \leq A_u = 1.0$, and a reasonably broad band of frequencies, $\omega_\ell = 0.063 \leq \omega \leq \omega_u = 1.88$; the latter is chosen to contain the frequencies of the least stable linear modes of the unperturbed flow in order to provoke subcritical resonance. Figure 2-2 shows the effect of the inputs, $p_1 = \ln A$ and $p_2 = \ln \omega$, on the flow isotherms.

We consider $K = 2$ system *outputs*. The first output, $\theta = \Theta^T(\mathbf{p})$, is the reduction in the eddy–promoter temperature difference relative to the uncontrolled case, $A = 0$, for a prescribed heat transfer rate. Here θ is the output and $\Theta^T(\mathbf{p})$ is the input–output function, $\Theta^T(\mathbf{p}) : \Omega \rightarrow I\!R$; the superscript T refers to "truth," as defined more precisely in Section 2.2. In order to calculate the temperature at fixed heat transfer rate from our fixed cylinder–temperature calculation, we need only evoke the linearity of the passive scalar equation (2.4). To wit, if we define the computed heat

transfer from the cylinder

$$b = \left\langle \int_{\partial \mathcal{D}_{\text{cyl}}} \nabla T \cdot \hat{n}\, d\zeta \right\rangle, \tag{2.5}$$

then

$$\Theta^T(\mathbf{p}) = \frac{b(A=0)}{b(\mathbf{p} = (\ln A, \ln \omega))}, \tag{2.6}$$

where $\partial \mathcal{D}_{\text{cyl}}$ is the cylinder surface, $d\zeta$ is a differential line element, $\hat{n}$ is the unit normal on $\partial \mathcal{D}_{\text{cyl}}$, and $< \cdot >$ refers to temporal average. Our second output, $\varpi = \Pi^T(\mathbf{p})$, is the nondimensional time–averaged power input per unit channel length required to rotate the eddy–promoter at the prescribed rate ω, given by

$$\Pi^T(\mathbf{p}) = c_0 \frac{1}{4} \frac{Re^2}{L} \left\langle A \sin \omega t \int_{\partial \mathcal{D}_{\text{cyl}}} \hat{\tau}_i \sigma_{ij} \hat{n}_j d\zeta \right\rangle, \tag{2.7}$$

where σ_{ij} is the viscous stress tensor and $c_0 = 1/1000$ is a convenient scaling parameter. (Note that $\Pi^T(\mathbf{p}) = \widehat{\Pi}^T(\mathbf{p})/(4\widehat{\rho}\widehat{\nu}^3/\widehat{h}^3)$, where $\widehat{\Pi}^T(\mathbf{p})$ is the dimensional power input.) We emphasize that no *regularity assumptions* are made on the input–output functions given the presence of subcritical resonance and highly nonlinear behavior.

2.2 Hierarchy of Models

We shall denote the model presented in (2.6) and (2.7) as the *truth model*, $\mathcal{M}^T$, from which we compute the truth input–output functions $\Theta^T(\mathbf{p}), \Pi^T(\mathbf{p})$. In fact, we cannot solve (2.6) and (2.7) analytically, so we appeal to a highly accurate 8^{th}–order spectral element simulation (described below) integrated for sufficiently long times such that all quantities of interest are effectively stationary. This truth model, deemed effectively exact, is *very expensive*, requiring approximately six workstation hours per input to compute the required outputs. Note that, in many cases, the truth model might be a *physical experiment*, perhaps noisy [13].

In what follows, we shall also need approximations to the truth model. At the next level in the hierarchy, we introduce the *high–fidelity* model, $\mathcal{M}$, from which

we compute the high–fidelity input–output functions $\Theta(\mathbf{p}), \Pi(\mathbf{p})$. The high–fidelity model is defined by a 4^{th}–order spectral element simulation integrated to a *specified*, reasonably long time, $t = t_f(\mathbf{p})$ ($t_f(\mathbf{p})$ ranges from 300 to 500 depending on the inputs). This model, which is considered accurate but not exact, is *expensive* but not *very expensive*, requiring roughly 40 workstation minutes per input to compute the required outputs.

Finally, at the lowest level in the hierarchy, we introduce the *optimization model*, $\widetilde{\mathcal{M}}$, from which we compute what we shall denote the surrogate input–output functions $\widetilde{\Theta}(\mathbf{p}), \widetilde{\Pi}(\mathbf{p})$. In contrast to the first two high–level models, the optimization model, at least in this thesis, is simply a response surface [1, 2]: the optimization model input–output functions, $\widetilde{\Theta}(\mathbf{p})$ and $\widetilde{\Pi}(\mathbf{p})$, are empirically constructed from corresponding high–fidelity–model data, with no independent underlying state–space representation. In general, the surrogate input–output functions will be *very inexpensive* to evaluate, requiring only a few workstation milliseconds per input; the fidelity of the surrogate outputs is the subject of our error analysis.

The truth (8^{th}–order, 11,000 spatial degrees–of–freedom) and high–fidelity (4^{th}–order, 3,300 spatial degrees–of–freedom) simulations are performed with the NEKTON code [14] on an HP735/9000 workstation. The NEKTON code is based on spectral element spatial discretizations [15], semi–implicit fractional time–stepping schemes [16, 17, 18], and preconditioned conjugate gradient iterative solution procedures [19]. For the high–fidelity model we take for initial condition the steady ($A = 0$) solution for $Re = 0$; for the truth model we take as initial condition the corresponding high–fidelity velocity and temperature fields at $t = t_f(\mathbf{p})$.

2.3 System Design

2.3.1 Optimization Problems

The optimization problems for our system can be generally stated as the minimization of an objective function

$$\Phi^T(\mathbf{p}) = \phi(\mathbf{p}, \Theta^T(\mathbf{p}), \Pi^T(\mathbf{p})), \tag{2.8}$$

subject to constraints

$$\Psi_i^T(\mathbf{p}) = \psi_i(\mathbf{p}, \Theta^T(\mathbf{p}), \Pi^T(\mathbf{p})) \leq 0, \quad i = 1, ..., n_c, \tag{2.9}$$

which may, and typically do, evolve during the design process. We impose the usual conditions on $\Phi^T(\mathbf{p})$ and the $\Psi_i^T(\mathbf{p})$ so that this problem is well–posed, such as lower semi-continuity of $\Phi^T(\mathbf{p})$. Note that we assume that the objective function and constraints depend only on the inputs and the *selected* outputs, not on the general system model $\mathcal{M}^T$.

One simple example of such an optimization problem is to search for the minimum eddy–promoter temperature

$$\theta_{\min}^T = \min_{\{\mathbf{p}\in\Omega\}} \Theta^T(\mathbf{p}), \tag{2.10}$$

and the corresponding minimizer, that is, the log–amplitude and log–frequency at which this minimum occurs,

$$\mathbf{p}^{*T} = \arg\min_{\{\mathbf{p}\in\Omega\}} \Theta^T(\mathbf{p}). \tag{2.11}$$

In general, $()_{\min}$ shall refer to the minimum, and $()^*$ to the minimizer, of any particular optimization problem.

A slightly more interesting example is a constrained problem in which we look for the minimum power subject to the temperature constraint $\Theta^T(\mathbf{p}) \leq \bar{\theta}$, where $\bar{\theta} \in I\!R_+$ is given. In this case we require

$$\varpi^T_{\min} = \min_{\{\mathbf{p}\in\Omega | \Theta^T(\mathbf{p}) \leq \bar{\theta}\}} \Pi^T(\mathbf{p}), \tag{2.12}$$

with corresponding minimizer

$$\mathbf{p}^{*T} = \arg \min_{\{\mathbf{p}\in\Omega | \Theta^T(\mathbf{p}) \leq \bar{\theta}\}} \Pi^T(\mathbf{p}). \tag{2.13}$$

Both these optimization problems reflect the general notion that better heat exchangers achieve lower temperatures at lower powers, a fact which will ultimately provide the basis for a more general multiobjective formulation of which (2.10)–(2.11) and (2.12)–(2.13) are special cases.

2.3.2 Optimization Frameworks

Most current simulation–based optimization frameworks rely on "direct insertion." In this approach the truth simulation appears as a function call to a mathematical programming procedure; typically, sensitivity derivatives [20] are also exploited to provide more rapid convergence to the minimizer. As mention in the Introduction, the advantages of this approach include predictive accuracy and, for local searches, effective treatment of high–dimensional design spaces. The major disadvantage is minimal control over simulation resources: a typical optimization procedure will require *many* evaluations of the objective function, $\Phi^T(\mathbf{p})$, each of which requires an appeal to $\mathcal{M}^T$ to compute (at least) $\Theta^T(\mathbf{p}), \Pi^T(\mathbf{p})$, each of which consumes *many* hours of computation. A solution to the simple optimization problems (2.10)–(2.11), which requires a global search over the input space Ω, is practically impossible with direct insertion (actually, it would take almost 7 years to find the minimum temperature of a 100×100 uniform input grid over the input space with the current

technology). Subsequently, this approach suffers from inefficiency in multipoint design studies, inadequate robustness, and little interactive or real–time capability.

In contrast, in the surrogate approach proposed here, the high–fidelity model $\mathcal{M}$ is evoked to construct the inexpensive surrogate input–output models $\tilde{\Theta}(\mathbf{p}), \tilde{\Pi}(\mathbf{p})$, which are then validated against the truth model, $\mathcal{M}^T$. The truth and high–fidelity models are then dismissed, and we consider a surrogate optimization problem in which $\Theta^T(\mathbf{p})$ and $\Pi^T(\mathbf{p})$ are replaced by $\tilde{\Theta}(\mathbf{p})$ and $\tilde{\Pi}(\mathbf{p})$ in the objective function (2.8) and constraints (2.9). For example, for (2.12), we look for the surrogate minimum (more precisely surrogate–proposed minimum),

$$\widetilde{\varpi}_{\min} = \min_{\{\mathbf{p}\in\Omega|\tilde{\Theta}(\mathbf{p})\leq\bar{\theta}\}} \tilde{\Pi}(\mathbf{p}), \tag{2.14}$$

and surrogate minimizer (more precisely surrogate–proposed minimizer)

$$\tilde{\mathbf{p}}^* = \arg \min_{\{\mathbf{p}\in\Omega|\tilde{\Theta}(\mathbf{p})\leq\bar{\theta}\}} \tilde{\Pi}(\mathbf{p}). \tag{2.15}$$

The advantages of this approach include direct control over expensive simulation resources, increased flexibility in global and multipoint design studies, and significant interactive and real–time potential. For example, different $\bar{\theta}$ in (2.14) may be investigated without re-appeal to the truth or high–fidelity simulations.

The primary disadvantage of the surrogate approach is the introduction of a new source of error, particularly problematic in high dimensional input spaces (large M). We can identify two types of errors [4]. The first, related to predictability, stability, and design localization, is concerned with the behavior of $\Theta^T(\mathbf{p})$ and $\Pi^T(\mathbf{p})$ for $\mathbf{p} = (\ln A, \ln \omega)$ in the vicinity of $\tilde{\mathbf{p}}^*$. The second, optimality, is related to the proximity of the surrogate minimizer $\tilde{\mathbf{p}}^*$ to the true minimizer $\mathbf{p}^{*T}$, and the proximity of the surrogate minimum, $\widetilde{\varpi}_{\min}$, or perhaps $\Pi^T(\tilde{\mathbf{p}}^*)$, to the true minimum $\varpi^T_{\min}$. These errors must be understood if the surrogate approach is to be useful. Unfortunately, classical approximation theory is not particularly appropriate since (i) the

expense of $\Theta^T(\mathbf{p})$ and $\Pi^T(\mathbf{p})$ permits only very few evaluations, and (ii) the complexity of $\Theta^T(\mathbf{p})$ and $\Pi^T(\mathbf{p})$ permits only minimal, typically non-quantitative, regularity assumptions. Furthermore, the unknown distribution of minimizers in multipoint design studies precludes standard average–case error analyses in design space, such as in Information–Based Complexity theory [21] or Bayesian loss–risk analysis [22, 23]. We thus pursue a nonparametric statistical approach in order to obtain, at a prescribed computational cost, *calculable* estimates based on *verifiable* hypotheses.

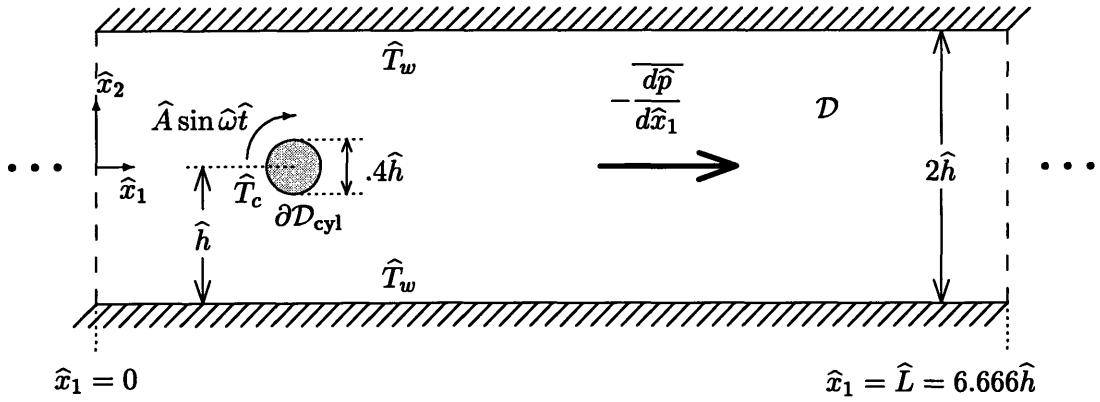


Figure 2-1: One periodic cell of the eddy–promoter channel flow.

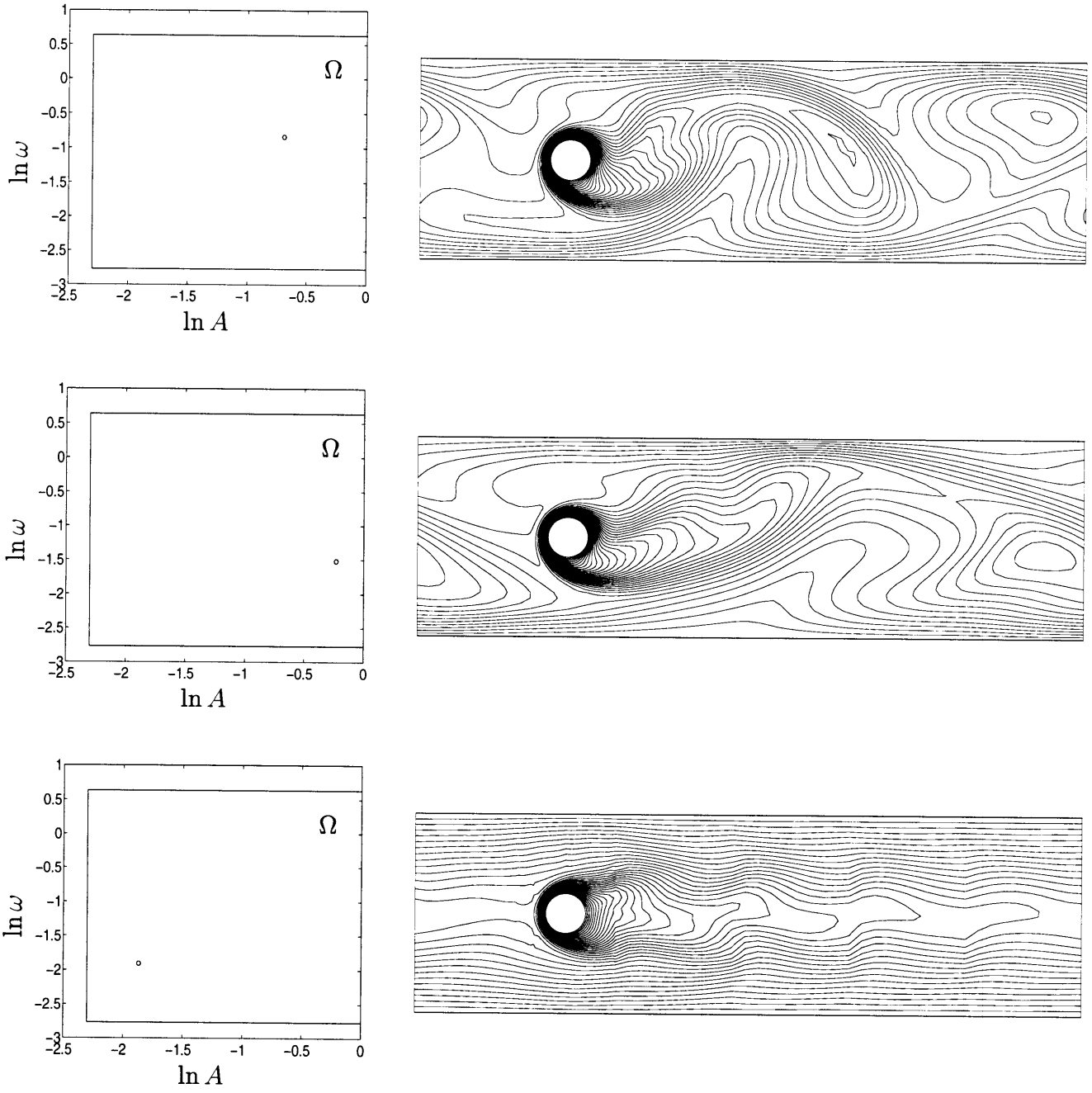


Figure 2-2: Instantaneous isotherms (right) for three different input pairs (left).

Chapter 3

Surrogate Basic Formulation

We describe the algorithm and theory in four steps: surrogate construction; surrogate validation; surrogate–based optimization; and error analysis. We consider in this chapter only the simpler optimization problem (2.10)–(2.11).

3.1 Surrogate Construction

There are various approaches to the construction of surrogates [24]. For a problem as complex as that considered here, a heuristic construction of $\tilde{\Theta}(\mathbf{p})$ and $\tilde{\Pi}(\mathbf{p})$, without reference to the high–fidelity or truth models, will certainly fail to be sufficiently accurate. We choose the simplest possible "empirical" construction, in which $\tilde{\Theta}(\mathbf{p})$ and $\tilde{\Pi}(\mathbf{p})$, are taken to be linear interpolants of corresponding high–fidelity (input, output) pairs on a triangulation of the design space Ω. Note that the surrogate framework makes no assumptions as to the construction method of the surrogate and will accept, and assess, any surrogate input–output functions. We choose to construct the surrogate input–output functions based on high–fidelity, rather than truth, data in order to minimize computational effort; this is discussed further in Section 3.2.

In our example here, we perform a triangulation of Ω based on $N^{co} = 138$ construction nodes $\mathbf{p}_1^{co}, ..., \mathbf{p}_N^{co}$, as shown in Figure 3-1. The placement of the nodes is

based on prior knowledge of anticipated resonance and *self*-generated "*a posteriori*" error estimates described further in Section 3.2. For each construction node, we perform a high–fidelity simulation to compute $\Theta(\mathbf{p}_j^{co})$ and $\Pi(\mathbf{p}_j^{co})$, $j = 1, ..., N^{co}$. We then explicitly form $\widetilde{\Theta}(\mathbf{p})$ and $\widetilde{\Pi}(\mathbf{p})$ as the linear interpolants of the high–fidelity input–output pairs $(\mathbf{p}_j^{co}, \Theta(\mathbf{p}_j^{co}))_{j=1,...,N^{co}}$ and $(\mathbf{p}_j^{co}, \Pi(\mathbf{p}_j^{co}))_{j=1,...,N^{co}}$, respectively. The interpolations, as well as all numerical calculations for this thesis are performed with MATLAB codes. The values of the high–fidelity input–output pairs for the construction nodes are given in Appendix A. Contour plots of $\widetilde{\Theta}(\mathbf{p})$ and $\widetilde{\Pi}(\mathbf{p})$ are shown in Figure 3-2. We see that the temperature is clearly affected by subcritical resonance, while the cylinder power is relatively insensitive to the resonance phenomenon, increasing monotonically with both amplitude and frequency.

3.2 Surrogate Validation

To describe our validation procedure we need to introduce two "prior" functions. First, we define the *importance function* [25] (or "Bayesian prior") $\rho(\mathbf{p}) : \Omega \rightarrow I\!R$ which satisfies

$$\rho(\mathbf{p}) \geq 0, \ \forall \mathbf{p} \in \Omega, \qquad \int_\Omega \rho(\mathbf{p}) d\mathbf{p} = 1. \tag{3.1}$$

The importance function reflects the designer's prejudice as to the ultimate distribution of design points visited; note the sharpness, but *not* the validity, of our error estimates will depend on judicious choice of $\rho(\mathbf{p})$. Associated with this importance function we define the *measure* of any subdomain $\mathcal{D}$ of Ω as

$$\mu(\mathcal{D} \subset \Omega) = \int_\mathcal{D} \rho(\mathbf{p}) d\mathbf{p}, \tag{3.2}$$

which is simply a generalized relative volume. Note the measure of the entire input domain is unity, $\mu(\Omega) = 1$. For our example here, we choose $\rho(\mathbf{p})$ uniform.

We introduce two strictly positive scaling functions $g_\Theta(\mathbf{p}) : \Omega \rightarrow I\!R_+$ and $g_\Pi(\mathbf{p}) :$

$\Omega \rightarrow I\!R_+$, which are prior estimates for the discrepancy between the truth–model input–output functions and the surrogate input–output functions, $|\Theta^T(\mathbf{p}) - \tilde{\Theta}(\mathbf{p})|$ and $|\Pi^T(\mathbf{p}) - \tilde{\Pi}(\mathbf{p})|$, respectively. The scaling functions are intended to serve two purposes: localize the model prediction error in the input space, that is, permit variations in our estimates over Ω; and normalize different outputs, that is, ensure that a large error in one output does not contaminate the joint bound to be developed below. These scaling function priors are also the basis for the error estimators with which the construction points are selected adaptively.

For our current example and the construction of Section 3.1, approximation theory [26] suggests appropriate scaling functions $g_\Theta(\mathbf{p})$ and $g_\Pi(\mathbf{p})$. We consider $\Theta(\mathbf{p})$, with analogous arguments applicable to $\Pi(\mathbf{p})$. First, we write

$$|\Theta^T(\mathbf{p}) - \tilde{\Theta}(\mathbf{p})| \leq |\Theta^T(\mathbf{p}) - \Theta(\mathbf{p})| + |\Theta(\mathbf{p}) - \tilde{\Theta}(\mathbf{p})|. \tag{3.3}$$

We then assume $|\Theta^T(\mathbf{p}) - \Theta(\mathbf{p})| \ll |\Theta(\mathbf{p}) - \tilde{\Theta}(\mathbf{p})|$, and recognize that the latter, since $\tilde{\Theta}(\mathbf{p})$ is a linear interpolant of $\Theta(\mathbf{p})$, is proportional to the Hessian of $\Theta(\mathbf{p})$. This Hessian contribution is then estimated by the Hessian of $\tilde{\Theta}(\mathbf{p})$. More precisely, on each edge γ of the triangulation we compute a scaled jump in $\nabla\tilde{\Theta}(\mathbf{p}) \cdot \hat{n}_p$, where $\hat{n}_p$ is the unit normal on γ; these jumps are then distributed to the nodes of the triangulation; finally, a small constant is included in $g_\Theta(\mathbf{p})$ to guard against unbounded amplification at points where $g_\Theta(\mathbf{p})$ is small or vanishes. Our estimator is very similar to the "jump" part of error estimators for first–order finite element approximations [27]; a detailed description of how the scaling functions are calculated for our example is provided in Appendix B. We calculate the values of the scaling functions at the construction nodes (given in Appendix A), and then explicitely compute $g_\Theta(\mathbf{p})$, for any input vector $\mathbf{p} \in \Omega$, as the linear interpolant of the pairs $(\mathbf{p}_j^{co}, g_\Theta(\mathbf{p}_j^{co}))$, $j = 1, ..., N^{co}$, in the same manner as with the input–output pairs. Contour plots of the functions $g_\Theta(\mathbf{p})$ and $g_\Pi(\mathbf{p})$ will be given in Section 3.4.

We now describe the procedure with which we validate the surrogate model against

the truth model. We first choose parameters $\varepsilon_1 \in]0,1[$ and $\varepsilon_2 \in]0,1[$, and set the validation sample size according to

$$N = \left\lceil \frac{\ln \varepsilon_2}{\ln(1-\varepsilon_1)} \right\rceil, \tag{3.4}$$

where $\lceil x \rceil$ is the smallest integer greater than x. We shall see in Section 3.4 that the parameter $\varepsilon_1 \approx \mathcal{O}(1/N)$ represents the measure of the uncharacterized region — the region of Ω over which the error in $\tilde{\Theta}(\mathbf{p})$ and $\tilde{\Pi}(\mathbf{p})$ is unknown — while $\varepsilon_2 \approx \mathcal{O}(e^{-\varepsilon_1 N})$ is the significance level of our validation statement. We then draw N independent identically distributed (i.i.d.) now–*random* inputs over the input space Ω, $\mathbf{P}_i = (\ln A_i, \ln \omega_i), i = 1, ..., N$, according to the importance function $\rho(\mathbf{p})$. Finally, we perform truth calculations for each of these input vectors to compute the *model prediction error*,

$$U = \max_{i \in \{1,...,N\}} \left(\max \left(\frac{|\Theta^T(\mathbf{P}_i) - \tilde{\Theta}(\mathbf{P}_i)|}{g_\Theta(\mathbf{P}_i)}, \frac{|\Pi^T(\mathbf{P}_i) - \tilde{\Pi}(\mathbf{P}_i)|}{g_\Pi(\mathbf{P}_i)} \right) \right), \tag{3.5}$$

which is simply the validation sample maximum over all (here $K = 2$) outputs.

For our particular example, we choose $\varepsilon_1 = 0.025$ and $\varepsilon_2 = 0.26$ which yields, from (3.4), a validation sample size $N = 53$. For the random validation sample shown in Figure 3-3, the corresponding model prediction error estimate is found from (3.5) to be $U = 0.53$. The magnitude and significance of U will be interpreted in Section 3.4.

Before proceeding, we comment on the relative roles of the high–fidelity and the truth models. Implicitly we assume that the high–fidelity model is, in fact, quite accurate; the very expensive truth model is thus after–the–fact assurance that this hypothesis is, in fact, valid — this final confirmation is critical if the surrogate is to be used with any confidence in the design process. The assumed accuracy of the high–fidelity model is reflected, first, in the construction of the surrogate from high–fidelity, not (much more expensive) truth data, and second, in the specification of the scaling functions $g_\Theta(\mathbf{p})$ and $g_\Pi(\mathbf{p})$. In fact, for the problems studied in this thesis, the

discrepancy between the truth model and the high–fidelity model is commensurate to the discrepancy between the high–fidelity model and the surrogate (0.37 and 0.28 respectively), in contradiction with our previous assumption.

Nested Validation

A more systematic, but also more expensive, procedure is *nested validation*, in which we first validate the high–fidelity outputs against the truth model to obtain $U_{\mathcal{M}^T,\mathcal{M}}$, and then validate the surrogate outputs against the high–fidelity model to obtain a model prediction error $U_{\mathcal{M},\widetilde{\mathcal{M}}}$; the total error is then given by $U = U_{\mathcal{M}^T,\mathcal{M}} + U_{\mathcal{M},\widetilde{\mathcal{M}}}$. To derive the sample sizes for each validation substep, we choose parameters $\varepsilon_{1_{\mathcal{M}^T,\mathcal{M}}}$, $\varepsilon_{1_{\mathcal{M},\widetilde{\mathcal{M}}}}$ and $\varepsilon_{2_{\mathcal{M}^T,\mathcal{M}}}$, $\varepsilon_{2_{\mathcal{M},\widetilde{\mathcal{M}}}}$ so that $\varepsilon_{1_{\mathcal{M}^T,\mathcal{M}}} + \varepsilon_{1_{\mathcal{M},\widetilde{\mathcal{M}}}} = \varepsilon_1$ and $\varepsilon_{2_{\mathcal{M}^T,\widetilde{\mathcal{M}}}} + \varepsilon_{2_{\widetilde{\mathcal{M}},\mathcal{M}}} = \varepsilon_2$, and then set $N_{\mathcal{M}^T,\mathcal{M}}$ and $N_{\mathcal{M},\widetilde{\mathcal{M}}}$ according to (3.4). The advantage of nested validation is that one can accurately decompose the error into high–fidelity and surrogate components. Thus one proceeds with surrogate construction only once a sufficiently accurate high–fidelity model is obtained, with the truth model then decoupled from all subsequent surrogate considerations. Combined with comparison, sequential, and adaptive construction–validation procedures discussed in [28, 29], this approach can be quite attractive. Unfortunately, the requisite sample sizes are considerably larger than those required by non-nested validation. For example, in the problem studied here, for an equal split of ε_1 and ε_2, giving $\varepsilon_{1_{\mathcal{M}^T,\mathcal{M}}} = \varepsilon_{1_{\mathcal{M},\widetilde{\mathcal{M}}}} = \frac{\varepsilon_1}{2} = 0.0125$ and $\varepsilon_{2_{\mathcal{M}^T,\mathcal{M}}} = \varepsilon_{2_{\mathcal{M},\widetilde{\mathcal{M}}}} = \frac{\varepsilon_2}{2} = 0.13$, we would need $N_{\mathcal{M}^T,\mathcal{M}} = 163$ very expensive and $N_{\mathcal{M},\widetilde{\mathcal{M}}} = 163$ expensive calculations (instead of only $N = 53$ very expensive ones).

3.3 Surrogate–Based Design

As our example here, we choose for the *design problem* the minimum temperature problem of (2.10)–(2.11). Following the recipe described on Section 2.3.2 we replace this difficult problem with the much simpler problem in which the surrogate is sub-

stituted for the truth model: find $\tilde{\mathbf{p}}^*$ such that

$$\tilde{\theta}_{\min} \equiv \tilde{\Theta}(\tilde{\mathbf{p}}^*) = \min_{\{\mathbf{p}\in\Omega\}} \tilde{\Theta}(\mathbf{p}). \tag{3.6}$$

In fact, for this particularly simple optimization problem and our piecewise–linear surrogate input–output functions, $\tilde{\theta}_{\min}$ is simply the minimum of $\tilde{\Theta}(\mathbf{p})$ over the construction points, $\mathbf{p}_j^{co}$, $j = 1, ..., N^{co}$. We thus readily find for the surrogate–based minimizer $\tilde{\mathbf{p}}^* = (\ln \tilde{A}^*, \ln \tilde{\omega}^*) = (-0.69, -0.84)$, corresponding to $\tilde{\theta}_{\min} = 0.82$ — an 18% improvement over the flow *without* cylinder oscillations (note that $\tilde{\theta}_{\max} \equiv \max_{\{\mathbf{p}\in\Omega\}}\tilde{\Theta}(\mathbf{p}) = 1.03$). The power requirement for the minimum temperature is $\tilde{\Pi}(\tilde{\mathbf{p}}^*) = 5.35$.

3.4 Error Analysis

We first derive the validation statement in Section 3.4.1 that then leads to the investigation of predictability and optimality in Sections 3.4.2 and 3.4.3, respectively.

3.4.1 Validation Statement

Critical to our analysis is the *uncharacterized region*, Υ, defined as the (random) subset of Ω for which the surrogate error is greater than U:

$$\Upsilon = \left\{ \mathbf{p} \in \Omega \,\middle|\, \max\left(\frac{|\Theta^T(\mathbf{p}) - \tilde{\Theta}(\mathbf{p})|}{g_\Theta(\mathbf{p})}, \frac{|\Pi^T(\mathbf{p}) - \tilde{\Pi}(\mathbf{p})|}{g_\Pi(\mathbf{p})} \right) > U \right\}. \tag{3.7}$$

We then define a random variable $Z = \mu(\Upsilon)$ which is the measure of the uncharacterized region Υ. It is shown in Appendix C that the cumulative distribution function of Z satisfies

$$Pr\{Z < z\} \equiv F_Z(z) \geq \hat{F}_Z(z), \tag{3.8}$$

where

$$\widehat{F}_Z(z) = 1 - (1-z)^N; \tag{3.9}$$

variants of (3.9), which is effectively a classical tolerance limit result [22, 30], are given in [4, 5].

It thus follows that for $z = \varepsilon_1$, and N given by (3.4), that $\widehat{F}_Z(\varepsilon_1) \geq 1 - \varepsilon_2$, and thus

$$Pr\{\mu(\Upsilon) < \varepsilon_1\} \geq 1 - \varepsilon_2, \tag{3.10}$$

with

$$\begin{cases} |\Theta^T(\mathbf{p}) - \tilde{\Theta}(\mathbf{p})| \leq U g_\Theta(\mathbf{p}) \\ |\Pi^T(\mathbf{p}) - \tilde{\Pi}(\mathbf{p})| \leq U g_\Pi(\mathbf{p}) \end{cases}, \quad \forall \mathbf{p} \in \Omega \setminus \Upsilon. \tag{3.11}$$

The probability in (3.10) is over the validation sample space: in a frequentistic interpretation, $\mu(\Upsilon) < \varepsilon_1$ will be true in a fraction $1 - \varepsilon_2$ of N_{ensemble} validation trials as $N_{\text{ensemble}} \to \infty$. The difficulty with (3.10)–(3.11) is that although we know that Υ is volumetrically small, we do not know its location. The cost of reducing the size of the uncharacterized region and/or increasing the confidence for its size, is reflected on the number of validation points required.

For our particular example, the *validation statement* reads: With probability greater than $1 - \varepsilon_2 = 0.74$, the surrogate error is given by

$$\begin{aligned} |\Theta^T(\mathbf{p}) - \tilde{\Theta}(\mathbf{p})| &\leq 0.53\, g_\Theta(\mathbf{p}), \\ |\Pi^T(\mathbf{p}) - \tilde{\Pi}(\mathbf{p})| &\leq 0.53\, g_\Pi(\mathbf{p}), \end{aligned} \tag{3.12}$$

for any input $\mathbf{p}$ in a region $\Omega \setminus \Upsilon$ which is *at least* of measure $1 - \varepsilon_1 = 0.975$ (97.5% of Ω). Figure 3-4 shows contours of the surrogate error $|\Theta^T(\mathbf{p}) - \tilde{\Theta}(\mathbf{p})|$, and an illustrative region of measure $\varepsilon_1 = 0.025$ which *may* constitute Υ. Note from (3.11) that contours of $|\Theta^T(\mathbf{p}) - \tilde{\Theta}(\mathbf{p})|$ are, in fact, contours of our scaling function $g_\Theta(\mathbf{p})$.

Nested–Validation Statement

We shall demonstrate here that we can provide analogous error statements for the nested validation procedure. We define $\Upsilon_{\mathcal{M}^T,\mathcal{M}}$ and $\Upsilon_{\mathcal{M},\widetilde{\mathcal{M}}}$ as the (random) subsets of Ω for which the surrogate error is greater than $U_{\mathcal{M}^T,\mathcal{M}}$ and $U_{\mathcal{M},\widetilde{\mathcal{M}}}$, respectively. It then follows that

$$\begin{aligned} Pr\{\mu(\Upsilon_{\mathcal{M}^T,\mathcal{M}}) < \varepsilon_{1\mathcal{M}^T,\mathcal{M}}\} \geq 1 - \varepsilon_{2\mathcal{M}^T,\mathcal{M}} \ , \\ Pr\{\mu(\Upsilon_{\mathcal{M},\widetilde{\mathcal{M}}}) < \varepsilon_{1\mathcal{M},\widetilde{\mathcal{M}}}\} \geq 1 - \varepsilon_{2\mathcal{M},\widetilde{\mathcal{M}}} \ . \end{aligned} \tag{3.13}$$

Given that $\varepsilon_{1\mathcal{M}^T,\mathcal{M}} + \varepsilon_{1\mathcal{M},\widetilde{\mathcal{M}}} = \varepsilon_1$, $\varepsilon_{2\mathcal{M}^T,\widetilde{\mathcal{M}}} + \varepsilon_{2\widetilde{\mathcal{M}},\mathcal{M}} = \varepsilon_2$, and that the validation steps are independent, we have

$$\begin{aligned} Pr\{\mu(\Upsilon_{\mathcal{M}^T,\mathcal{M}} \cup \Upsilon_{\mathcal{M},\widetilde{\mathcal{M}}}) < \varepsilon_1\} &= (Pr\{\mu(\Upsilon_{\mathcal{M}^T,\mathcal{M}}) < \varepsilon_{1\mathcal{M}^T,\mathcal{M}}\})(Pr\{\mu(\Upsilon_{\mathcal{M},\widetilde{\mathcal{M}}}) < \varepsilon_{1\mathcal{M},\widetilde{\mathcal{M}}}\}) \\ &\geq (1 - \varepsilon_{2\mathcal{M}^T,\mathcal{M}})(1 - \varepsilon_{2\mathcal{M},\widetilde{\mathcal{M}}}) \\ &\geq 1 - (\varepsilon_{2\mathcal{M}^T,\mathcal{M}} + \varepsilon_{2\mathcal{M},\widetilde{\mathcal{M}}}) + \varepsilon_{2\mathcal{M}^T,\mathcal{M}} \cdot \varepsilon_{2\mathcal{M},\widetilde{\mathcal{M}}} \\ &> 1 - \varepsilon_2. \end{aligned} \tag{3.14}$$

We can therefore state that with probability greater than $1 - \varepsilon_2$, the surrogate error is given by

$$\begin{aligned} |\Theta^T(\mathbf{p}) - \tilde{\Theta}(\mathbf{p})| \leq (U_{\mathcal{M}^T,\mathcal{M}} + U_{\mathcal{M},\widetilde{\mathcal{M}}})\, g_\Theta(\mathbf{p}), \\ |\Pi^T(\mathbf{p}) - \tilde{\Pi}(\mathbf{p})| \leq (U_{\mathcal{M}^T,\mathcal{M}} + U_{\mathcal{M},\widetilde{\mathcal{M}}})\, g_\Pi(\mathbf{p}) \end{aligned} \tag{3.15}$$

for any input $\mathbf{p}$ in a region $\Omega \setminus (\Upsilon_{\mathcal{M}^T,\mathcal{M}} \cup \Upsilon_{\mathcal{M},\widetilde{\mathcal{M}}})$ which is at least of measure $1 - \varepsilon_1$.

3.4.2 Predictability Analysis

The deficiency with the validation statement is that, since the location of Υ is unknown, $\tilde{\mathbf{p}}^*$ might be within Υ, and the design process could be seriously misled. To partially remedy this situation, in *predictability analysis* we examine how the truth model performs in the vicinity of the minimizer $\tilde{\mathbf{p}}^*$. To this end we define a *predic-*

tion neighborhood, $\mathcal{P}(\tilde{\mathbf{p}}^*, z)$, as a region of measure $z = \kappa\varepsilon_1$, $\kappa \geq 1$, that (typically) contains $\tilde{\mathbf{p}}^*$. To obtain a unique prediction neighborhood we introduce a distance function $\delta(\mathbf{p})$, $\mathbf{p} \in \Omega$, and select $\mathcal{P}(\tilde{\mathbf{p}}^*, z)$ to be the region $\mathcal{R}$ of measure z within some class of domains $\mathcal{C}$ which minimizes

$$r_{\mathcal{R}} = \max_{\{\mathbf{p}\in\mathcal{R}\}} \delta(\mathbf{p}). \tag{3.16}$$

A schematic representation of a prediction neighborhood is shown in Figure 3-5.

We can then make the following *nonparametric predictability statement*: With probability greater than $1 - \varepsilon_2$, there is a subset Γ of $\mathcal{P}(\tilde{\mathbf{p}}^*, \kappa\varepsilon_1)$, $\Gamma = \mathcal{P}(\tilde{\mathbf{p}}^*, \kappa\varepsilon_1) \setminus \Upsilon$, of measure *strictly greater* than zero, such that

$$\begin{cases} B_\theta^\ell \leq \Theta^T(\mathbf{p}) - \tilde{\theta}_{\min} \leq B_\theta^u \\ B_\varpi^\ell \leq \Pi^T(\mathbf{p}) - \tilde{\Pi}(\tilde{\mathbf{p}}^*) \leq B_\varpi^u \end{cases}, \ \forall \mathbf{p} \in \Gamma. \tag{3.17}$$

Here B_θ^ℓ, B_θ^u, and B_ϖ^ℓ, B_ϖ^u, are lower and upper bounds given by

$$B_\theta^\ell = \min_{\{\mathbf{p}\in\mathcal{P}(\tilde{\mathbf{p}}^*,\kappa\varepsilon_1)\}} (-U g_\Theta(\mathbf{p}) + [\tilde{\Theta}(\mathbf{p}) - \tilde{\theta}_{\min}]), \tag{3.18}$$

$$B_\theta^u = \max_{\{\mathbf{p}\in\mathcal{P}(\tilde{\mathbf{p}}^*,\kappa\varepsilon_1)\}} (U g_\Theta(\mathbf{p}) + [\tilde{\Theta}(\mathbf{p}) - \tilde{\theta}_{\min}]), \tag{3.19}$$

and

$$B_\varpi^\ell = \min_{\{\mathbf{p}\in\mathcal{P}(\tilde{\mathbf{p}}^*,\kappa\varepsilon_1)\}} (-U g_\Pi(\mathbf{p}) + [\tilde{\Pi}(\mathbf{p}) - \tilde{\Pi}(\tilde{\mathbf{p}}^*)]), \tag{3.20}$$

$$B_\varpi^u = \max_{\{\mathbf{p}\in\mathcal{P}(\tilde{\mathbf{p}}^*,\kappa\varepsilon_1)\}} (U g_\Pi(\mathbf{p}) + [\tilde{\Theta}(\mathbf{p}) - \tilde{\Pi}(\tilde{\mathbf{p}}^*)]), \tag{3.21}$$

respectively. Note that no appeals to the truth or high–fidelity model are required to compute (3.18)–(3.21). We can, of course, also compute the predictability of any *function* of the selected outputs.

The proof of (3.18)–(3.21) is very simple given the validation statement (3.10)–

(3.11). First, we note that for $\mathbf{p} \in \mathcal{P}(\tilde{\mathbf{p}}^*, k\varepsilon_1)$,

$$\Theta^T(\mathbf{p}) - \tilde{\theta}_{\min} = [\Theta^T(\mathbf{p}) - \tilde{\Theta}(\mathbf{p})] + [\tilde{\Theta}(\mathbf{p}) - \tilde{\theta}_{\min}], \tag{3.22}$$

$$\Pi^T(\mathbf{p}) - \tilde{\Pi}(\tilde{\mathbf{p}}^*) = [\Pi^T(\mathbf{p}) - \tilde{\Pi}(\mathbf{p})] + [\tilde{\Pi}(\mathbf{p}) - \tilde{\Pi}(\tilde{\mathbf{p}}^*)]. \tag{3.23}$$

Now, at all points $\mathbf{p} \in \Gamma$, $-Ug_\Theta(\mathbf{p}) < \Theta^T(\mathbf{p}) - \tilde{\Theta}(\mathbf{p}) < Ug_\Theta(\mathbf{p})$ and $-Ug_\Pi(\mathbf{p}) < \Pi^T(\mathbf{p}) - \tilde{\Pi}(\mathbf{p}) < Ug_\Pi(\mathbf{p})$. Furthermore, at all points $\mathbf{p} \in \Gamma$, $[\tilde{\Theta}(\mathbf{p}) - \tilde{\theta}_{\min}]$ is certainly bounded from below by

$$\min_{\{\mathbf{p} \in \mathcal{P}(\tilde{\mathbf{p}}^*, k\varepsilon_1)\}} [\tilde{\Theta}(\mathbf{p}) - \tilde{\theta}_{\min}],$$

from which (3.18) directly follows; similar arguments prove (3.19), (3.20), and (3.21). Note that it is critical that the *same* Γ applies to both the temperature and power; this in turn follows from the choice

$$\mathcal{E}_Y(\mathbf{p}) = \max\left(\frac{|\Theta^T(\mathbf{p}) - \tilde{\Theta}(\mathbf{p})|}{g_\Theta(\mathbf{p})}, \frac{|\Pi^T(\mathbf{p}) - \tilde{\Pi}(\mathbf{p})|}{g_\Pi(\mathbf{p})}\right), \tag{3.24}$$

in Appendix A.

We demonstrate the concept of predictability with two particular examples from our illustrative application.

Example *a*

In the first example, we make no assumptions on the properties of $\Theta^T(\mathbf{p})$ and $\Pi^T(\mathbf{p})$. We set $\kappa = 1$, and choose our prediction neighborhood $\mathcal{P}^a(\tilde{\mathbf{p}}^*, \varepsilon_1)$ to be the ellipse that minimizes the upper bound B_θ^u given by (3.19); $\mathcal{P}^a(\tilde{\mathbf{p}}^*, \varepsilon_1)$ is shown in Figure 3-6. We can then state that, with probability greater than $1 - \varepsilon_2 = 0.74$, there is a region $\Gamma \subset \mathcal{P}^a(\tilde{\mathbf{p}}^*, \varepsilon_1)$ of positive measure, $\mu(\Gamma) > 0$, such that for all $\mathbf{p} \in \Gamma$,

$$\begin{cases} B_\theta^\ell = -0.08 \;\leq\; \Theta^T(\mathbf{p}) - \tilde{\theta}_{\min} \;\leq\; B_\theta^u = 0.13 \\ B_\varpi^\ell = -4.56 \;\leq\; \Pi^T(\mathbf{p}) - \tilde{\Pi}(\tilde{\mathbf{p}}^*) \;\leq\; B_\varpi^u = 5.30. \end{cases} \tag{3.25}$$

Additionally, we can show that the expected measure of the region Γ is 61% of the

prediction neighborhood $\mathcal{P}^a(\tilde{\mathbf{p}}^*, \varepsilon_1)$ [31]. Recalling that $\tilde{\theta}_{\min} = 0.82$ and $\tilde{\theta}_{\max} = 1.03$, we see that the predictability in the temperature is not overly good, though the bounds on the power are relatively sharp.

The statement includes several notions. First, there is the notion of predictability: there *exists* a $\mathbf{p} \in \Gamma$ such that the bounds (3.25) hold. We denote the predictability gap (say, for the temperature) as $B_\theta^u - B_\theta^\ell$, and note that this gap has two contributions: the first term, which measures the *accuracy* of the surrogate and is proportional to U; and the second term, which measures the *sensitivity* of the surrogate about $\mathbf{p} = \tilde{\mathbf{p}}^*$. The former vanishes as we improve our construction, that is, increase N^{co}; the latter vanishes as we perform additional validation, that is, increase N so as to decrease ε_1. Note that as we decrease ε_1, $\mathcal{P}^a(\tilde{\mathbf{p}}^*, \varepsilon_1)$ shrinks to $\tilde{\mathbf{p}}^*$, and hence for $\mathbf{p} \in \Gamma$, $\Theta^T(\mathbf{p}) - \tilde{\theta}_{\min} \to 0$, assuming continuity of $\tilde{\Theta}(\mathbf{p})$ at $\tilde{\mathbf{p}}^*$.

In addition to predictability, we have a notion of *genericity*, in that (3.17) applies for all $\mathbf{p} \in \Gamma$, a region of *non-zero* measure. We can even derive a weak sense of *stability* by appealing to a theorem of set theory [32] related to density points: if we introduce $h \times h$ open neighborhoods of $\mathbf{p}$, $\mathcal{N}(\mathbf{p})$, then for all $\mathbf{p} \in \Gamma$, except perhaps a set of measure zero,

$$\frac{\mu(\Gamma \cap \mathcal{N}(\mathbf{p}))}{\mu(\mathcal{N}(\mathbf{p}))} \to 1 \quad \text{as} \quad h \to 0. \tag{3.26}$$

In words, (3.26) ensures that most $\mathbf{p} \in \Gamma$ are surrounded by mostly other $\mathbf{p} \in \Gamma$, thus providing stability to random infinitesimal disturbances. Stability is critical in engineering analyses in which manufacturing variation and external disturbances may introduce input noise.

Finally, we note that (3.17) provides a sense of design localization, that is, an indication of the "ball" in which designs can be found for which predicted behavior will obtain. In summary, we can state that, although $\Theta^T(\tilde{\mathbf{p}}^*)$ will *not* equal $\tilde{\Theta}(\tilde{\mathbf{p}}^*) = \tilde{\theta}_{\min}$, there are many design points $\mathbf{p}'$ "near" $\tilde{\mathbf{p}}^*$ — in $\mathcal{P}^a(\tilde{\mathbf{p}}^*, \varepsilon_1)$ — for which system performance $\Theta^T(\mathbf{p}')$ is "near" $\tilde{\theta}_{\min}$, as measured by the predictability gap $B_\theta^u - B_\theta^\ell$.

Example *b*

In our second example we focus exclusively on the temperature, and we further assume that $\Theta^T(\mathbf{p})$ is locally quasi-convex [33]. We recall that if $\Theta^T(\mathbf{p})$ is quasi-convex in a convex region Ω_{qc}, the level sets of $\Theta^T(\mathbf{p})$ in Ω_{qc} are convex, and that for any $\mathbf{p}_1 \in \Omega_{qc}$, $\mathbf{p}_2 \in \Omega_{qc}$, and $\alpha \in]0,1[$, $\Theta^T(\alpha\mathbf{p}_1 + (1-\alpha)\mathbf{p}_2) \leq \max(\Theta^T(\mathbf{p}_1), \Theta^T(\mathbf{p}_2))$. We set $\kappa = 2$, and choose $\mathcal{P}^b(\tilde{\mathbf{p}}^*, 2\varepsilon_1)$ to be that ellipse of measure $2\varepsilon_1$ with center $\tilde{\mathbf{p}}^*$ that minimizes B_θ^u of (3.19); $\mathcal{P}^b(\tilde{\mathbf{p}}^*, 2\varepsilon_1)$ is shown in Figure 3-7.

We can then state that, with probability greater than $1 - \varepsilon_2 = 0.74$, there exists a convex region $\Gamma_{Co} \subset \mathcal{P}^b(\tilde{\mathbf{p}}^*, 2\varepsilon_1)$ of measure $\mu(\Gamma_{Co}) > \varepsilon_1$ such that, for all $\mathbf{p} \in \Gamma_{Co}$,

$$\Theta^T(\mathbf{p}) - \tilde{\theta}_{\min} \leq B_\theta^{u,qc}, \tag{3.27}$$

where

$$B_\theta^{u,qc} = \max_{\{\mathbf{p} \in \mathcal{P}(\tilde{\mathbf{p}}^*, 2\kappa\varepsilon_1)\}} (U g_\Theta(\mathbf{p}) + \tilde{\Theta}(\mathbf{p}) - \tilde{\theta}_{\min}) = 0.16. \tag{3.28}$$

Furthermore, we can state that $\tilde{\mathbf{p}}^* \in \Gamma_{Co}$. We thus obtain a much sharper sense of stability in this case, as there must exist an open neighborhood about $\tilde{\mathbf{p}}^*$ of nonzero measure in which (3.27) is satisfied. There is no improvement, however, in design localization or predictability gap.

The proof of (3.27) is relatively simple. First, we note that (3.27) applies for $\Gamma = \mathcal{P}^b(\tilde{\mathbf{p}}^*, 2\varepsilon_1) \backslash \Upsilon$ *without* the assumption of quasi-convexity. But by quasi-convexity, (3.27) thus also applies in Γ_{Co} = Convex Hull(Γ). To see this, we recall that Γ_{Co} is the union of of all vectors $(\alpha\mathbf{p}_1 + (1-\alpha)\mathbf{p}_2)$, $(\mathbf{p}_1 \in \Gamma, \mathbf{p}_2 \in \Gamma)$, and hence, for any $\mathbf{p} \in \Gamma_{Co}$, $\Theta^T(\mathbf{p}) \leq \tilde{\theta}_{\min} + B_\theta^{u,qc}$, since $\Theta^T(\mathbf{p})$ is bounded on any segment $\overline{\mathbf{p_1 p_2}}$ by $\max(\Theta^T(\mathbf{p}_1), \Theta^T(\mathbf{p}_2))$. Note that this does *not* require that $|\Theta^T(\mathbf{p}) - \tilde{\Theta}(\mathbf{p})|$ be quasi-convex. Since $\Gamma \subset \Gamma_{Co}$, $\mu(\Gamma_{Co}) \geq \mu(\Gamma) > \varepsilon_1$, which completes the first part of the proof.

For the second part of the proof, we show that $\tilde{\mathbf{p}}^* \in \Gamma_{Co}$. This part of the proof, unlike the first part, requires $\kappa \geq 2$ and point–symmetry of $\mathcal{P}^b(\tilde{\mathbf{p}}^*, 2\varepsilon_1)$ about $\tilde{\mathbf{p}}^*$.

We then proceed by contradiction, and assume that $\tilde{\mathbf{p}}^* \notin \Gamma_{Co}$. There is, then, a separating hyperplane through $\tilde{\mathbf{p}}^*$ for which (say) Γ_{Co} lies entirely to the "left." But by point–symmetry of $\mathcal{P}^b(\tilde{\mathbf{p}}^*, 2\varepsilon_1)$, this implies that $\mu(\Gamma_{Co}) \leq \varepsilon_1$, which contradicts the first part of the proof. (The argument is, clearly, more complex if $\rho(\mathbf{p})$ is not uniform.) Note that the proof as stated only requires $\Theta^T(\mathbf{p})$ to be quasi-convex in $\mathcal{P}^b(\tilde{\mathbf{p}}^*, 2\varepsilon_1)$. The proof can be extended to permit the selection of a specified footprint $\mathcal{F}$ such that $\mathcal{F} \subset \Gamma_{Co}$, which is related to the problem of bounded but unknown noise [6].

Finally, we remark that with the assumption of quasi-convexity, results similar to (3.27) can now be obtained by ε–net techniques [34]. In particular, if we enumerate those M–simplices constructed from vertices of the validation sample which contain $\tilde{\mathbf{p}}^*$, then from Jensen's inequality [22], $\Theta^T(\tilde{\mathbf{p}}^*)$ can be bounded by the minimum over all such simplices of the maximum of $\Theta^T(\mathbf{p})$ at the associated $M+1$ vertices. We have not investigated the relative efficiency of the probabilistic and deterministic approaches.

3.4.3 Optimality Analysis

We now turn to the issue of optimality, focusing exclusively on the temperature. In earlier work [4, 5] constructive results for a region which contains the true minimizer $\tilde{\mathbf{p}}^{*T}$ based on assumptions of quasi-convexity on $\Theta^T(\mathbf{p})$ were obtained. In practice, these regions are quite large, and the assumption of quasi-convexity is thus dubious. We consider a different approach here, in which we ascertain the additional computations required to *improve* upon the surrogate result.

We first compute

$$\Phi_o = \min_{\{\mathbf{p}\in\Omega,\, \widetilde{\Theta}(\mathbf{p})-Ug_\Theta(\mathbf{p})\leq\theta\leq\widetilde{\Theta}(\mathbf{p})+Ug_\Theta(\mathbf{p}),\, \widetilde{\Pi}(\mathbf{p})-Ug_\Pi(\mathbf{p})\leq\varpi\leq\widetilde{\Pi}(\mathbf{p})+Ug_\Pi(\mathbf{p})\}} \phi(\mathbf{p},\theta,\varpi). \quad (3.29)$$

We next consider a new sequence of i.i.d. random vectors $\widehat{\mathbf{P}}_1, ..., \widehat{\mathbf{P}}_j$ drawn according

to the importance function $\rho(\mathbf{p})$ with which we have previously validated $\tilde{\Theta}(\mathbf{p})$. We then introduce the random variable

$$L = \min j \text{ such that } \Phi(\hat{\mathbf{P}}_j) < \Phi_o, \tag{3.30}$$

for which we will show that

$$Pr\{L > \ell\} \geq \frac{N}{N+\ell}, \tag{3.31}$$

where N is the validation sample size. From (3.31) it follows that, if we select $\varepsilon_L \in]0,1[$, and set

$$m \leq \left(\frac{1}{\varepsilon_L} - 1\right) N, \tag{3.32}$$

then the probability of drawing a sequence of random input vectors $\hat{\mathbf{P}}_1, ..., \hat{\mathbf{P}}_m$ according to $\rho(\mathbf{p})$ and finding a point such that $\Phi(\hat{\mathbf{P}}_j) < \Phi_o$ is less than $1 - \varepsilon_L$. Note that the sample space associated with (3.31) is now the combined validation sample $\hat{\mathbf{P}}_i$, $i, ..., N$, and post-validation sample $\hat{\mathbf{P}}_1, ..., \hat{\mathbf{P}}_m$.

To prove (3.32), we first note that $\mathbf{p}$ such that $\Phi^T(\mathbf{p}) = \phi(\mathbf{p}, \Theta^T(\mathbf{p}), \Pi^T(\mathbf{p}))$ $< \Phi_o$ must reside in some subset of Υ, $\mathcal{R}$, since if $\mathbf{p} \in \Omega \setminus \Upsilon$, $\tilde{\Theta}(\mathbf{p}) - U g_\Theta(\mathbf{p}) \leq \Theta^T(\mathbf{p}) \leq \tilde{\Theta}(\mathbf{p}) + U g_\Theta(\mathbf{p})$, and $\tilde{\Pi}(\mathbf{p}) - U g_\Pi(\mathbf{p}) \leq \Pi^T(\mathbf{p}) \leq \tilde{\Pi}(\mathbf{p}) + U g_\Pi(\mathbf{p})$. (It follows that if $\mu(\Upsilon) = 0$, then $\Phi_o \leq \operatorname{ess\,inf}_{\mathbf{p}\in\Omega} \Phi^T(\mathbf{p})$.) With some attention to detail, we can identify mappings $z \to \mathcal{R}(z)$ and $z \to h(z)$, where $z = \mu(\Upsilon)$, and $h(z) = \mu(\mathcal{R}) \leq z$. We can then write

$$\begin{aligned} Pr\{L > \ell\} &= \int_0^1 Pr\{L > \ell \mid z\} dF_Z & (3.33) \\ &= \int_0^1 (1 - h(z))^\ell dF_Z & (3.34) \\ &\geq \int_0^1 (1 - z)^\ell dF_Z & (3.35) \\ &\geq \int_0^1 (1 - z)^\ell d\hat{F}_Z & (3.36) \end{aligned}$$

$$= \int_0^1 (1-z)^\ell N(1-z)^{N-1} dz \tag{3.37}$$

$$= \frac{N}{N+\ell}. \tag{3.38}$$

(The slow $1/\ell$ decay results in logarithmic divergence: the expectation of L is infinite.) Here (3.33) denotes the Riemann–Stieltjes integral; (3.34) reflects $Pr\{\widehat{\mathbf{P}}_j \notin \mathcal{R}(z), j = 1, ..., \ell\} = (1-h(z))^\ell$, which is then bounded from below in (3.35) by $(1-z)^\ell$; finally, (3.36) recognizes that for a decreasing function of z, $q(z)$, $\int_0^1 q(z) dF_Z \geq \int_0^1 q(z) d\widehat{F}_Z$, as F_Z never lies underneath $\widehat{F}_Z$ (see Appendix C). The evaluation of $\widehat{F}_Z(z)$ then gives (3.37) and (3.38).

As a concrete example, we continue our eddy–promoter application, with $\Phi^T(\mathbf{p}) = \Theta^T(\mathbf{p})$ as in the test case of Sections 3.3 and Section 3.4.2. It is readily computed from (3.29) that $\Phi_o = 0.77$ for the surrogate and $N = 53$–point validation of Sections 3.1 and 3.2, respectively. We then set $\varepsilon_L = 0.2$, from which it follows that, even if we perform $m = (1/\varepsilon_L - 1)\, N = 212$ truth evaluations, with probability greater than 0.2 there will be no points $\widehat{\mathbf{P}}_j$ which improve upon Φ_o. If Φ_o is considered sufficiently *unimproved* with respect to $\widetilde{\theta}_{\min}$ (or perhaps, more conservatively, $\widetilde{\theta}_{\min} + B_\theta^u$), then the additional expense is clearly not warranted. If $\Phi_o \ll \widetilde{\theta}_{\min}$, then our surrogate model is clearly not adequate, and must be improved. Note that we do not actually advocate random search; rather, we exploit random search as a measure of computational effort. This hypothesis is not too naive if, in fact, the surrogate model captures the "smooth" behavior of $\Theta^T(\mathbf{p})$, with only isolated basins unresolved; the latter is as easily found by random search as by more sophisticated procedures. Indeed, (3.31) may be viewed as an alternative to the usual basin of attraction results for (multistart) random search procedures [25].

3.5 Surrogate Limitations

While the surrogate approach offers a complete framework for engineering design, it will only be effective for low–dimensional input spaces, that is, M small. To understand why, we recall that the upper predictability bound for (say) the temperature is given by

$$B^u_\theta = \max_{\{\mathbf{p}\in\mathcal{P}(\widetilde{\mathbf{p}}^*,\kappa\varepsilon_1)\}} (Ug_\Theta(\mathbf{p}) + [\widetilde{\Theta}(\mathbf{p}) - \widetilde{\theta}_{\min}]).$$

The second, sensitivity, term will typically scale with the "radius" of the prediction neighborhood as defined by (3.16); the radius is, thus, a measure of both design localization and predictability.

We now take the case in which $\Omega =]0,1[^M$, the prediction neighborhoods are hypercubes, and the importance function is uniform. The hypervolume of the prediction neighborhood must be at least of measure ε_1. It follows that the side of the hypercube must be at least $(\varepsilon_1)^{1/M}$. By definition, the radius of the prediction neighborhoods, $r_\mathcal{P}$, must then be at least $\frac{1}{2}(\varepsilon_1)^{1/M}$)$r_\mathcal{P}$ is minimized when $\widetilde{\mathbf{p}}^*$ is located in the center of the hypercube, case in which it is equal to $\frac{\sqrt{M}}{2}(\varepsilon_1)^{1/M}$). Typically, $r_\mathcal{P}$, is much larger, especially in the case where $\widetilde{\mathbf{p}}^*$ is near a boundary of Ω. In fact, for *any* prediction neighborhood $\mathcal{P}$ there must exist a point $\mathbf{p}' \in \mathcal{P}$ such that, in at least one input coordinate direction j , $|p'_j - \widetilde{p}^*_j| \geq \frac{1}{2}(\varepsilon_1)^{1/M}$. It is simple to see why this is true by contradiction: let's assume that for all coordinate directions j and for all $\mathbf{p}' \in \mathcal{P}$, $|p'_j - \widetilde{p}^*_j| < \frac{1}{2}(\varepsilon_1)^{1/M}$. Then, $r_\mathcal{P} < \frac{1}{2}(\varepsilon_1)^{1/M}$, and consequently the prediction neighborhood can be confined in a hypercube of side–length $= 2r_\mathcal{P} < (\varepsilon_1)^{1/M}$. The volume of the bounding hypercube must then be less than ε_1; this cannot hold since the volume of the prediction neighborhood is at least ε_1. Therefore, $r_\mathcal{P} \geq \frac{1}{2}(\varepsilon_1)^{1/M}$. Since $\varepsilon_1 \sim O(\frac{1}{N})$, it follows that at fixed ε_2, $(\varepsilon_1)^{1/M} \sim (\frac{1}{N})^{1/M}$. Thus as $M \to \infty$ we must either choose $N \sim n^M, n$ a fixed positive integer — which is clearly prohibitive — or accept that $r_\mathcal{P} \to O(1)$ — clearly unattractive, as we will lose design localization, and through the sensitivity contribution to (3.19), predictability. (Note that ε_2

decreases rapidly with N; ε_1, and hence the size of the uncharacterized region, is the problem.)

A critical assumption in the above analysis is that $\rho(\mathbf{p})$ is uniform. It is clear that if we can anticipate where the $\tilde{\mathbf{p}}^*$ will lie, $\rho(\mathbf{p})$ can be focused in that area, and design localization and predictability can be greatly improved. One approach is reliance on the designer to specify an appropriate $\rho(\mathbf{p})$; this, however, permits no *a posteriori* analysis of when the technique will perform well. A more systematic approach is described in the next chapter.

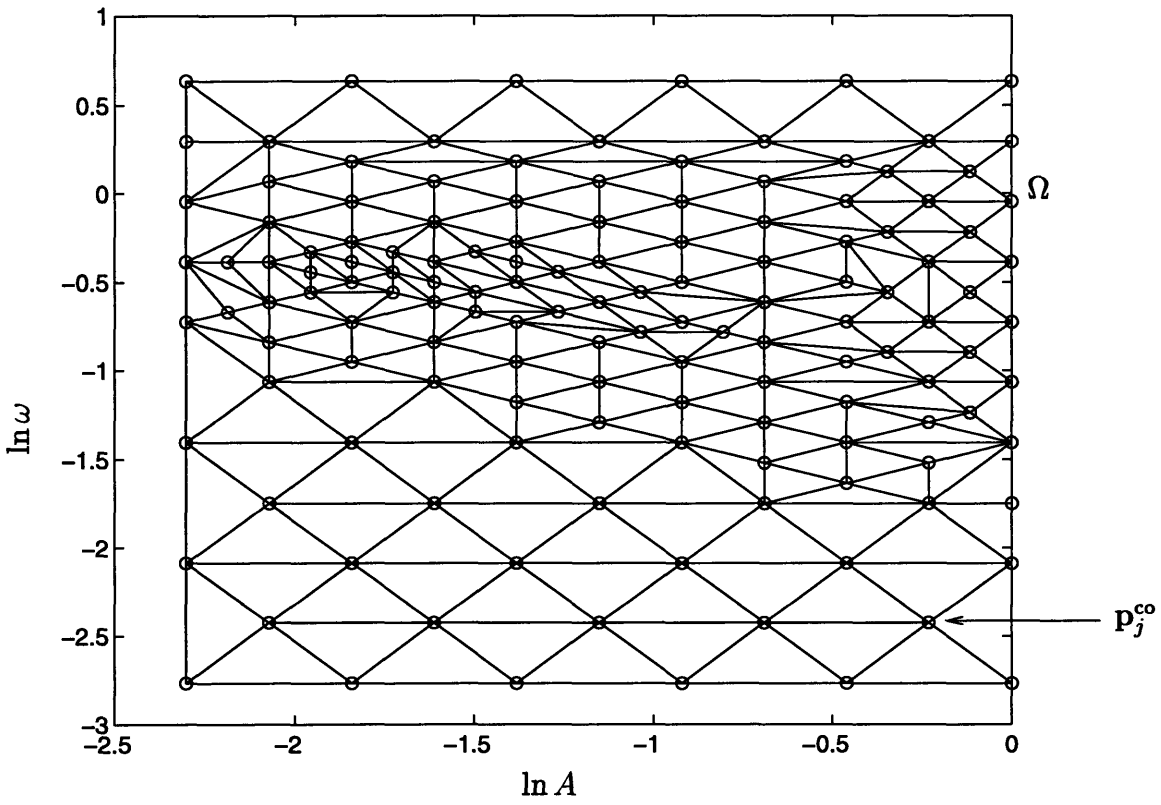


Figure 3-1: Triangulation of the input domain (design space) Ω based on $N^{co} = 138$ construction nodes $\mathbf{p}_1^{co}, ..., \mathbf{p}_N^{co}$.

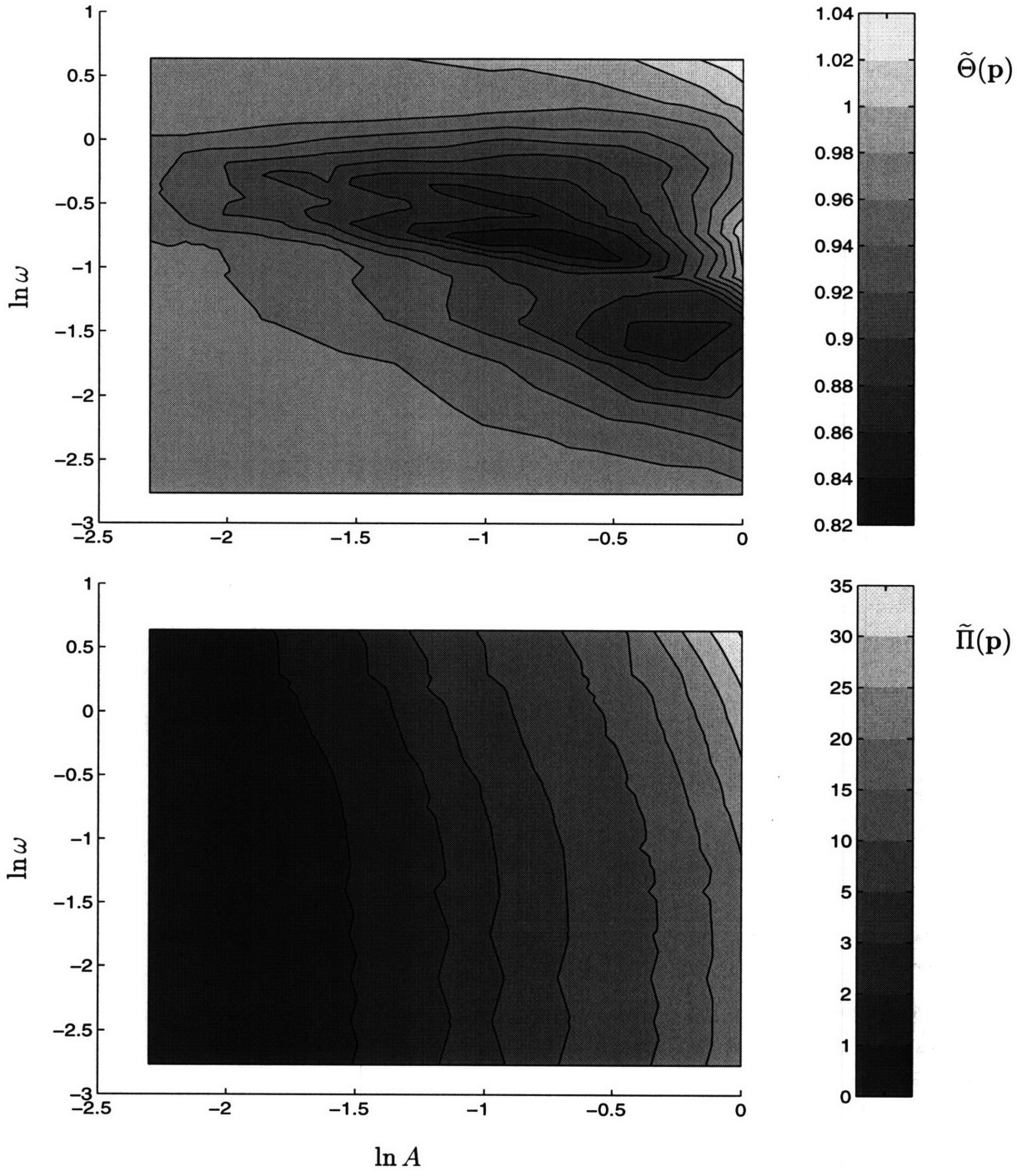


Figure 3-2: Surrogate input–outputs functions $\widetilde{\Theta}(\mathbf{p})$ (top) and $\widetilde{\Pi}(\mathbf{p})$ (bottom).

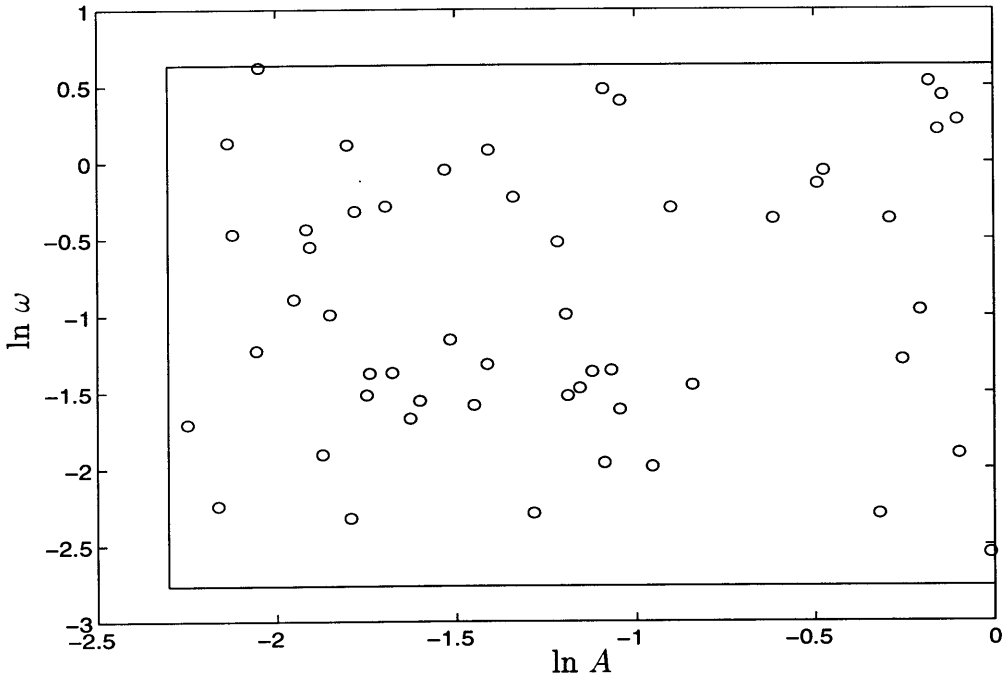


Figure 3-3: Validation sample consisting of $N = 53$ i.i.d. random input vectors $\mathbf{P}_i = (\ln A_i, \ln \omega_i), i = 1, ..., N$.

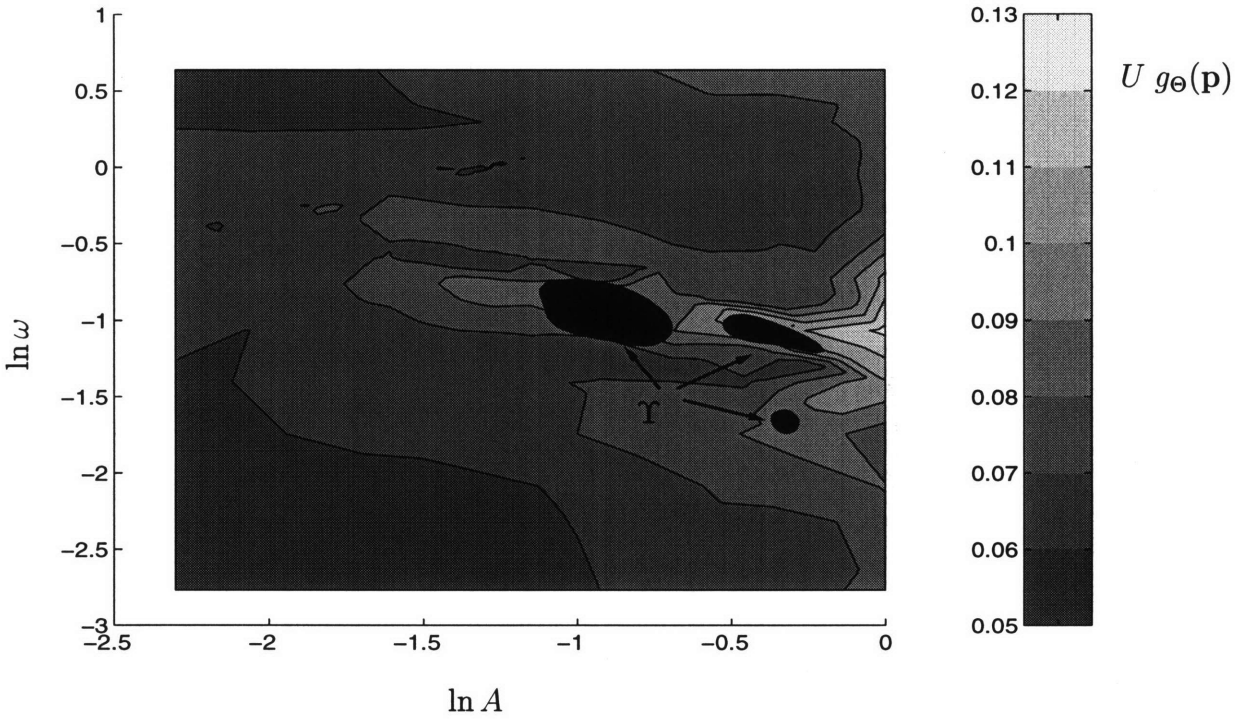


Figure 3-4: Distribution of surrogate error for the temperature, and a *possible* location of the uncharacterized region, Υ.

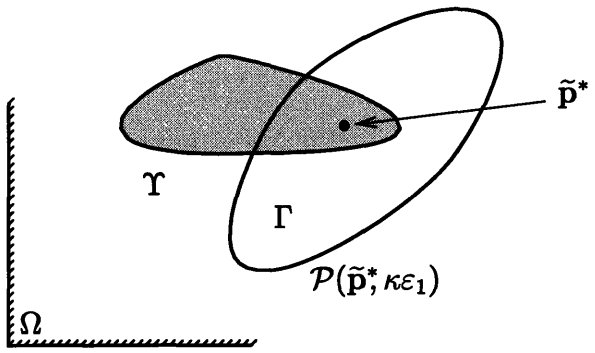


Figure 3-5: Schematic representation of a prediction neighborhood $\mathcal{P}(\tilde{\mathbf{p}}^*, z)$, $z = \kappa\varepsilon_1$, $\kappa \geq 1$. Here Υ is the uncharacterized region and $\Gamma = \mathcal{P}(\tilde{\mathbf{p}}^*, \kappa\varepsilon_1) \setminus \Upsilon$.

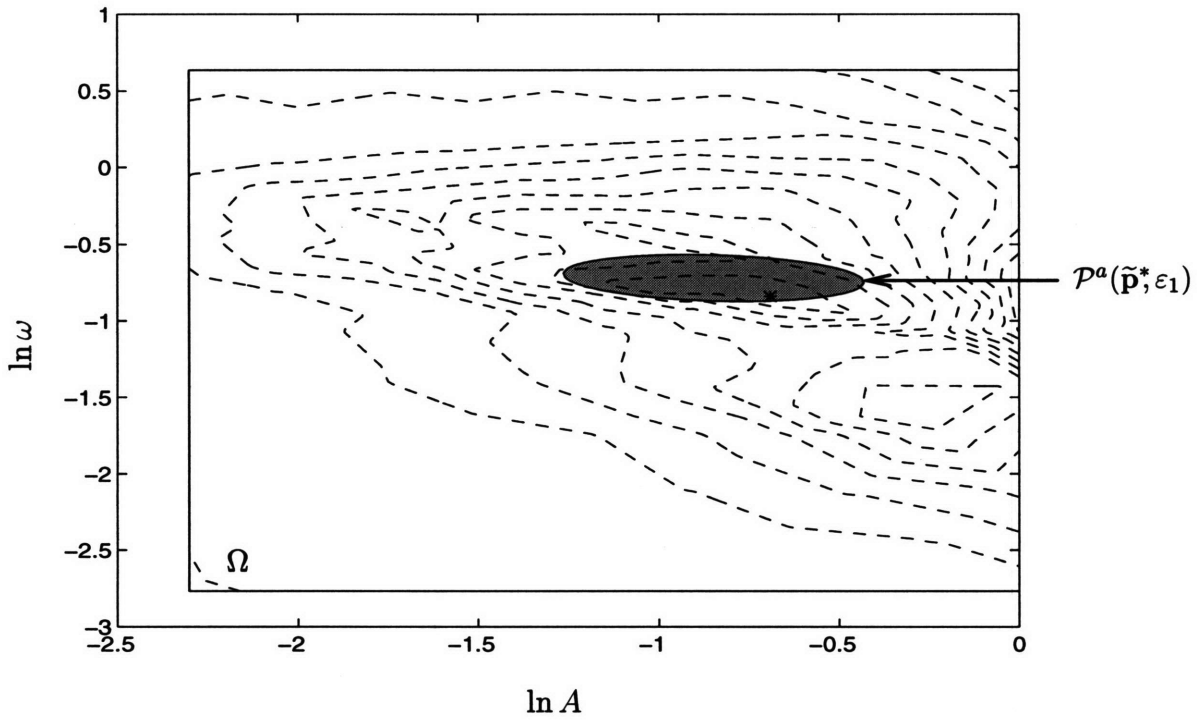


Figure 3-6: Surrogate minimizer $\tilde{\mathbf{p}}^*$ $(*)$ and prediction neighborhood $\mathcal{P}^a(\tilde{\mathbf{p}}^*, \varepsilon_1)$ (shaded). Dashed lines are contours of $\tilde{\Theta}(\mathbf{p})$.

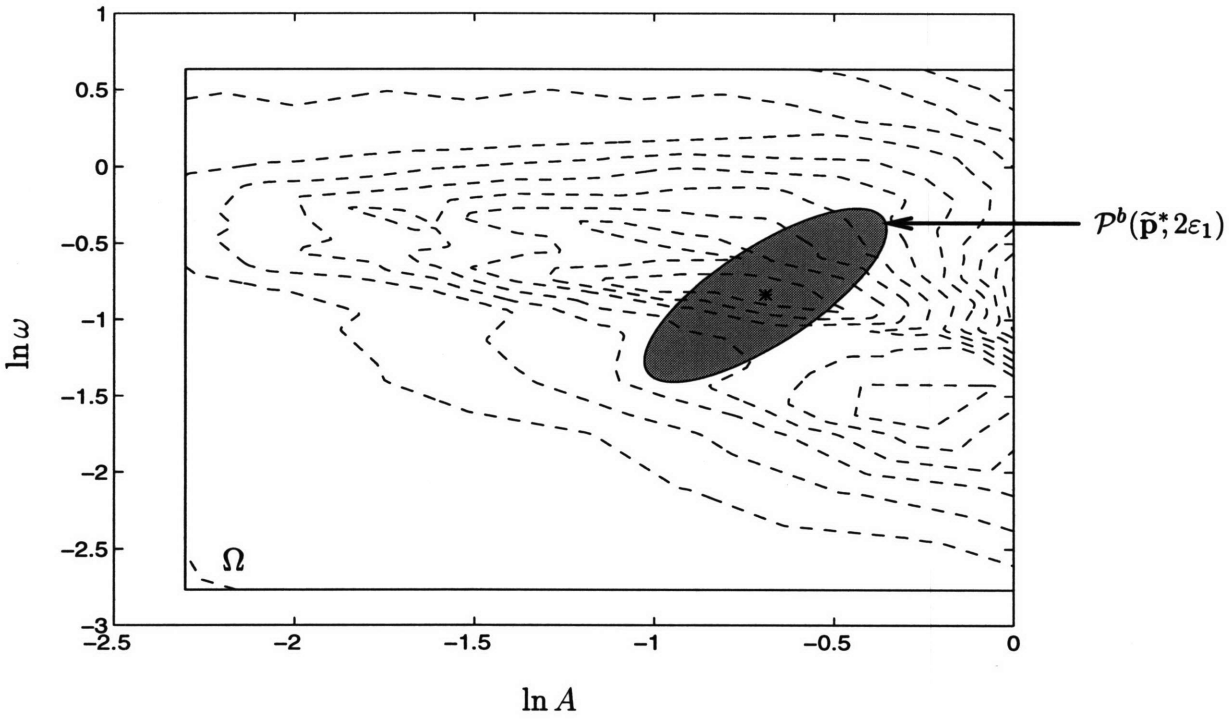


Figure 3-7: Surrogate minimizer $\tilde{\mathbf{p}}^*$ $(*)$ and prediction neighborhood $\mathcal{P}^b(\tilde{\mathbf{p}}^*, 2\varepsilon_1)$ (shaded) for the locally quasi-convex example. Dashed lines are contours of $\tilde{\Theta}(\mathbf{p})$.

Chapter 4

Surrogate Pareto Formulation

In Section 4.1 we briefly review several essential Pareto concepts, and conclude by relating these concepts to our surrogate framework. In Sections 4.2, 4.3, 4.4, 4.5 we then detail the steps of the surrogate Pareto formulation: surrogate construction; surrogate validation; surrogate–based design; and error analysis.

4.1 Introduction

The notion of Pareto Analysis was first developed as an optimality concept in the field of economics by the eponymous V. Pareto [9], who continued and expanded the equilibrium concept introduced by Edgeworth [35] in 1881. In 1906 V. Pareto defined an optimum for n consumers as a position that enjoys *maximum ophelimity*; and then stated that "any displacement in parting from that position necessarily has the effect of increasing the opfelimity that certain individuals enjoy, of being agreeable to some and disagreeable to others."

The above statement has been the basis of a class of optimal solutions defined as Pareto optimal by one of the founding fathers of quality improvement, J. Juran [36] in 1951. The concept has also a separate mathematical history that was eventually united with economics in game theory by E. Borel [37] in 1921. The history of the

subject has been traced by W. Stadler [38].

The application of Pareto optimization in engineering is a relatively recent event, first referred to in the 1970's [39, 40]. Multicriteria optimization is certainly no longer a novelty in engineering [10, 41, 42]. The widest application has been observed in resources management and structural design. In disciplines such as fluid mechanics, however, applications remain extremely rare — for reasons that will become clear below.

The fundamental premise in Pareto analysis is that one can articulate preferences on the selected outputs, or performance metrics: in our application, it is clear that a lower temperature, θ, and a lower power, ϖ, will typically represent better thermal–hydraulic design. One then introduces the notion of the *achievable set*,

$$\mathcal{A}^T = \{\mathbf{s} \in I\!R^2 \mid \exists \mathbf{p} \in \Omega \text{ s.t. } \Theta^T(\mathbf{p}) \leq s_1, \Pi^T(\mathbf{p}) \leq s_2\}. \tag{4.1}$$

where $(\ )^T$ here refers, as before, to truth. Our interest is in the boundary of $\mathcal{A}^T$ (more precisely, that part of the boundary not at infinity), $\partial\mathcal{A}^T$, which we denote the *PO (Pareto Optimal) output manifold.* This manifold is widely known in finance as the efficient frontier; it was first identified by H. Markowitz [43], the founder of portfolio optimization theory, in 1952. In engineering applications it is refered to as the functional–efficient boundary, or weight–compliance diagram, or trade–off curve [10]. The set of points in the input space, Ω, which are pre-images of points $\mathbf{s} \in \partial\mathcal{A}^T$ under the mapping $[\Theta^T(\mathbf{p}), \Pi^T(\mathbf{p})]$, is denoted the *PO input manifold,*

$$\mathcal{L}^T = \{\mathbf{p} \in \Omega \mid \exists \mathbf{s} \in \partial\mathcal{A}^T \text{ s.t. } \Theta^T(\mathbf{p}) = s_1, \Pi^T(\mathbf{p}) = s_2\}. \tag{4.2}$$

The sets $\partial\mathcal{A}^T$ and $\mathcal{L}^T$ are shown schematically in Figure 4-1. Note that although we optimistically denote $\partial\mathcal{A}^T$ and $\mathcal{L}^T$ as $K-1$ dimensional manifolds, where K is the number of outputs (in our example $K = 2$), we do not in fact know that these sets, in particular $\mathcal{L}^T$, have nice topological properties; however we contend that for many

engineering applications this will be the case.

The interest in $\partial\mathcal{A}^T$, and hence $\mathcal{L}^T$, is that all other outputs, and hence inputs, correspond to design points which can be improved with respect to our stated preferences [44]. More precisely, for an output pair (θ', ϖ') not on $\partial\mathcal{A}^T$, there exists a $\mathbf{p} \in \Omega$ in $\mathcal{L}^T$ such that $\Theta^T(\mathbf{p}) < \theta'$ and $\Pi^T(\mathbf{p}) \leq \varpi'$ or $\Theta^T(\mathbf{p}) \leq \theta'$ and $\Pi^T(\mathbf{p}) < \varpi'$. For our purposes here, the most critical point concerning $\partial\mathcal{A}^T$ and $\mathcal{L}^T$ is that, for a large class of optimization problems, $\mathbf{p}^{*T} \in \mathcal{L}^T$, and hence $(\Theta^T(\mathbf{p}^{*T}), \Pi^T(\mathbf{p}^{*T})) \in \partial\mathcal{A}^T$. In particular, for any objective function $\phi(\mathbf{p}, s_1, s_2)$ and constraint function $\psi(\mathbf{p}, s_1, s_2)$ in (2.8) and (2.9) that do not depend explicitly on the first variable ($\mathbf{p}$) and are non-decreasing functions of s_1 and s_2, a (presumed feasible) minimizer will always lie on $\mathcal{L}^T$. (In fact, the class of objective functions and constraints can be enlarged even further depending on the structure of $\mathcal{L}^T$.) Both our design problems of Section 2.3.1 satisfy these conditions, as does $\phi(\mathbf{p}, s_1, s_2) = \alpha s_1 + \beta s_2$, in which α and β represent trade–off coefficients.

It should be clear from the above that Pareto analysis will often be an ideal framework for engineering analysis, in which a wide class of objective functions may be readily considered. It is typically the case that (say) trade–off coefficients are not known *a priori*, because the sensitivity of the problem is not yet understood; Pareto analysis accommodates this uncertainty, and provides a flexible environment in which to both articulate and realize design objectives.

The reason, or a reason, Pareto analysis is not more widespread is the difficulty in determining $\partial\mathcal{A}^T$ and $\mathcal{L}^T$. To construct $\partial\mathcal{A}^T$, we must reduce the vector multi-criteria minimization problem to a series of scalar problems through a *scalarization* process. Scalarization, also known as *parameterization*, can be carried out in several ways: the most widely used methods are objective weighting, distance functions, constrained oriented transformation (trade–off method), and the min–max formulation [10]. We choose here the min–max formulation (the one most frequently used), as this is the only technique that can treat non-convexity; we have no reason to suspect,

and certainly cannot prove, that $\mathcal{A}^T$ will be convex for our problem. Variations of the min–max formulation and other scalarization methods can be found in [45].

The min–max method is based on the property that, at any point s in $\partial\mathcal{A}^T$, all points s' such that $s'_1 < s_1, s'_2 < s_2, ..., s'_K < s_K$ are *not* in $\mathcal{A}^T$ (if s' was in $\mathcal{A}^T$, then s would not be Pareto optimal and hence not in $\partial\mathcal{A}^T$). To find the point s for any direction in $\mathcal{A}^T$, it then suffices to find the vector which has at least one component s_i ,$i = 1, ..., K$, (say the maximum), that cannot be further minimized; in other words s is the point closer to the origin in a particular vector direction. We then construct $\partial\mathcal{A}^T$, by finding s for all directions. To do that, we first define a scalarization parameter w, and scalarization space $\mathcal{W}^T = [0, 1]$. We then introduce the function $V_w^T(\mathbf{p}) = v_w(\Theta^T(\mathbf{p}), \Pi^T(\mathbf{p}))$, where $v_w(s_1, s_2) = \max(ws_1, (1-w)s_2)$. The solution to the min–max problem for a particular w is a *PO input*, defined as

$$\xi^T(w) = \arg \min_{\{\mathbf{p}\in\Omega\}} V_w^T(\mathbf{p}) = \arg \min_{\{\mathbf{p}\in\Omega\}} \max(w\Theta^T(\mathbf{p}), (1-w)\Pi^T(\mathbf{p})). \tag{4.3}$$

The solution is illustrated graphically in Figure 4-2. The corresponding *PO* output pair is then given by

$$(\Theta^T(\xi^T(w)), \Pi^T(\xi(w))) = \min_{\{(\theta,\varpi)\in\mathcal{A}^T\}} \max(w\theta, (1-w)\varpi). \tag{4.4}$$

The *PO* sets are then defined as $\mathcal{L}^T = \xi^T(\mathcal{W}^T)$ and $\partial\mathcal{A}^T = (\Theta^T(\mathcal{L}^T), \Pi^T(\mathcal{L}^T))$, which represent the solution of many (in theory, infinite) min–max problems — a nontrivial task given the expense of $\Theta^T(\mathbf{p}), \Pi^T(\mathbf{p})$.

The method is completed by executing three final steps. First, we have been somewhat imprecise in our definition of the *PO* output manifold. In fact, "vertical" and "horizontal" segments of $\partial\mathcal{A}^T$ are *not* part of the *PO* output set; we must redefine $\partial\mathcal{A}^T$ as $\partial\mathcal{A}^T$ with these segments removed. Second, and relatedly, these excised vertical and horizontal segments correspond to intervals of w in $[0, 1]$ for which (4.3)–(4.4) does not have a unique solution; we redefine $\mathcal{W}^T$ as $[0, 1]$ with these intervals

removed. Third, and finally, if after the first two steps there still exist w for which $\xi^T(w)$ is nonunique, we force uniqueness by selecting that solution of the min–max problem (4.3) which is (say) smallest in the Euclidean norm. A consistent choice of $\xi^T(w)$ is important in obtaining a smooth $\mathcal{L}^T$.

It is clear from the above analysis that the Pareto input manifold $\mathcal{L}^T$ is precisely the (or a) "special region" of Ω described in the Introduction on which we can focus attention with some surety that we have, indeed, identified the design–relevant subset of Ω. For problems originally posed as multicriteria problems, or objectives and constraints in (2.8)–(2.9) that satisfy the non-decreasing conditions described above, there is no loss in performance; if these conditions are not satisfied, the results will, admittedly, be sub-optimal. In any event, it can be easily inferred that Pareto analysis offers the surrogate approach a dimension reduction from $I\!R^M$ (Ω) to $I\!R^{K-1}$ ($\mathcal{L}^T$), (in our example from $I\!R^2$ to $I\!R^1$); the benefit of this reduction will be improved localization and predictability. Conversely, the benefit of the surrogate approach to Pareto analysis is the reduction of the scalarization problem to tractable form.

4.2 Surrogate Construction

Surrogate construction in the Pareto context involves two steps. The first step is construction of a surrogate over all of Ω, as in the basic formulation; it is essential to note that this construction is based on the high–fidelity model, not the truth model, and thus the need for increasing numbers of construction points with increasing input dimension will be less debilitating. For our purpose here, we simply import the surrogates $\tilde{\Theta}(\mathbf{p}), \tilde{\Pi}(\mathbf{p})$ of Section 3.1.

The second step is the construction of the surrogate approximation to the truth *PO* input and output manifolds. To this end, we define the surrogate achievable set $\tilde{\mathcal{A}}$ as

$$\tilde{\mathcal{A}} = \{\mathbf{s} \in I\!R^2 \mid \exists \mathbf{p} \in \Omega \text{ s.t. } \tilde{\Theta}(\mathbf{p}) \leq s_1, \tilde{\Pi}(\mathbf{p}) \leq s_2\}. \tag{4.5}$$

Our interest is, of course, in isolating the boundary of $\tilde{\mathcal{A}}$, $\partial\tilde{\mathcal{A}}$ (appropriately sanitized as described in the previous section), and the pre-image of $\tilde{\mathcal{A}}$ in Ω, $\tilde{\mathcal{L}}$. We thus introduce the scalarization space $\widetilde{\mathcal{W}} \subset [0,1]$, and the min–max problem

$$\tilde{\xi}(w) = \arg \min_{\{\mathbf{p}\in\Omega\}} \max(w\tilde{\Theta}(\mathbf{p}), (1-w)\tilde{\Pi}(\mathbf{p})), \tag{4.6}$$

which represents a point in $\tilde{\mathcal{L}}$. The corresponding *PO* output pair is then given by

$$(\tilde{\Theta}(\tilde{\xi}(w)), \tilde{\Pi}(\tilde{\xi}(w))) = \min_{\{(\theta,\varpi)\in\tilde{\mathcal{A}}\}} \max(w\theta, (1-w)\varpi). \tag{4.7}$$

The solution for all $w \in \widetilde{\mathcal{W}}$ gives $\tilde{\mathcal{L}} = \tilde{\xi}(\widetilde{\mathcal{W}})$ and $\partial\tilde{\mathcal{A}} = (\tilde{\Theta}(\tilde{\mathcal{L}}), \tilde{\Pi}(\tilde{\mathcal{L}}))$. Note that $\widetilde{\mathcal{W}}$ is not known *a priori*, but is determined as part of the solution procedure. The elimination of redundant intervals of w in $[0,1]$ is intended to economize on subsequent expensive truth validation computations, not on the inexpensive surrogate min–max problems (4.6)–(4.7).

In the particular example of this thesis, $\tilde{\Theta}(\mathbf{p})$ and $\tilde{\Pi}(\mathbf{p})$ are sufficiently inexpensive that we may find $\partial\tilde{\mathcal{A}}$ and $\tilde{\mathcal{L}}$ by direct search. To wit, we construct a uniform 300×300 grid in the input space Ω, and evaluate $\tilde{\Theta}(\mathbf{p})$ and $\tilde{\Pi}(\mathbf{p})$ at each point. For each $w = k/1000, k = 0, 1, ..., 1000$, we then compute $\max(w\tilde{\Theta}(\mathbf{p}), (1-w)\tilde{\Pi}(\mathbf{p}))$ over the input grid, and find the minimum by simple search. The resulting *PO* input and output manifolds are shown in Figure 4-3; note $\tilde{\mathcal{A}}$ is, indeed, non-convex.

4.3 Surrogate Validation

As for surrogate construction, surrogate validation also now entails two steps. The first validation step, related to optimality, is over the entire input domain Ω with respect to a prescribed importance and scaling functions, $\rho(\mathbf{p}) : \Omega \to I\!R$ and $g_\Theta(\mathbf{p})$, $g_\Pi(\mathbf{p})$, respectively, and a specified number of validation points, N. Although this validation is over the entire input space, the results will serve only in "volumetric"

optimality analyses akin to that of Section 3.4.3; localization and predictability are not, therefore, an issue, and N need not grow exponentially with M. (Note, in particular, that ε_1 and ε_2 are not required, but only N, the sample size.)

For our particular eddy–promoter example, we simply take $\rho(\mathbf{p})$ uniform, $g_\Theta(\mathbf{p})$, $g_\Pi(\mathbf{p})$ as given in Section 3.2, and $N = 53$. This permits us to recycle the validation–sample realization of Section 3.2 and Figure 3-3; we thus obtain $U = 0.53$ for our global prediction error.

The second validation step is related to predictability. Here, we define a second importance function $\rho'(\mathbf{p}) : \Omega \to I\!R$ which is only nonzero over a subdomain $\Omega' \subset \Omega$. We then introduce the associated normalized measure for any domain $\mathcal{D} \subset \Omega'$, $\mu'(\mathcal{D} \subset \Omega') = \int_{\mathcal{D}} \rho'(\mathbf{p})\mathbf{dp}$. Note that now $\mu'(\Omega') = 1$. In general, $\rho'(\mathbf{p})$ is concentrated about $\widetilde{\mathcal{L}}$, that is, Ω' is a tube or sheath that surrounds $\widetilde{\mathcal{L}}$. A convenient way to define $\rho'(\mathbf{p})$ is to introduce a random scalarization parameter $W \in \widetilde{\mathcal{W}}$ with density $f_W(w)$, a random vector V distributed uniformly over the unit disk, and a radius parameter r' (independent of the validation parameters). The importance function $\rho'(\mathbf{p})$ is then defined as the probability density of the random input vector $\mathbf{P}'$ given by

$$\mathbf{P}' = \widetilde{\xi}(W) + Vr'. \tag{4.8}$$

The corresponding input space Ω' is then given by

$$\Omega' = \{\mathbf{p} \in \Omega \mid \|\mathbf{p} - \widetilde{\xi}(w)\| < r', \forall w \in \widetilde{\mathcal{W}}\}, \tag{4.9}$$

where $\|\cdot\|$ is the (here) Euclidean norm. We depict Ω' in Figure 4-4 for $r' = 0.5$. Once $\rho'(\mathbf{p})$ is chosen, the remainder of the validation procedure follows according to Section 3.2.

We illustrate the validation procedure for our particular example. We set $f_W(w)$ such that $\ln(\widetilde{\Pi}(\mathbf{p}))$ is uniform, in order to provide adequate resolution over the wide range of powers. The $f_W(w)$ is simple to sample if we enumerate ϖ on the surrogate

PO output manifold, $\partial\widetilde{\mathcal{A}}$, as a function of $w = k/1000, k = 0, 1, ..., 1000$. We then readily find W by linearly interpolating those $k/1000$ that bracket the uniformly chosen sample values of $\ln(\widetilde{\Pi}(\mathbf{p}))$ over the range from $\ln(\varpi_{\min})$ to $\ln(\varpi_{\max})$. The positive scaling functions $g'_\Theta(\mathbf{p})$ and $g'_\Pi(\mathbf{p})$ are taken to be $g_\Theta(\mathbf{p})$ and $g_\Pi(\mathbf{p})$ from the basic formulation, now restricted to Ω'. We set the parameters ε'_1 and ε'_2 to 0.06 and 0.26, respectively, which therefore requires, from (3.4), a validation sample size $N' = \lceil \ln \varepsilon'_2 / \ln(1 - \varepsilon'_1) \rceil = 22$. We then draw $N' = 22$ i.i.d. input vectors $\mathbf{P}'_i = (\ln \mathcal{A}'_i, \ln \omega'_i)$, $i = 1, ..., N'$ according to the density $\rho'(\mathbf{p})$; our particular realization is shown in Figure 4-4. Finally, we perform $N' = 22$ expensive truth calculations, and compute the model prediction error

$$U' = \max_{i \in \{1,...,N'\}} \left(\max \left(\frac{|\Theta^T(\mathbf{P}'_i) - \widetilde{\Theta}(\mathbf{P}'_i)|}{g'_\Theta(\mathbf{P}'_i)}, \frac{|\Pi^T(\mathbf{P}'_i) - \widetilde{\Pi}(\mathbf{P}'_i)|}{g'_\Pi(\mathbf{P}'_i)} \right) \right) = 0.32. \qquad (4.10)$$

(See Appendix A for the validation data.) The local model prediction error, $U' = 0.32$, is less than the global model prediction error, $U = 0.53$; this is expected, since $\Omega' \subset \Omega$, and $N' < N$. Note that $N' < N$ since we can now afford a larger ε'_1; we could also keep $N' \sim N$ and $\varepsilon'_1 = 0.06$, but reduce ε'_2 to 0.037 to obtain greater significance.

4.4 Surrogate–Based Design

For our design problem we now choose (2.12)–(2.13). In particular we look for the lowest possible power requirements for given temperatures $\bar{\theta}^q, q \in \{1, ..., Q\} \equiv \mathcal{Q}$. The Q different temperature requirements are intended to reflect the changing specifications typical of most design studies. In the surrogate–based design procedure, we substitute $\widetilde{\Theta}(\mathbf{p})$ for $\Theta(\mathbf{p})$ and $\widetilde{\Pi}(\mathbf{p})$ for $\Pi(\mathbf{p})$ to find

$$\widetilde{\varpi}^q_{\min} = \min_{\{\mathbf{p} \in \Omega | \widetilde{\Theta}(\mathbf{p}) \le \bar{\theta}^q\}} \widetilde{\Pi}(\mathbf{p}), \quad \forall q \in \mathcal{Q}, \qquad (4.11)$$

and corresponding $(\ln A, \ln \omega)$ design points,

$$\tilde{\mathbf{p}}^{*q} = \arg \min_{\{\mathbf{p} \in \Omega | \tilde{\Theta}(\mathbf{p}) \leq \bar{\theta}^q\}} \tilde{\Pi}(\mathbf{p}), \quad \forall q \in \mathcal{Q}. \tag{4.12}$$

By the arguments of Section 4.1, $\tilde{\mathbf{p}}^{*q} \in \tilde{\mathcal{L}} \subset \Omega'$, and $(\bar{\theta}^q, \varpi^q_{\min}) \in \partial\tilde{\mathcal{A}}$, $\forall q \in \mathcal{Q}$. (In order to assure a feasible point for each specified temperature, we assume that $\bar{\theta}^q \geq \tilde{\theta}_{\min}$.)

In our particular example here, we take $Q = 2$, with $\bar{\theta}^1 = 0.8245$ and $\bar{\theta}^2 = 0.93$. The solutions are found to be $\widetilde{\varpi}^1_{\min} = 3.74, \widetilde{\varpi}^2_{\min} = 0.33$, and $\tilde{\mathbf{p}}^{*1} = (-0.89, -0.78), \tilde{\mathbf{p}}^{*2} = (-2.15, -0.53)$, as shown in Figure 4-6. Note that these solutions can also be readily found if we enumerate θ and ϖ on the trade–off curve as a function of $w = k/1000, k = 0, 1, ..., 1000$. In particular, we first find those $k/1000$ which bracket $\bar{\theta}$, and then perform (say) linear interpolation to obtain $\tilde{\mathbf{p}}^*$ and $\widetilde{\varpi}^q_{\min}$.

4.5 Error Analysis

4.5.1 Validation Statement

The optimality validation will be discussed in Section 4.5.3; we pass directly here to the predictability–related validation statement. In complete analogy with the earlier analysis of Section 3.4.1 we define the uncharacterized region Υ' as

$$\Upsilon' = \left\{ \mathbf{p} \in \Omega' \left| \max \left(\frac{|\Theta^T(\mathbf{p}) - \tilde{\Theta}(\mathbf{p})|}{g'_\Theta(\mathbf{p})}, \frac{|\Pi^T(\mathbf{p}) - \tilde{\Pi}(\mathbf{p})|}{g'_\Pi(\mathbf{p})} \right) > U' \right. \right\}. \tag{4.13}$$

It can then be shown that

$$Pr\{\mu'(\Upsilon') < \varepsilon'_1\} \geq 1 - \varepsilon'_2, \tag{4.14}$$

and

$$\begin{cases} |\Theta^T(\mathbf{p}) - \tilde{\Theta}(\mathbf{p})| & \leq & U' g'_\Theta(\mathbf{p}) \\ |\Pi^T(\mathbf{p}) - \tilde{\Pi}(\mathbf{p})| & \leq & U' g'_\Pi(\mathbf{p}) \end{cases}, \quad \forall \mathbf{p} \in \Omega' \setminus \Upsilon'. \tag{4.15}$$

As before, the difficulty with (4.15) is that we do not know the location of Υ'.

For our particular example the validation statement reads: With probability greater than $1 - \varepsilon'_2 = 0.74$, the surrogate error is given by

$$\begin{cases} |\Theta^T(\mathbf{p}) - \tilde{\Theta}(\mathbf{p})| & \leq & 0.32\, g'_\Theta(\mathbf{p}) \\ |\Pi^T(\mathbf{p}) - \tilde{\Pi}(\mathbf{p})| & \leq & 0.32\, g'_\Pi(\mathbf{p}) \end{cases}, \tag{4.16}$$

for any input $\mathbf{p}$ in a region $\Omega' \setminus \Upsilon'$ which is *at least* of ρ'–measure $1 - \varepsilon'_1 = 0.94$ (94% of Ω'). Figure 4-5 shows contours of surrogate error $|\Theta^T(\mathbf{p}) - \tilde{\Theta}(\mathbf{p})|$, and an illustrative region of measure $\varepsilon'_1 = 0.006$ that may constitute Υ'. We recall that contours of $|\Theta^T(\mathbf{p}) - \tilde{\Theta}(\mathbf{p})|$ are, in fact, contours of our scaling function $g'_\Theta(\mathbf{p})$.

4.5.2 Predictability Analysis

We now form the predictability statement, analogous to the predictability statement of Section 3.4.2. As before, we first introduce prediction neighborhoods $\mathcal{P}'^q(\tilde{\mathbf{p}}^{*q}, \kappa\varepsilon'_1)$, $\kappa \geq 1$, $\forall q \in \mathcal{Q}$. We can then state that, with probability greater than $1 - \varepsilon'_2$, $\forall q \in \mathcal{Q}$ there exists a region $\Gamma'^q = \mathcal{P}'^q(\tilde{\mathbf{p}}^{*q}, \kappa\varepsilon'_1) \setminus \Upsilon'$ of positive measure, $\mu'(\Gamma'^q) > (\kappa - 1)\varepsilon'_1$, such that

$$\begin{cases} |\Theta^T(\mathbf{p}) - \bar{\theta}^q| & \leq & B'^q_\theta \\ |\Pi^T(\mathbf{p}) - \widetilde{\varpi}^q_{\min}| & \leq & B'^q_\varpi, \end{cases}, \quad \forall \mathbf{p} \in \Gamma'^q, \tag{4.17}$$

where B'^q_θ and B'^q_ϖ are given by

$$B'^q_\theta = \max_{\{\mathbf{p} \in \mathcal{P}'^q(\tilde{\mathbf{p}}^{*q}, \varepsilon'_1)\}} (U' g'_\Theta(\mathbf{p}) + |\tilde{\Theta}(\mathbf{p}) - \bar{\theta}^q|), \tag{4.18}$$

and

$$B'^q_\varpi = \max_{\{\mathbf{p} \in \mathcal{P}'^q(\tilde{\mathbf{p}}^{*q}, \varepsilon'_1)\}} (U' g'_\Pi(\mathbf{p}) + |\tilde{\Pi}(\mathbf{p}) - \widetilde{\varpi}^q_{\min}|), \tag{4.19}$$

respectively. Note that in multipoint studies, the measure of of the region Γ'^q does not decrease with q — the events of the predictability statement are not independent and hence the bounds are jointly valid for all q. Note also that for simplicity we choose here to pursue symmetric bounds, rather than the sharper non-symmetric bounds of Section 3.4.2.

We now turn to our particular calculation. For each prediction neighborhood $\mathcal{P}'^q(\tilde{\mathbf{p}}^{*q}, \varepsilon'_1)$ we choose that region $\mathcal{R}'$ of measure ε'_1 that minimizes $\max_{\{\mathbf{p}' \in \mathcal{R}'\}} \delta'^q(\mathbf{p}')$, where

$$\delta'^q(\mathbf{p}') = \max\left(\frac{(U' g'_\Theta(\mathbf{p}') + |\tilde{\Theta}(\mathbf{p}') - \bar{\theta}^q|)}{\ell_\theta}, \frac{(U' g'_\Pi(\mathbf{p}') + |\tilde{\Pi}(\mathbf{p}') - \widetilde{\varpi}^q_{\min}|)}{\ell_\varpi} \right). \quad (4.20)$$

We choose for the temperature and power scalings $\ell_\theta = 0.2$ and $\ell_\varpi = 3$, respectively, in order to provide roughly the same relative predictability in both outputs. We emphasize that our estimates are joint over any choice of ℓ_θ, ℓ_ϖ, and we can thus interactively determine an appropriate distance function. The $\mathcal{P}'^q(\tilde{\mathbf{p}}^{*q}, \varepsilon'_1)$ are readily found by the Monte–Carlo method: $\delta'(\cdot)$ is first evaluated at N_{MC} points $\mathbf{P}'$ distributed according to $\rho'(\mathbf{p})$; these points are sorted in order of increasing $\delta'(\cdot)$; the first $\varepsilon_1 N_{MC}$ points in this sorted list then "define" the prediction neighborhood. Here, $N_{MC} = 10,000$; the $\varepsilon_1 N_{MC} = 600$ points around each minimizer with the smallest $\delta'(\cdot)$ constitute the prediction neighborhoods, $\mathcal{P}'^q(\tilde{\mathbf{p}}^{*q}, \varepsilon'_1)$ shown in Figure 4-6.

The associated predictability statement is then: With probability greater than $1 - \varepsilon'_2 = 0.74$, there exists a region $\Gamma'^{1[2]} \subset \mathcal{P}'^{1[2]}(\tilde{\mathbf{p}}^{*1[2]}, \varepsilon'_1)$ of positive measure, $\mu'(\Gamma'^{1[2]}) > 0$, such that

$$\begin{cases} |\Theta^T(\mathbf{p}) - \bar{\theta}^{1[2]}| & \leq & B'^{1[2]}_\theta & = & 0.054\,[0.044] \\ |\Pi^T(\mathbf{p}) - \tilde{\Pi}(\tilde{\mathbf{p}}^{*1[2]})| & \leq & B'^{1[2]}_\varpi & = & 0.81\,[0.21], \end{cases} \quad (4.21)$$

where [ ] in (4.21) refers to the case $q = 2$. These results are illustrated graphically in Figure 4-6.

We see that the predictability gaps do not overlap in θ or ϖ, which indicates that (with probability $\geq 1-\varepsilon_2' = 0.74$) there exist design points which yield distinguishable performance. The predictability is much better than for the basic formulation; note that the $q = 2$ study is, in effect, equivalent to the design problem studied in Section 3.3. However, the predictability in the temperature is still not particularly good, the reasons for which will be discussed in Chapter 5. As before, we also obtain notions of genericity and stability, the latter with respect to *all* directions about $\tilde{\mathbf{p}}^*$; the particular choice of r' must be informed by design considerations. Finally, design localization is, of course, much improved over the basic formulation (see Figure 4-6), even though $\varepsilon_1' > \varepsilon_1$, due to the dimension reduction intrinsic to the Pareto formulation.

In summary, the Pareto analysis tells us that we may not be able to achieve *actual* (true) performance of precisely $\widetilde{\varpi}^1_{\min}$, $\bar{\theta}^1$ (respectively, $\widetilde{\varpi}^2_{\min}$, $\bar{\theta}^2$) at precisely $\tilde{\mathbf{p}}^{*1}$ (respectively, $\tilde{\mathbf{p}}^{*2}$), but that there are regions of nonzero measure near $\tilde{\mathbf{p}}^{*1}$ (respectively, $\tilde{\mathbf{p}}^{*2}$), points $\mathbf{p}'$ of which will yield $(\Theta^T(\mathbf{p}'), \Pi^T(\mathbf{p}'))$ pairs *close* to $\widetilde{\varpi}^1_{\min}$, $\bar{\theta}^1$ (respectively, $\widetilde{\varpi}^2_{\min}$, $\bar{\theta}^2$). The question remains, however, as to how close $\tilde{\mathcal{L}}$ and $\partial\tilde{\mathcal{A}}$ are to $\mathcal{L}^T$ and $\partial\mathcal{A}^T$.

4.5.3 Optimality Analysis

Although we cannot answer the question of how close $\partial\tilde{\mathcal{A}}$ is to $\partial\mathcal{A}^T$, we can, in a fashion analogous to that of Section 3.4.3, indicate how much computational effort is required to find a design point $\mathbf{p}$ which improves upon $\partial\tilde{\mathcal{A}}$, or more precisely, improves upon an expanded achievable set $\mathcal{A}_o$ which differs from $\tilde{\mathcal{A}}$ by the order of the model prediction error. To wit, we define a lower–bound achievable set as

$$\mathcal{A}_o = \{\mathbf{s} \in I\!R^2 \mid \exists \mathbf{p} \in \Omega \text{ s.t. } \tilde{\Theta}(\mathbf{p}) - Ug_\Theta(\mathbf{p}) \leq s_1, \tilde{\Pi}(\mathbf{p}) - Ug_\Pi(\mathbf{p}) \leq s_2\}, \quad (4.22)$$

the boundary of which is denoted $\partial\mathcal{A}_o$; note that $\tilde{\mathcal{A}} \subset \mathcal{A}_o$. It follows that if a point $\mathbf{p}$ is found such that $(\theta = \Theta^T(\mathbf{p}),\ \varpi = \Pi^T(\mathbf{p})) \notin \mathcal{A}_o$, then $\mathbf{p}$ must lie in Υ' (from which it also follows that if $\mu'(\Upsilon') = 0$, then $\mathcal{A}_o$ includes $\mathcal{A}^T$). To show this, we note that if $(\theta, \varpi) \notin \mathcal{A}_o$, there exists *no* point $\mathbf{p}'$ for which $\tilde{\Theta}(\mathbf{p}') - Ug_\Theta(\mathbf{p}') \leq \theta$, $\tilde{\Pi}(\mathbf{p}') - Ug_\Pi(\mathbf{p}') \leq \varpi$, and particularly at $\mathbf{p}' = \mathbf{p}$ we must have

$$\tilde{\Theta}(\mathbf{p}) - Ug_\Theta(\mathbf{p}) > \Theta^T(\mathbf{p}) \tag{4.23}$$

or

$$\tilde{\Pi}(\mathbf{p}) - Ug_\Pi(\mathbf{p}) > \Pi^T(\mathbf{p}). \tag{4.24}$$

But (4.23)–(4.24) implies $\mathbf{p} \in \Upsilon'$, as desired.

The results of Section 3.4.3 can thus be directly imported. In particular, if we set $\varepsilon_L \in [0, 1]$ and m according to (3.32), then the probability of drawing a sequence of i.i.d. random input vectors $\hat{\mathbf{P}}_1, ..., \hat{\mathbf{P}}_m$ according to $\rho(\mathbf{p})$ and finding a point $\hat{\mathbf{P}}_j$ such that $(\Theta^T(\hat{\mathbf{P}}_j),\ \Pi^T(\hat{\mathbf{P}}_j)) \notin \mathcal{A}_o$ is less than $1 - \varepsilon_L$; conversely, with probability greater than ε_L, we will at best find an improvement to $\partial\tilde{\mathcal{A}}$ which lies in $\mathcal{A}_o \setminus \tilde{\mathcal{A}}$. The conclusions which may be drawn are summarized in Section 3.4.3.

We show $\partial\mathcal{A}_o$ for our particular example in Figure 4-7. If we set (say) $\varepsilon_L = 0.5$, then $m = 53$, implying that after $m = 53$ additional (post-validation) appeals to the truth model there is a greater than 0.5 probability that we will have improved upon $\partial\tilde{\mathcal{A}}$ by no more than $\mathcal{A}_o \setminus \tilde{\mathcal{A}}$. However, as in our example the potential reductions $\mathcal{A}_o \setminus \tilde{\mathcal{A}}$ are considerable, we conclude that our models are not yet sufficiently accurate; nontrivial improvement requires the sequential or adaptive procedures described in [28, 29]. When $\mathcal{A}_o \setminus \tilde{\mathcal{A}}$ *is* sufficiently small, and we additionally hypothesize that $\tilde{\Theta}(\mathbf{p}), \tilde{\Pi}(\mathbf{p})$ captures the smooth parts of $\Theta^T(\mathbf{p}), \Pi^T(\mathbf{p})$, any further *global* improvement would be difficult, and thus presumably ill–advised unless system performance is critical.

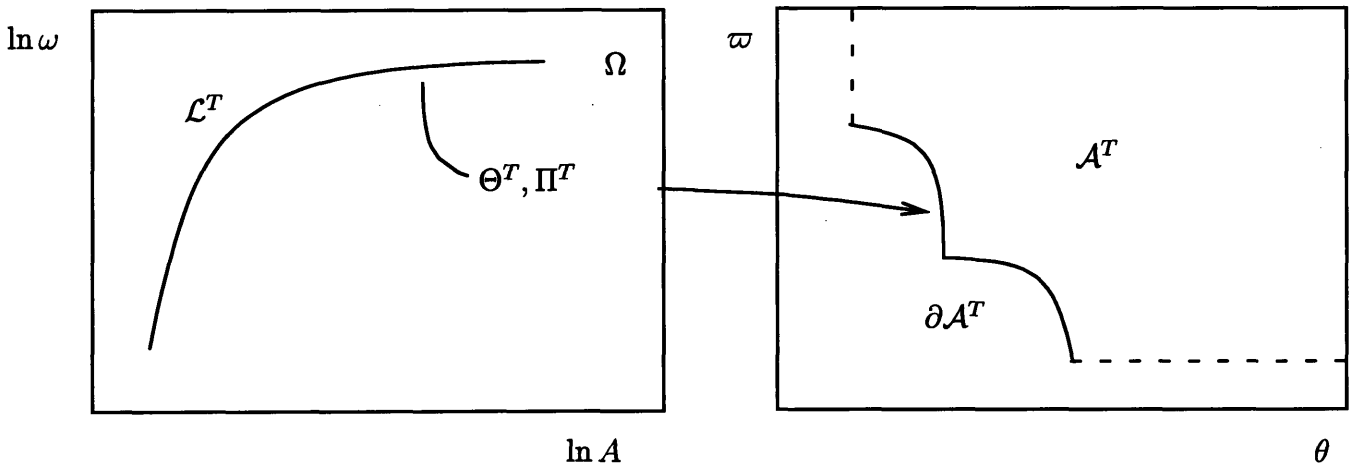


Figure 4-1: Schematic representation of a PO input manifold $\mathcal{L}^T$ (left) and output manifold $\partial\mathcal{A}^T$ (right).

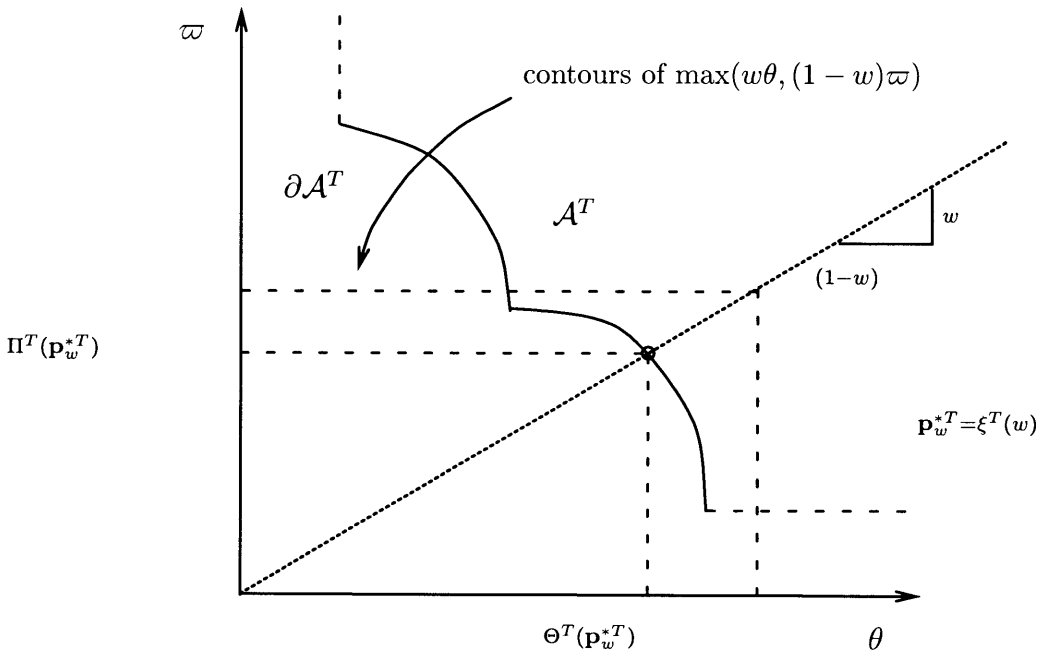


Figure 4-2: For a particular w, the solution to the min–max problem is given by the minimum contour of $\max(w\theta, (1-w)\varpi)$.

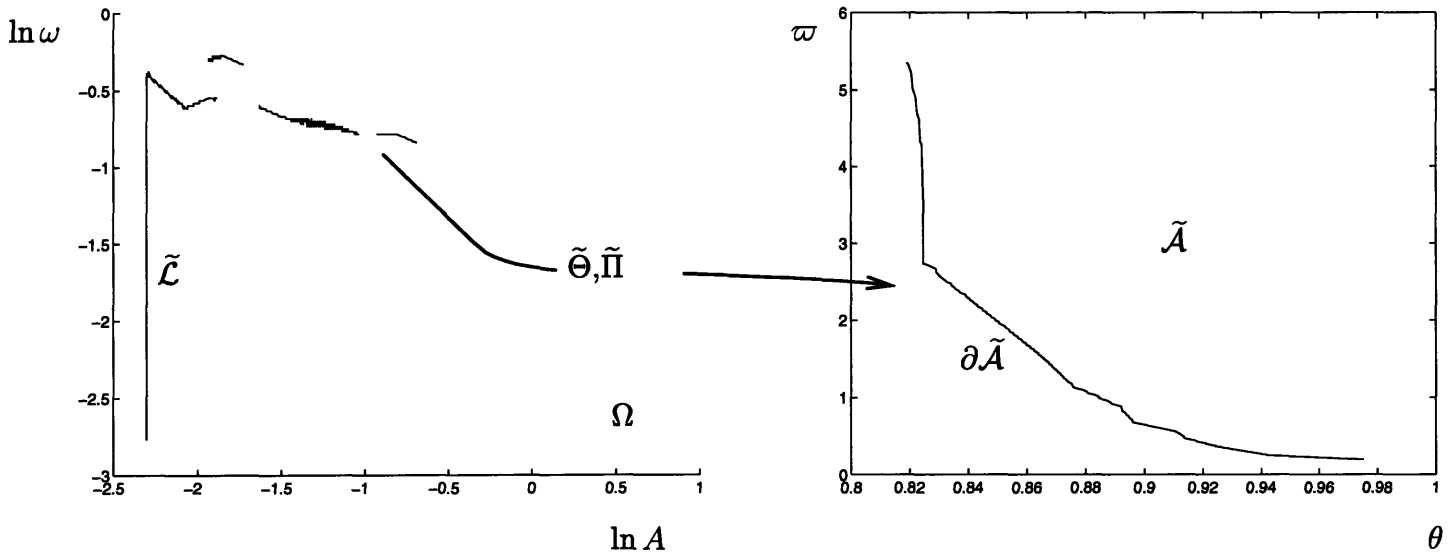


Figure 4-3: Surrogate PO input manifold $\tilde{\mathcal{L}}$ (left) and output manifold $\partial\tilde{\mathcal{A}}$ (right).

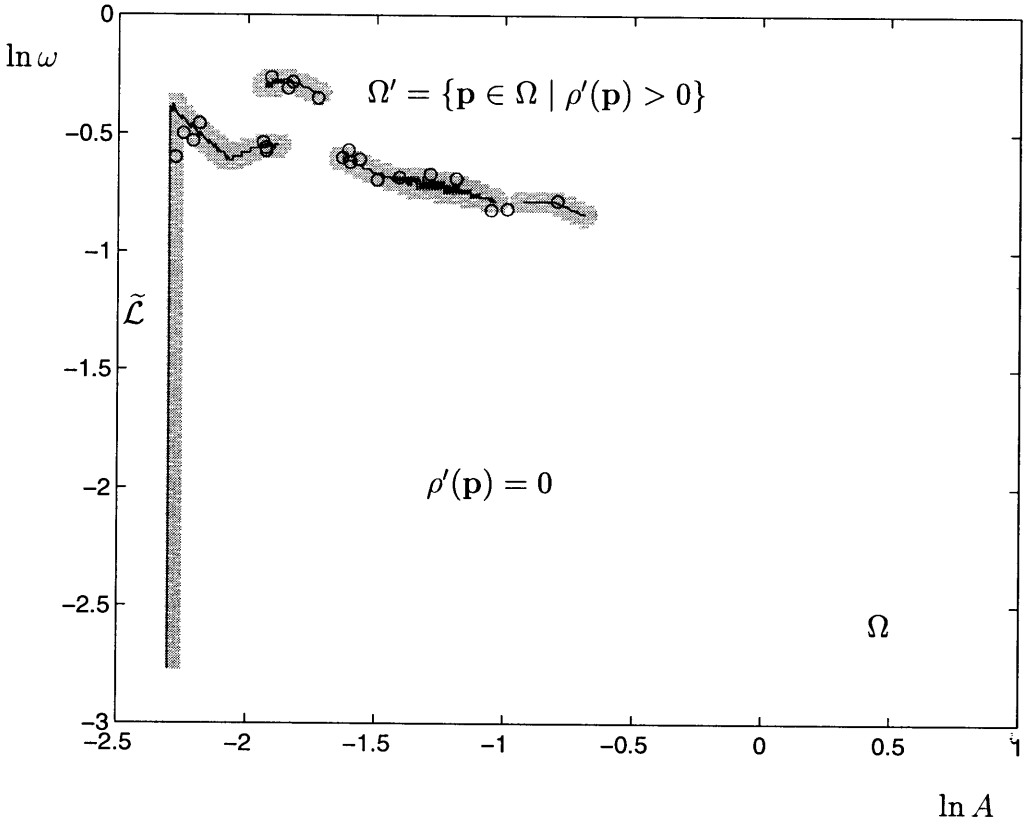


Figure 4-4: Region $\Omega' \subset \Omega$ over which the importance function $\rho'(\mathbf{p})$ is nonzero, and associated validation sample points $\mathbf{P}'_i$ ($\circ$).

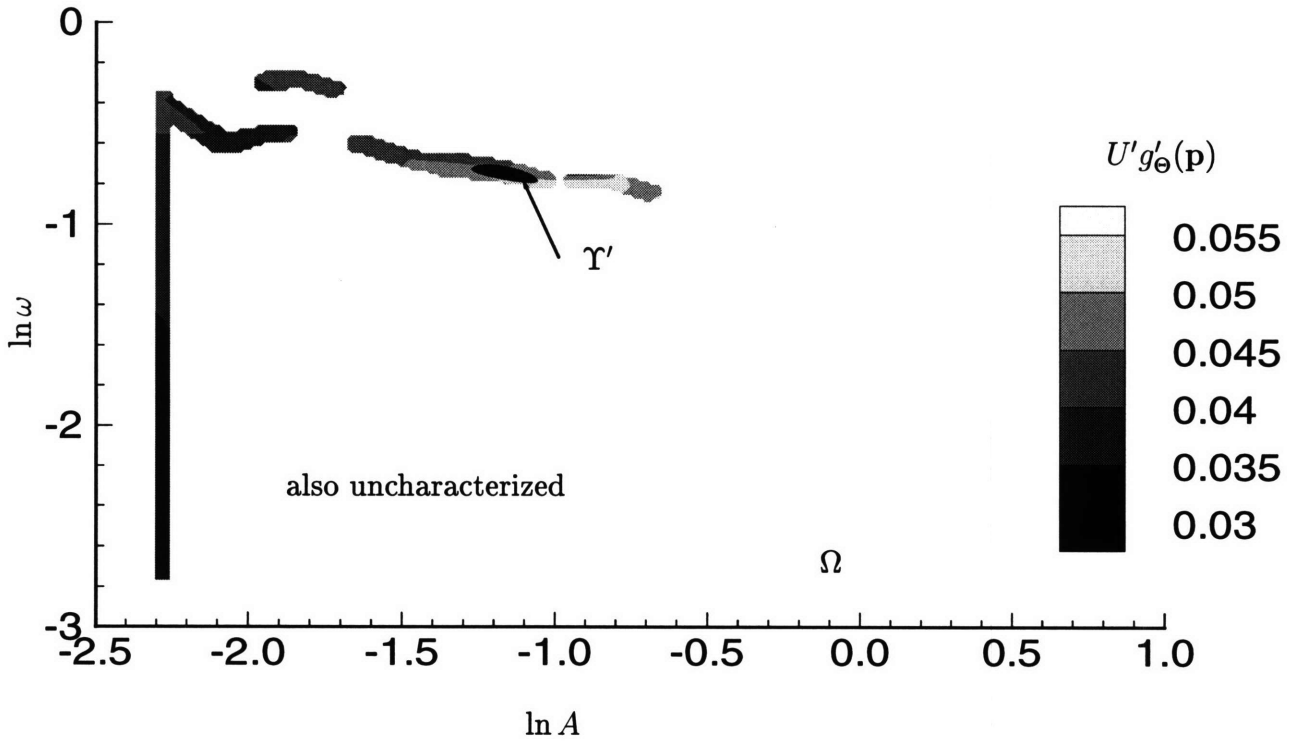


Figure 4-5: Distribution of surrogate error for the temperature in Ω', and a *possible* location of the uncharacterized region, Υ'.

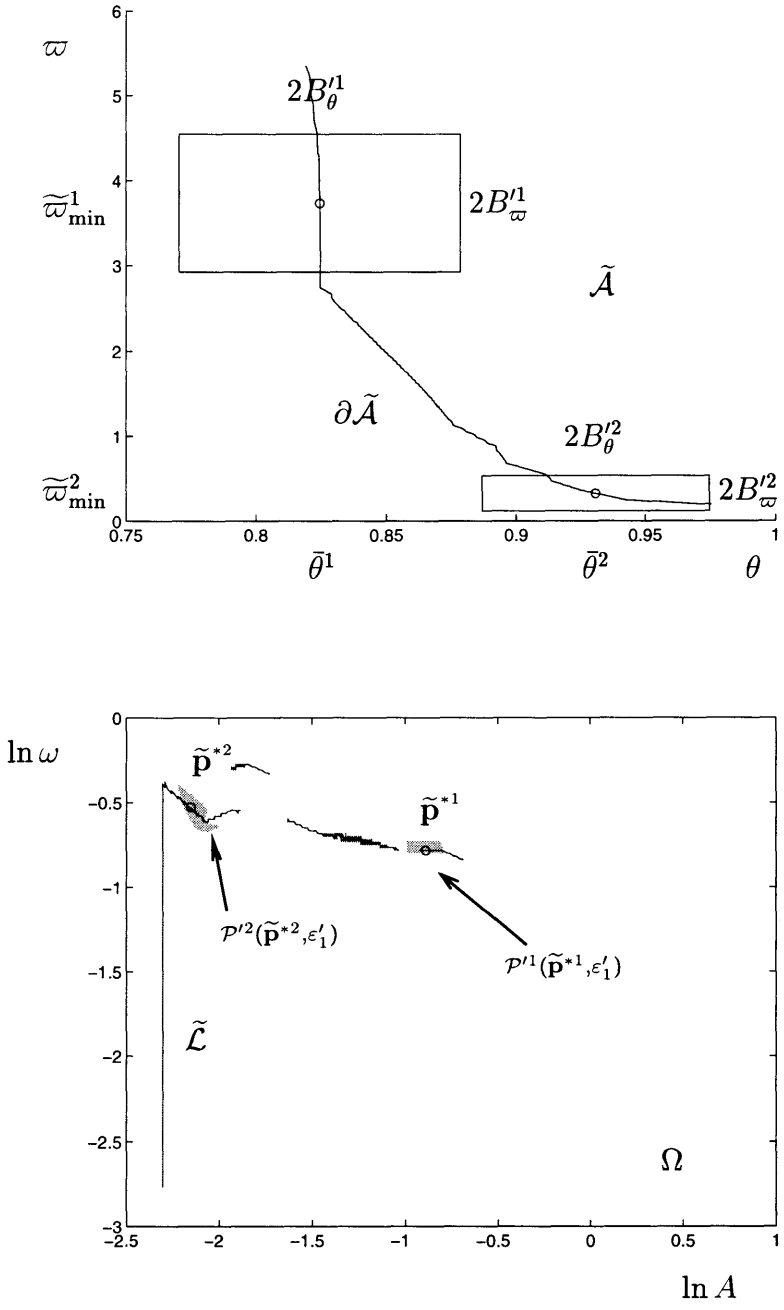


Figure 4-6: Surrogate minimizers $(\widetilde{\mathbf{p}}^{*1}, \widetilde{\mathbf{p}}^{*2})$ and minima $((\bar{\theta}^1, \widetilde{\varpi}^1_{\min}), (\bar{\theta}^2, \widetilde{\varpi}^2_{\min}))$, and associated prediction neighborhoods $\mathcal{P}'^q(\widetilde{\mathbf{p}}^{*q}, \varepsilon'_1)$, $q = 1, 2$, and predictability gaps (boxes).

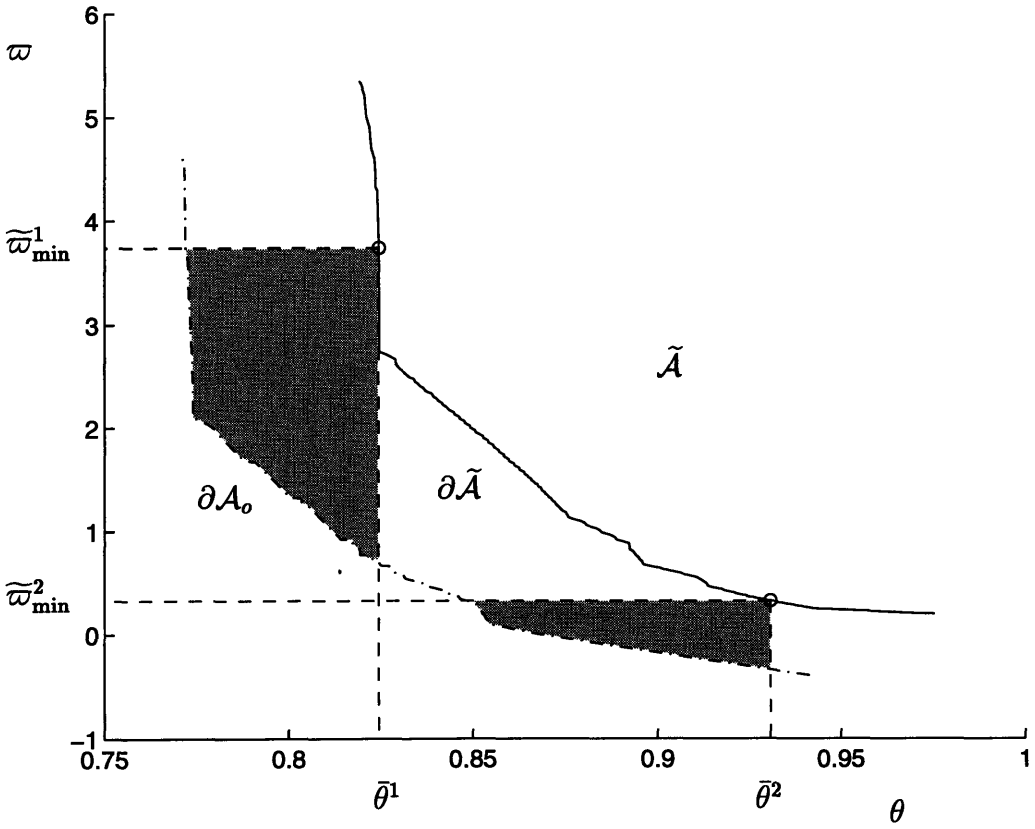


Figure 4-7: Lower bound Pareto manifold $\partial\mathcal{A}_o$ and associated improvement (shaded) for the two designs of Figure 4–4.

Chapter 5

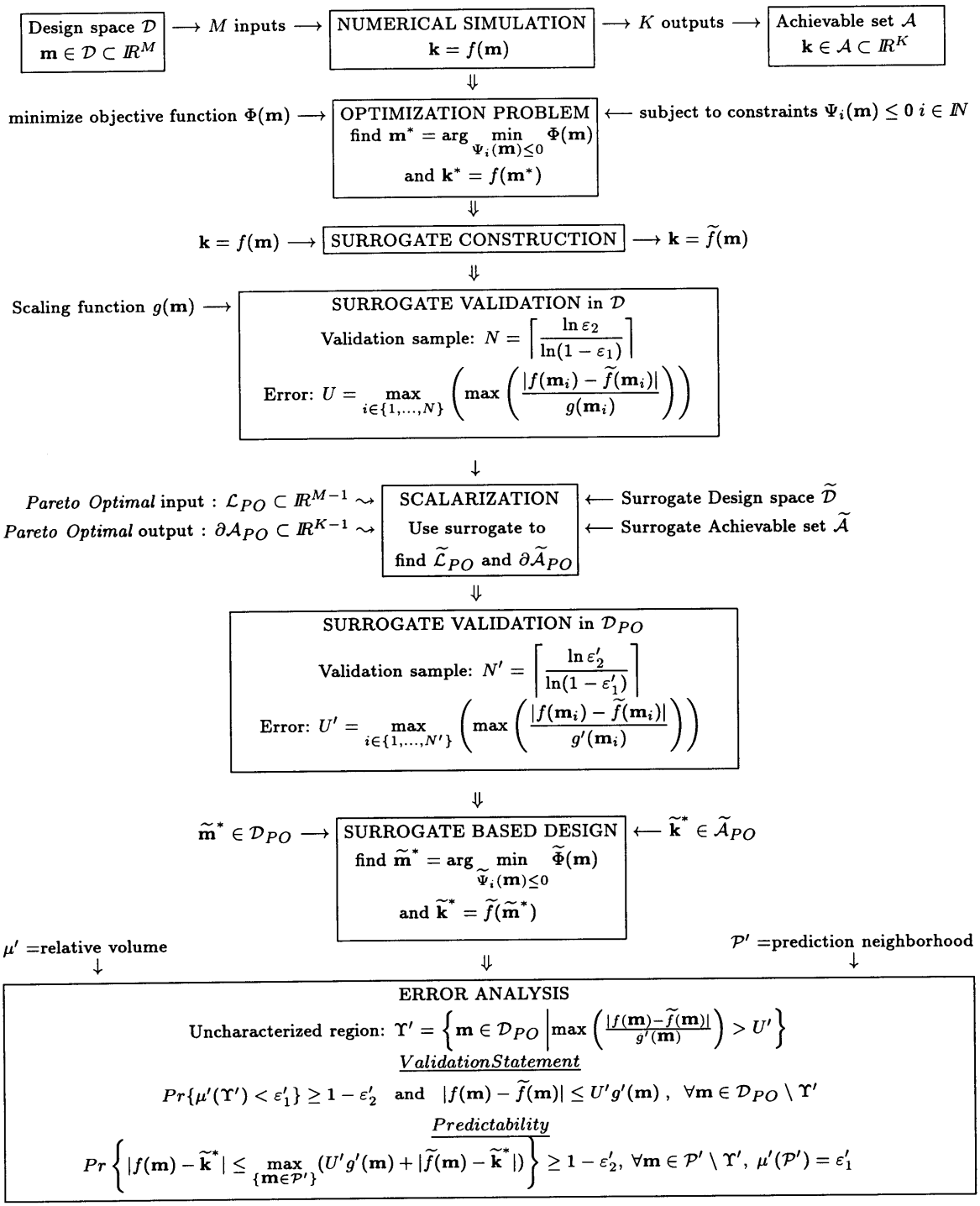
Final Remarks

5.1 Summary

We have presented basic and Pareto formulations of a nonparametric–validated computer–simulation surrogate approach through an illustrative application from fluid dynamics. The Pareto approach, when applicable, permits the consideration of a much larger class of applications both as regards multicriteria problem statement and input (design–variable) dimensionality.

The critical ingredients of both formulations have been developed and demonstrated through the optimization problem associated with the eddy–promoter channel heat–exchanger. Many of the techniques used were problem–specific; we need to emphasize that the user has significant freedom in adjusting the procedure according to the requirements of the problem. Therefore, instead of summarizing our work by refering again to the application of this thesis, we present in the next page an abstract layout of the Surrogate Pareto formulation, where all the assumptions concerning Pareto optimality hold.

SURROGATE PARETO FORMULATION



5.2 Comments and Conclusions

In our particular application, the system was characterized by $M = 2$ inputs and $K = 2$ outputs. The PO sets ($\partial\mathcal{A}^T$ and $\mathcal{L}^T$) had nice topological properties and consequently the graphical representation of the Pareto formulation was simple. In higher dimensions, though, there is no guarantee that the PO input manifold will be "connected"; it might be scattered in the design space. The sparsity of the PO input points might weaken stability and amplify the sensitivity contribution. The latter will result to higher predictability gaps that might yield undistinguishable performance. Signs of such a pitfall appeared in practice. When a third input was introduced (the Reynolds number), the resulting $\mathcal{L}^T$ (now a two–dimensional surface) had a few scattered pieces in the design space ($I\!R^{M=3}$); some of them were actually single points. Partially, this problem can be cured by introducing an appropriate input space Ω' that captures the scattered pieces (or points.) Another possible palliative would be to divide Ω' into subdomains with different density functions.

Our results indicate another outstanding problem — the validation cost. Validation against the truth model is, by definition, very expensive. In the final Pareto analysis of Section 4.5 the predominant contribution to the predictability gap is, in fact, the surrogate error, not the sensitivity contribution; however, for $K > 2$, ε_1 must certainly be decreased, and hence N, the number of truth evaluations, increased. In fact, for purposes of *validation*, we need not evaluate $\Theta^T(\mathbf{p}), \Pi^T(\mathbf{p})$ exactly, but only obtain (sufficiently) sharp bounds on these quantities. Suggested methods to obtain precise quantitative bounds are given in [46].

One other problem, associated with cost, is the fact that the high–fidelity model is (here, and often) too inaccurate and too expensive. As described above, the major contribution to the predictability gap of (3.25) are the terms $|\Theta^T(\mathbf{p}) - \Theta(\mathbf{p})|$ and $|\Theta(\mathbf{p}) - \tilde{\Theta}(\mathbf{p})|$; the former reflects the inaccuracy of $\Theta(\mathbf{p})$, the latter reflects the expense of $\Theta(\mathbf{p})$ (which limits the construction sample size N^{co}). We believe that Proper Orthogonal Decomposition (POD) techniques [47, 48] can be of relevance

here. Further discussion is again provided in [46].

In conclusion, the surrogate Pareto formulation developed in this thesis, although at an early stage of evolution, offers a complete optimization framework. As it stands, it is general enough to be applied to disciplines outside the engineering domain; or lend some of its principles to other numerical optimization routines that face dimensionality hurdles. At the same time, it is flexible enough to borrow ideas from other work in the numerical optimization field. The combination of innovative bounding and approximation techniques with the surrogate Pareto notions discussed here may ultimately permit expensive simulations to participate in high–dimensional, multicriteria optimization problems.

Appendix A

Tables of data points for the illustrative example

A.1 Surrogate construction data

j	p_1	p_2	$\Theta(\mathbf{p}_j^{co})$	$\Pi(\mathbf{p}_j^{co})$	$g_\Theta(\mathbf{p}_j^{co})$	$g_\Pi(\mathbf{p}_j^{co})$
1	-2.3025851e+00	-2.7672931e+00	9.7515526e-01	1.9604966e-01	1.0549496e-01	1.8538666e+00
2	-2.3025851e+00	-2.0892596e+00	9.7426909e-01	1.9747462e-01	1.0498366e-01	1.8690480e+00
3	-2.3025851e+00	-1.4063166e+00	9.6885322e-01	2.0383915e-01	1.1030071e-01	1.8717428e+00
4	-2.3025851e+00	-7.2607279e-01	9.5894777e-01	2.2180047e-01	1.2124586e-01	1.8299312e+00
5	-2.3025851e+00	-4.6655803e-02	9.5597759e-01	2.9214965e-01	1.2803880e-01	1.8521908e+00
6	-2.3025851e+00	6.3390426e-01	9.7561132e-01	3.7346645e-01	1.0561583e-01	1.9115034e+00
7	-1.8420007e+00	-2.7672931e+00	9.7412984e-01	4.8524615e-01	1.0475893e-01	1.8850619e+00
8	-1.8420007e+00	-2.0892596e+00	9.7174154e-01	4.8943105e-01	1.0549469e-01	1.9200140e+00
9	-1.8420007e+00	-1.4063166e+00	9.5946566e-01	5.0630550e-01	1.1614759e-01	1.9231544e+00
10	-1.8420007e+00	-7.2607279e-01	9.3629848e-01	5.5695028e-01	1.2834408e-01	1.8205171e+00
11	-1.8420007e+00	-4.6655803e-02	9.4196015e-01	7.3561262e-01	1.1389242e-01	1.8362460e+00
12	-1.8420007e+00	6.3390426e-01	9.7698879e-01	9.3831178e-01	1.0961114e-01	2.0325452e+00
13	-1.3815058e+00	-2.7672931e+00	9.7286543e-01	1.2032547e+00	1.0766624e-01	2.0337801e+00
14	-1.3815058e+00	-2.0892596e+00	9.6646189e-01	1.2135544e+00	1.0760021e-01	2.1202173e+00
15	-1.3815058e+00	-1.4063166e+00	9.3943878e-01	1.2546230e+00	1.2600721e-01	2.0733865e+00
16	-1.3815058e+00	-7.2607279e-01	8.6906632e-01	1.3982374e+00	1.5595140e-01	1.9765162e+00
17	-1.3815058e+00	-4.6655803e-02	9.2450551e-01	1.8486834e+00	1.1209197e-01	1.9103112e+00
18	-1.3815058e+00	6.3390426e-01	9.7924221e-01	2.3475698e+00	1.1698386e-01	2.4105247e+00
19	-9.2105205e-01	-2.7672931e+00	9.7116093e-01	2.9798720e+00	1.1273930e-01	2.4009458e+00
20	-9.2105205e-01	-2.0892596e+00	9.5482744e-01	3.0057202e+00	1.1649738e-01	2.6128366e+00

21	-9.2105205e-01	-1.4063166e+00	9.0264668e-01	3.0993735e+00	1.3294025e-01	2.1322611e+00
22	-9.2105205e-01	-7.2607279e-01	8.3471187e-01	3.4881390e+00	1.1321970e-01	2.0143479e+00
23	-9.2105205e-01	-4.6655803e-02	9.0886123e-01	4.5950960e+00	1.1809260e-01	2.0865610e+00
24	-9.2105205e-01	6.3390426e-01	9.8448267e-01	5.8930475e+00	1.2896092e-01	3.3108744e+00
25	-4.6044942e-01	-2.7672931e+00	9.7187185e-01	7.3709305e+00	1.1884880e-01	3.3055804e+00
26	-4.6044942e-01	-2.0892596e+00	9.3227409e-01	7.4297255e+00	1.3715349e-01	3.7440693e+00
27	-4.6044942e-01	-1.4063166e+00	8.6197488e-01	7.6133980e+00	1.2856515e-01	3.1139216e+00
28	-4.6044942e-01	-7.2607279e-01	8.6060138e-01	8.6358435e+00	1.3633901e-01	2.7428702e+00
29	-4.6044942e-01	-4.6655803e-02	9.2015138e-01	1.1281968e+01	1.1912484e-01	2.7597074e+00
30	-4.6044942e-01	6.3390426e-01	9.9720587e-01	1.4552053e+01	1.3580806e-01	5.5274464e+00
31	0.0000000e+00	-2.7672931e+00	9.6909631e-01	1.8204320e+01	1.3658389e-01	5.2888721e+00
32	0.0000000e+00	-2.0892596e+00	9.1125339e-01	1.8083822e+01	1.5093211e-01	5.4887010e+00
33	0.0000000e+00	-1.4063166e+00	8.6312545e-01	1.8540733e+01	1.8597217e-01	4.5071925e+00
34	0.0000000e+00	-7.2607279e-01	1.0148229e+00	2.2433865e+01	1.9296996e-01	4.6879636e+00
35	0.0000000e+00	-4.6655803e-02	9.6958979e-01	2.7412750e+01	1.4325132e-01	4.9654274e+00
36	0.0000000e+00	6.3390426e-01	1.0330104e+00	3.5676558e+01	1.4115709e-01	6.5793725e+00
37	-2.0722673e+00	-2.4237034e+00	9.7353546e-01	3.0847323e-01	1.0431214e-01	1.8693328e+00
38	-2.0722673e+00	-1.7484458e+00	9.7146408e-01	3.1305168e-01	1.1098659e-01	1.8705388e+00
39	-2.0722673e+00	-1.0661880e+00	9.6201766e-01	3.3162745e-01	1.1250447e-01	1.8616760e+00
40	-2.0722673e+00	-3.8682149e-01	9.3052078e-01	3.9200143e-01	1.1803443e-01	1.7976394e+00
41	-2.0722673e+00	2.9375862e-01	9.7395260e-01	5.2579070e-01	1.0977930e-01	1.8943505e+00
42	-1.6119410e+00	-2.4237034e+00	9.7190910e-01	7.6659015e-01	1.0548507e-01	1.9963171e+00
43	-1.6119410e+00	-1.7484458e+00	9.6674863e-01	7.7890669e-01	1.1699983e-01	1.9977118e+00
44	-1.6119410e+00	-1.0661880e+00	9.4664239e-01	8.2604043e-01	1.1941609e-01	1.8784055e+00
45	-1.6119410e+00	-3.8682149e-01	8.9688265e-01	9.9464359e-01	1.4607948e-01	1.8155977e+00
45	-1.6119410e+00	-3.8682149e-01	8.9688265e-01	9.9464359e-01	1.4607948e-01	1.8155977e+00
46	-1.6119410e+00	2.9375862e-01	9.7194368e-01	1.3162789e+00	1.1034040e-01	2.2199990e+00
47	-1.1513804e+00	-2.4237034e+00	9.6774667e-01	1.8980833e+00	1.0937315e-01	2.3103595e+00
48	-1.1513804e+00	-1.7484458e+00	9.5489154e-01	1.9298546e+00	1.2564492e-01	2.3131503e+00
49	-1.1513804e+00	-1.0661880e+00	9.2174111e-01	2.0420024e+00	1.3472865e-01	1.8964742e+00
50	-1.1513804e+00	-3.8682149e-01	8.5174757e-01	2.5619312e+00	1.4562327e-01	1.9474633e+00
51	-1.1513804e+00	2.9375862e-01	9.6817481e-01	3.3041313e+00	1.1385657e-01	2.8287739e+00
52	-6.9075006e-01	-2.4237034e+00	9.6205390e-01	4.7002372e+00	1.1924541e-01	3.0811751e+00
53	-6.9075006e-01	-1.7484458e+00	9.2361771e-01	4.7736621e+00	1.4328826e-01	2.8611890e+00
54	-6.9075006e-01	-1.0661880e+00	8.8977923e-01	5.0226231e+00	1.6868793e-01	2.8198741e+00
55	-6.9075006e-01	-3.8682149e-01	8.7333391e-01	6.3355295e+00	1.1962884e-01	2.2027219e+00
56	-6.9075006e-01	2.9375862e-01	9.6458043e-01	8.1723783e+00	1.2644382e-01	4.3186658e+00
57	-2.3029406e-01	-2.4237034e+00	9.5420485e-01	1.1612770e+01	1.2274984e-01	3.8361876e+00
58	-2.3029406e-01	-1.7484458e+00	8.6203002e-01	1.1680458e+01	1.5796035e-01	3.1185961e+00
59	-2.3029406e-01	-1.0661880e+00	9.0322346e-01	1.2770114e+01	2.0944709e-01	3.2460651e+00
60	-2.3029406e-01	-3.8682149e-01	9.3150590e-01	1.5728464e+01	1.2529881e-01	2.5646666e+00

61	-2.3029406e-01	2.9375862e-01	9.8292847e-01	2.0080173e+01	1.2634924e-01	3.8884627e+00
62	-3.4531119e-01	-8.9549094e-01	8.6186400e-01	1.0476015e+01	1.3540618e-01	2.6875686e+00
63	-3.4531119e-01	-5.5682332e-01	9.0751873e-01	1.1887442e+01	1.3346943e-01	2.5524126e+00
64	-3.4531119e-01	-2.1706700e-01	9.1022129e-01	1.3384343e+01	1.2191070e-01	2.9270573e+00
65	-3.4531119e-01	1.2363404e-01	9.4991805e-01	1.5055468e+01	1.1938727e-01	3.0915519e+00
66	-2.3016817e-01	-7.2607279e-01	9.2202917e-01	1.4109071e+01	1.3763741e-01	2.5445548e+00
67	-2.3016817e-01	-4.6655803e-02	9.4046937e-01	1.7692573e+01	1.1293426e-01	2.7712903e+00
68	-1.1518641e-01	-1.2368984e+00	8.6574355e-01	1.5046033e+01	1.9342175e-01	3.1434817e+00
69	-1.1518641e-01	-8.9549094e-01	9.3194224e-01	1.6621637e+01	1.5908297e-01	2.7865052e+00
70	-1.1518641e-01	-5.5682332e-01	9.6353235e-01	1.8685525e+01	1.4095557e-01	2.5905198e+00
71	-1.1518641e-01	-2.1706700e-01	9.4301571e-01	2.0864720e+01	1.2554065e-01	2.7650206e+00
72	-1.1518641e-01	1.2363404e-01	9.6462604e-01	2.3473180e+01	1.2593825e-01	2.8599390e+00
73	0.0000000e+00	-1.0661880e+00	9.9651911e-01	2.0250623e+01	2.2973782e-01	4.6772213e+00
74	0.0000000e+00	-3.8682149e-01	9.7300045e-01	2.4689525e+01	1.4256726e-01	4.6884103e+00
75	-2.0722673e+00	-6.1320803e-01	9.2442253e-01	3.6527421e-01	1.2421846e-01	1.7974893e+00
76	-1.8420007e+00	-2.7326156e-01	8.9613513e-01	6.7601441e-01	1.3275624e-01	1.8136880e+00
77	-1.3815058e+00	-2.7326156e-01	8.7731516e-01	1.6882407e+00	1.3380338e-01	1.8698529e+00
78	-1.3815058e+00	-5.0033520e-01	8.9031628e-01	1.4890098e+00	1.3327255e-01	1.8439417e+00
79	-1.1513804e+00	-6.1320803e-01	8.6769630e-01	2.2728223e+00	1.2912302e-01	1.9197172e+00
80	-1.9568700e+00	-5.5682332e-01	9.1374529e-01	4.6742590e-01	1.2401346e-01	1.8073073e+00
81	-1.6119410e+00	-6.1320803e-01	8.8898069e-01	9.1487863e-01	1.3318671e-01	1.8098343e+00
82	-1.6119410e+00	-1.6016884e-01	9.1101442e-01	1.1046688e+00	1.2940634e-01	1.8441630e+00
83	-1.8420007e+00	-5.0033520e-01	9.1192591e-01	5.9523748e-01	1.1530333e-01	1.8001935e+00
84	-2.3025851e+00	-3.8682149e-01	9.4211847e-01	2.4849445e-01	1.3019381e-01	1.8343672e+00
85	-2.0722673e+00	-1.6016884e-01	9.2697083e-01	4.4330196e-01	1.2796169e-01	1.8126271e+00
86	-1.1513804e+00	-1.6016884e-01	9.0259492e-01	2.7759616e+00	1.1886023e-01	1.9877286e+00
87	-1.1513804e+00	-8.4012901e-01	8.8140577e-01	2.1300148e+00	1.5602542e-01	1.9425628e+00
88	-1.8420007e+00	-9.5246838e-01	9.5650131e-01	5.2965799e-01	1.2137604e-01	1.8390433e+00
89	-1.3815058e+00	-9.5246838e-01	9.3551950e-01	1.3193556e+00	1.4646955e-01	1.8622060e+00
90	-9.2105205e-01	-9.5246838e-01	8.8945472e-01	3.2631735e+00	1.5955226e-01	1.9880213e+00
91	-1.6119410e+00	-8.4012901e-01	9.4122497e-01	8.5332276e-01	1.3694577e-01	1.8348027e+00
92	-9.2105205e-01	-1.1801008e+00	9.0474646e-01	3.1646646e+00	1.1843431e-01	1.9498051e+00
93	-1.1513804e+00	-1.2925301e+00	9.1983459e-01	1.9946438e+00	1.1294204e-01	1.9776905e+00
94	-1.3815058e+00	-1.1801008e+00	9.2957148e-01	1.2843602e+00	1.1633299e-01	1.8620993e+00
95	-2.0722673e+00	6.7096004e-02	9.6183228e-01	4.8580598e-01	1.2051928e-01	1.8196196e+00
96	-1.6119410e+00	6.7096004e-02	9.5006934e-01	1.2109082e+00	1.1314063e-01	1.9207750e+00
97	-1.1513804e+00	6.7096004e-02	9.3955157e-01	3.0326436e+00	1.1262040e-01	2.1178670e+00
98	-2.0722673e+00	-8.4012901e-01	9.6062824e-01	3.4183815e-01	1.2610963e-01	1.8024101e+00
99	-1.8420007e+00	1.8029878e-01	9.6848178e-01	7.9398126e-01	1.1553633e-01	1.9933559e+00
100	-1.3815058e+00	1.8029878e-01	9.6411868e-01	1.9992224e+00	1.1616441e-01	2.3243848e+00

101	-9.2105205e-01	1.8029878e-01	9.5657514e-01	4.9874683e+00	1.1919489e-01	3.1088230e+00
102	-6.9075006e-01	6.7096004e-02	9.3730204e-01	7.5046668e+00	1.1757306e-01	2.9908483e+00
103	-6.9075006e-01	-1.6016884e-01	8.9357316e-01	6.9443679e+00	1.2151528e-01	2.5380073e+00
104	-9.2105205e-01	-2.7326156e-01	8.8157372e-01	4.1776895e+00	1.2236945e-01	2.1107303e+00
105	-9.2105205e-01	-5.0033520e-01	8.4993203e-01	3.7677565e+00	1.4846986e-01	2.1645321e+00
106	-6.9075006e-01	-6.1320803e-01	8.5684865e-01	5.6577808e+00	1.4165351e-01	2.4555197e+00
107	-6.9075006e-01	-8.4012901e-01	8.1918146e-01	5.3506497e+00	1.5148492e-01	2.3567955e+00
108	-6.9075006e-01	-1.2925301e+00	8.9565113e-01	4.9148495e+00	1.2242882e-01	2.0345736e+00
109	-4.6044942e-01	-1.1801008e+00	8.8392096e-01	7.7589724e+00	1.4503466e-01	2.8842949e+00
110	-6.9075006e-01	-1.5202608e+00	8.8450173e-01	4.8239740e+00	1.4144203e-01	2.0717400e+00
111	-4.6044942e-01	-1.6358910e+00	8.6012548e-01	7.5163329e+00	1.4697707e-01	2.4853995e+00
112	-2.3029406e-01	-1.5202608e+00	8.3949726e-01	1.1795145e+01	1.7331664e-01	3.1956885e+00
113	-4.6044942e-01	1.8029878e-01	9.5165807e-01	1.2310677e+01	1.2486251e-01	3.8852962e+00
114	-4.6044942e-01	-2.7326156e-01	9.0170252e-01	1.0420700e+01	1.2237001e-01	2.4752954e+00
115	-4.6044942e-01	-5.0033520e-01	8.8957375e-01	9.6315023e+00	1.2169738e-01	2.3744788e+00
116	-4.6044942e-01	-9.5246838e-01	8.3818291e-01	8.2380648e+00	2.0034997e-01	2.7627952e+00
117	-2.3029406e-01	-1.2925301e+00	8.7246428e-01	1.1998112e+01	1.2074123e-01	3.1357701e+00
118	0.0000000e+00	-1.7484458e+00	8.9117761e-01	1.8165834e+01	1.3980681e-01	5.5621713e+00
119	0.0000000e+00	2.9375862e-01	1.0079612e+00	3.1201727e+01	1.4623276e-01	6.0647891e+00
120	-2.3025851e+00	2.9375862e-01	9.7428257e-01	3.3189004e-01	1.1026124e-01	1.8953131e+00
121	-1.9568700e+00	-3.3017713e-01	9.1514635e-01	5.1491577e-01	1.2339666e-01	1.7983512e+00
122	-1.7270960e+00	-4.4392549e-01	9.1042412e-01	7.6301051e-01	1.2077721e-01	1.7988252e+00
123	-1.4965558e+00	-5.5682332e-01	8.9114481e-01	1.1653517e+00	1.1857487e-01	1.8230608e+00
124	-1.4965558e+00	-3.3017713e-01	8.7173193e-01	1.3145310e+00	1.4220664e-01	1.8133823e+00
125	-1.2665577e+00	-4.4392549e-01	8.7418754e-01	1.9297047e+00	1.4555742e-01	1.8600627e+00
126	-1.2665577e+00	-6.6927519e-01	8.5734207e-01	1.7860977e+00	1.2588122e-01	1.8918732e+00
127	-1.0362010e+00	-5.5682332e-01	8.7083419e-01	2.9034706e+00	1.4278083e-01	2.0535183e+00
128	-1.0362010e+00	-7.8353683e-01	8.2465044e-01	2.7416273e+00	1.5944206e-01	2.0184508e+00
129	-8.0586805e-01	-7.8353683e-01	8.2370588e-01	4.3237375e+00	1.6287394e-01	2.0279496e+00
130	-2.1874723e+00	-3.8682149e-01	9.3686157e-01	3.1310297e-01	1.0990623e-01	1.8015022e+00
131	-1.9568700e+00	-4.4392549e-01	9.2218430e-01	4.8174599e-01	1.1698537e-01	1.7997092e+00
132	-1.7270960e+00	-3.3017713e-01	8.9245315e-01	8.2348601e-01	1.2974015e-01	1.7953063e+00
133	-1.6119410e+00	-5.0033520e-01	9.0155310e-01	9.4094669e-01	1.2974610e-01	1.8090779e+00
134	-2.1874723e+00	-6.6927519e-01	9.4611178e-01	2.8638619e-01	1.2304255e-01	1.7929290e+00
135	-1.7270960e+00	-5.5682332e-01	9.0191250e-01	7.3795927e-01	1.2825712e-01	1.8189770e+00
136	-1.4965558e+00	-6.6927519e-01	8.7572911e-01	1.1319600e+00	1.3859668e-01	1.8458037e+00
137	-1.3815058e+00	-3.8682149e-01	8.6848544e-01	1.5988181e+00	1.3384691e-01	1.8085060e+00
138	-1.8420007e+00	-3.8682149e-01	9.1660726e-01	6.2451770e-01	1.2861420e-01	1.8028019e+00

End of the table.

A.2 Surrogate Basic formulation validation data

#	p_1	p_2	$\Theta^T(\mathbf{p})$	$\Pi^T(\mathbf{p})$	$\widetilde{\Theta}(\mathbf{p})$	$\widetilde{\Pi}(\mathbf{p})$	$\frac{\lvert\Theta^T(\mathbf{p})-\widetilde{\Theta}(\mathbf{p})\rvert}{g_\Theta(\mathbf{p})}$	$\frac{\lvert\Pi^T(\mathbf{p})-\widetilde{\Pi}(\mathbf{p})\rvert}{g_\Pi(\mathbf{p})}$
1	-1.16	-1.47	0.93	2.11	0.94	1.87	0.07	0.11
2	-0.14	0.44	1.00	26.22	0.94	24.20	0.44	0.41
3	-1.68	-1.37	0.95	0.77	0.96	0.67	0.08	0.05
4	-2.12	-0.47	0.93	0.35	0.93	0.32	0.03	0.01
5	-1.75	-1.52	0.96	0.63	0.97	0.57	0.12	0.03
6	-1.09	0.48	0.98	4.23	0.97	3.70	0.05	0.18
7	-1.79	-2.32	0.97	0.57	0.98	0.51	0.12	0.03
8	-0.29	-0.37	0.92	14.27	0.86	13.89	0.52	0.15
9	-2.06	-1.23	0.96	0.35	0.97	0.32	0.07	0.02
10	-2.13	0.13	0.97	0.45	0.96	0.41	0.05	0.02
11	-1.60	-1.56	0.96	0.86	0.97	0.77	0.13	0.04
12	-1.29	-2.29	0.97	1.52	0.98	1.41	0.08	0.05
13	-1.19	-1.52	0.94	1.94	0.95	1.75	0.09	0.09
14	-1.80	0.11	0.96	0.86	0.94	0.79	0.20	0.03
15	-1.87	-1.91	0.97	0.49	0.98	0.44	0.12	0.03
16	-1.52	-1.16	0.94	1.01	0.95	0.94	0.03	0.04
17	-1.74	-1.38	0.95	0.67	0.96	0.59	0.08	0.04
18	-0.01	-2.55	0.95	17.86	0.92	18.28	0.22	0.08
19	-2.25	-1.71	0.97	0.23	0.99	0.21	0.14	0.01
20	-1.95	-0.89	0.96	0.44	0.97	0.41	0.11	0.02
21	-1.69	-0.29	0.90	0.90	0.85	0.85	0.40	0.03
22	-0.26	-1.29	0.87	11.52	0.85	11.49	0.20	0.01
23	-0.49	-0.14	0.91	10.68	0.84	10.15	0.53	0.19
24	-1.41	-1.32	0.94	1.20	0.94	1.14	0.03	0.03
25	-0.47	-0.05	0.92	11.03	0.85	10.74	**0.53**	0.11
26	-1.07	-1.36	0.91	2.43	0.92	2.26	0.02	0.08
27	-1.92	-0.44	0.92	0.53	0.92	0.49	0.01	0.02
28	-0.10	-1.90	0.89	15.37	0.87	15.35	0.10	0.01
29	-2.05	0.62	0.98	0.68	0.99	0.57	0.11	0.06
30	-1.45	-1.59	0.95	1.14	0.97	1.04	0.13	0.05
31	-0.10	0.28	0.99	26.10	0.93	25.02	0.48	0.22
32	-1.04	-1.61	0.93	2.49	0.95	2.33	0.10	0.07
33	-1.85	-0.99	0.96	0.54	0.97	0.49	0.11	0.02
34	-1.22	-0.52	0.87	2.07	0.86	1.94	0.08	0.07
35	-1.34	-0.23	0.89	1.91	0.84	1.76	0.35	0.08

36	-1.53	-0.05	0.93	1.41	0.88	1.31	0.43	0.05
37	-1.63	-1.67	0.96	0.78	0.98	0.72	0.11	0.03
38	-0.61	-0.37	0.88	7.63	0.83	7.23	0.46	0.17
39	-1.09	-1.97	0.96	2.29	0.96	2.11	0.02	0.08
40	-0.32	-2.30	0.95	9.99	0.93	9.78	0.16	0.05
41	-0.21	-0.96	0.91	13.88	0.84	13.67	0.37	0.07
42	-1.12	-1.37	0.92	2.23	0.92	2.03	0.03	0.10
43	-1.04	0.40	0.97	4.52	0.96	3.98	0.09	0.17
44	-1.91	-0.55	0.91	0.53	0.91	0.49	0.02	0.02
45	-0.16	0.21	0.98	22.63	0.91	21.97	0.48	0.18
46	-0.84	-1.45	0.90	3.70	0.90	3.53	0.01	0.08
47	-2.16	-2.24	0.97	0.27	0.99	0.24	0.13	0.01
48	-0.95	-1.99	0.95	2.90	0.95	2.75	0.01	0.06
49	-0.90	-0.30	0.88	4.35	0.82	4.15	0.44	0.09
50	-1.78	-0.32	0.90	0.75	0.86	0.70	0.25	0.03
51	-1.41	0.08	0.95	1.85	0.90	1.73	0.44	0.06
52	-0.18	0.53	1.01	25.83	0.95	23.10	0.40	0.50
53	-1.19	-0.99	0.91	1.93	0.93	1.82	0.09	0.05

End of the table.

A.3 Surrogate Pareto formulation validation data

#	p_1	p_2	$\Theta^T(\mathbf{p})$	$\Pi^T(\mathbf{p})$	$\widetilde{\Theta}(\mathbf{p})$	$\widetilde{\Pi}(\mathbf{p})$	$\frac{\lvert\Theta^T(\mathbf{p})-\widetilde{\Theta}(\mathbf{p})\rvert}{g_\Theta(\mathbf{p})}$	$\frac{\lvert\Pi^T(\mathbf{p})-\widetilde{\Pi}(\mathbf{p})\rvert}{g_\Pi(\mathbf{p})}$
1	-1.05	-0.82	0.82	2.52	0.85	2.65	0.16	0.23
2	-1.93	-0.57	0.91	0.47	0.91	0.50	0.03	0.09
3	-1.83	-0.28	0.86	0.65	0.90	0.69	0.28	0.11
4	-0.99	-0.82	0.80	2.88	0.84	3.00	0.22	0.23
5	-1.41	-0.68	0.86	1.27	0.87	1.36	0.10	0.20
6	-1.29	-0.67	0.84	1.63	0.86	1.72	0.16	0.20
7	-1.60	-0.62	0.88	0.88	0.89	0.93	0.07	0.14
8	-2.21	-0.53	0.93	0.27	0.94	0.29	0.06	0.06
9	-2.28	-0.60	0.95	0.22	0.95	0.24	0.04	0.05
10	-1.61	-0.57	0.89	0.88	0.89	0.93	0.07	0.14
11	-0.80	-0.78	0.79	4.31	0.82	4.43	0.23	0.19
12	-1.93	-0.56	0.91	0.47	0.91	0.50	0.01	0.09
13	-1.73	-0.35	0.88	0.75	0.90	0.81	0.16	0.18
14	-2.25	-0.50	0.93	0.25	0.94	0.27	0.06	0.05
15	-1.63	-0.60	0.88	0.84	0.89	0.88	0.06	0.14
16	-2.19	-0.46	0.93	0.28	0.93	0.31	0.01	0.07
17	-1.85	-0.31	0.87	0.61	0.90	0.66	0.26	0.12
18	-1.57	-0.61	0.88	0.95	0.89	1.01	0.08	0.16
19	-1.91	-0.26	0.86	0.56	0.91	0.60	**0.32**	0.11
20	-1.19	-0.69	0.82	1.99	0.85	2.11	0.19	0.25
21	-1.94	-0.54	0.91	0.46	0.91	0.49	0.02	0.08
22	-1.50	-0.70	0.88	1.06	0.88	1.13	0.05	0.16

End of the table.

Appendix B

Scaling Functions

We describe here the procedure by which we obtain the appropriate element–based scaling function $g_\Theta(\mathbf{p})$, after we have performed a triangulation of our input space Ω based on N^{co} construction nodes $\mathbf{p}_j^{co}$, $j = 1, ..., N^{co}$, as shown in Figure 3-1. The same procedure applies to $g_\Pi(\mathbf{p})$.

For each edge γ of the triangualation, we find e_γ^1 and e_γ^2, the two triangular elements to which it belongs. Then, we compute the magnitude of the p_1-component, $(n_\gamma)_{p_1}$, and p_2-component, $(n_\gamma)_{p_2}$, of the unit normal of each γ, based on the position (p_1, p_2) of its vertices (endpoints), ν_γ^1 and ν_γ^2. For every element e, we compute the area A^e, and we find the coefficients a^e and b^e that satisfy the system of equations of the element's vertex values which have the general form

$$a^e p_1 + b^e p_2 + c^e = \Theta(\mathbf{p}). \tag{B.1}$$

Then, for every edge γ we calculate a norm H_γ, given by

$$H_\gamma = \left(A^{e_\gamma^1} A^{e_\gamma^2}\right)^{\frac{1}{4}} \left|(a^{e_\gamma^1} - a^{e_\gamma^2})(n_\gamma)_{p_1} + (b^{e_\gamma^1} - b^{e_\gamma^2})(n_\gamma)_{p_2}\right|, \tag{B.2}$$

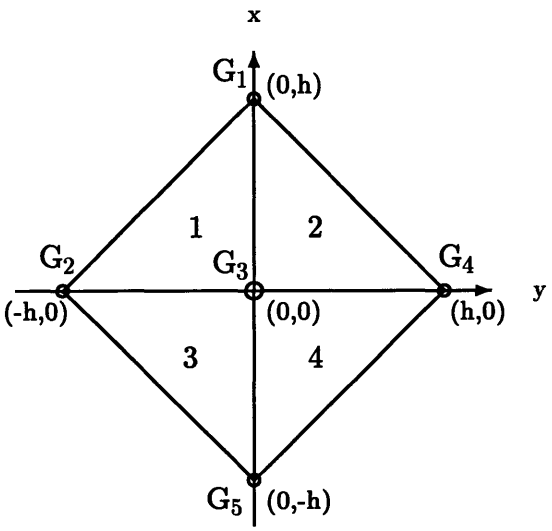
the value of which we accumulate for both vertices ν_γ^1 and ν_γ^2. The norm H_γ is basically a multiple of the absolute difference of the lengths of the normal vectors to

both sides of edge γ: one normal vector lies in e_γ^1 and the other one in e_γ^2, and their lengths are given by $|a^{e_\gamma^1}(n_\gamma)_{p_1} + b^{e_\gamma^1}(n_\gamma)_{p_2}|$ and $|a^{e_\gamma^2}(n_\gamma)_{p_1} + b^{e_\gamma^2}(n_\gamma)_{p_2}|$, respectively. For each construction node, $\mathbf{p}_j^{co}$, the scaling function value, $g_\Theta(\mathbf{p}_j^{co})$, is the cumulative value of H_γ for the corresponding vertex, divided by the number of elements to which it belongs. In other words, the cumulative value of H_γ at each vertex is the sum of the values of the norms H_γ of all the edges that share that vertex. In the case where γ lies on the boundary of Ω and consequently it belongs to only one element, e_γ^1, the value of the norm H_γ is given by

$$H_\gamma = \left(A^{e_\gamma^1}\right)^{\frac{1}{2}} \left|(a^{e_\gamma^1})(n_\gamma)_{p_1} + (b^{e_\gamma^1})(n_\gamma)_{p_2}\right|. \tag{B.3}$$

We can then explicitely compute $g_\Theta(\mathbf{p})$, for any input vector $\mathbf{p} \in \Omega$, as the linear interpolant of the pairs $(\mathbf{p}_j^{co}, g_\Theta(\mathbf{p}_j^{co}))$, $j = 1, ..., N^{co}$. To guard against unbounded amplification at points where $g_\Theta(\mathbf{p})$ is small or vanishes, we add a small constant. For our problem, we found by performing numerical optimization of pseudo-validation results, that by adding the value $0.05 \cdot \max(\Theta(\mathbf{p}))$ to $g_\Theta(\mathbf{p})$ we avoid singularities.

We will demonstrate now, through an example, that our scaling functions accurately reflect the sum of the partial second derivatives of the input-output function. We consider the following configuration:



The central node, has value G_3, and is at a distance h from the other four nodes. To obtain the scaling function value g_3, we first compute the coefficients, a^e, b^e, and c^e, for the four elements, $e = 1, 2, 3, 4$. Solving the matrix equation (B.1) for every element, we find that

$$\begin{aligned}
a^1 &= \frac{G_3 - G_2}{h}, b^1 = \frac{G_1 - G_3}{h}, c^1 = G_3, \\
a^2 &= \frac{G_4 - G_3}{h}, b^2 = \frac{G_1 - G_3}{h}, c^2 = G_3, \\
a^3 &= \frac{G_3 - G_2}{h}, b^3 = \frac{G_3 - G_5}{h}, c^3 = G_3, \\
a^4 &= \frac{G_4 - G_3}{h}, b^4 = \frac{G_3 - G_5}{h}, c^4 = G_3.
\end{aligned}$$

We then compute the norm H_γ for all edges that have the central node as a vertex, $\gamma = N, S, E, W$ (the letters stand for the directions). The area of every element is $\frac{h^2}{2}$, and thus $(A^{e_\gamma^1} A^{e_\gamma^2})^{\frac{1}{4}} = \frac{h}{\sqrt{2}}$. We then have

$$\begin{aligned}
H_N &= \frac{h}{\sqrt{2}} \left| (a^2 - a^1)1 + (b^2 - b^1)0 \right| = \frac{1}{\sqrt{2}} \left| G_2 - 2G_3 + G_4 \right|, \\
H_S &= \frac{h}{\sqrt{2}} \left| (a^4 - a^3)1 + (b^4 - b^3)0 \right| = \frac{1}{\sqrt{2}} \left| G_2 - 2G_3 + G_4 \right|, \\
H_E &= \frac{h}{\sqrt{2}} \left| (a^1 - a^3)0 + (b^1 - b^3)1 \right| = \frac{1}{\sqrt{2}} \left| G_1 - 2G_3 + G_5 \right|, \\
H_W &= \frac{h}{\sqrt{2}} \left| (a^2 - a^4)0 + (b^2 - b^4)1 \right| = \frac{1}{\sqrt{2}} \left| G_1 - 2G_3 + G_5 \right|.
\end{aligned}$$

The number of elements that the central node belongs to is 4, and therefore

$$g_3 = \frac{1}{4} \left(\frac{2}{\sqrt{2}} \left| G_2 - 2G_3 + G_4 \right| + \frac{2}{\sqrt{2}} \left| G_1 - 2G_3 + G_5 \right| \right). \tag{B.4}$$

From the finite difference theory, we know that $|G_{xx}| \equiv \frac{|G_2 - 2G_3 + G_4|}{h^2}$ and $|G_{yy}| \equiv \frac{|G_1 - 2G_3 + G_5|}{h^2}$ (second term of the Taylor series); hence, B.4 is equivalent to

$$g_3 = \frac{h^2}{2\sqrt{2}} (|G_{xx}| + |G_{yy}|). \tag{B.5}$$

Note that this example is chosen specifically to demonstrate that our norm is in fact better than the Laplacian norm, $|\nabla^2| = |G_{xx} + G_{yy}| = |G_1 + G_2 - 4G_3 + G_4 + G_5|$, since the latter could be zero in the case where $G_{xx} = -G_{yy}$ (a saddle node).

Appendix C

Validation Proof

Given a truth input–output function $Y^T(\mathbf{p}) : \Omega \to I\!R$ we define the model prediction error function, $\mathcal{E}_Y(\mathbf{p}) : \Omega \to I\!R$ as

$$\mathcal{E}_Y(\mathbf{p}) = |Y^T(\mathbf{p}) - \tilde{Y}(\mathbf{p})|, \tag{C.1}$$

where $\tilde{Y}(\mathbf{p})$ is the proposed surrogate. The function $\mathcal{E}_Y(\mathbf{p})$ need not be smooth or even continuous. We then define a function $\mathcal{Z}(x) : [0, \infty) \to [0, 1]$ which gives the measure of a subset of Ω in which $\mathcal{E}_Y(\mathbf{p})$ is strictly greater than $x \in [0, \infty)$, that is

$$\mathcal{Z}(x) = \mu(\{\mathbf{p} \in \Omega \mid \mathcal{E}_Y(\mathbf{p}) > x\}). \tag{C.2}$$

It is readily shown that the function $\mathcal{Z}(x)$ is right–continuous and non-increasing. For a plateau of the function $\mathcal{E}_Y(\mathbf{p})$, for which $\mathcal{E}_Y(\mathbf{p}) = x_c$, $\forall \mathbf{p} \in \Omega_c \subset \Omega$, the jump in $\mathcal{Z}(x)$ is given by

$$\lim_{y \searrow 0} (\mathcal{Z}(x_c - y) - \mathcal{Z}(x_c)) = \mu(\Omega_c), \tag{C.3}$$

the measure of the set for which $\mathcal{E}_Y(\mathbf{p}) = x_c$.

The model prediction error is given by

$$U_Y = \max_{j \in \{1,...,N\}} |Y^T(\mathbf{P}_j) - \tilde{Y}(\mathbf{P}_j)|, \tag{C.4}$$

where $\mathbf{P}_1, ..., \mathbf{P}_N$ are random input vectors distributed according to the probability density function $\rho(\mathbf{p}) : \Omega \to I\!R$. The random uncharacterized region is given by

$$\Upsilon_Y = \{\mathbf{p} \in \Omega \mid \mathcal{E}_Y(\mathbf{p}) > U_Y\}. \tag{C.5}$$

We then define a random variable Z which represents the measure of this random region Υ_Y,

$$Z = \mu(\Upsilon_Y = \{\mathbf{p} \in \Omega \mid \mathcal{E}_Y(\mathbf{p}) > U_Y\}) = \mathcal{Z}(U_Y). \tag{C.6}$$

We are interested in computing the cumulative distribution function for $Z, F_Z(z) = Pr\{Z \leq z\}$.

To proceed, we set, for any $z \in [0, 1]$,

$$x_z = \min_{\{x \in [0,\infty) | \mathcal{Z}(x) \leq z\}} x. \tag{C.7}$$

Since $\mathcal{Z}(x)$ is a non-increasing function, we have from (C.7) that

$$Pr\{\mathcal{Z}(U_Y) \leq z\} = Pr\{U_Y \geq x_z\}. \tag{C.8}$$

Now, let $\mathcal{D}$ be a subset of Ω given by

$$\mathcal{D} = \{\mathbf{p} \in \Omega \mid \mathcal{E}_Y(\mathbf{p}) \geq x_z\}. \tag{C.9}$$

The measure of the subset $\mathcal{D}$ is given by

$$\mu(\mathcal{D}) = \inf_{x < x_z} \mathcal{Z}(x), \tag{C.10}$$

where from (C.2), (C.7), and the properties of $\mathcal{Z}(x)$,

$$\mu(\mathcal{D}) \geq z. \tag{C.11}$$

From the definition of U_Y we have

$$\begin{aligned} Pr\{U_Y \geq x_z\} &= Pr\{\exists j \in \{1,...,N\} \mid \mathbf{P}_j \in \mathcal{D}\} \\ &= 1 - Pr\{\mathbf{P}_j \in \Omega \setminus \mathcal{D}, \forall j \in \{1,...,N\}\}, \end{aligned} \tag{C.12}$$

and it thus only remains to compute the probability that all the $\mathbf{P}_j$ are not in $\mathcal{D}$.

To proceed, we note that since the $\mathbf{P}_j$ are i.i.d., the probability that all validation points are outside the subset $\mathcal{D}$ is simply

$$\begin{aligned} Pr\{\mathbf{P}_j \in \Omega \setminus \mathcal{D},\ \forall j \in \{1,...,N\}\} &= (Pr\{\mathbf{P}_1 \in \Omega \setminus \mathcal{D}\})^N \\ &= (\mu(\Omega \setminus \mathcal{D}))^N \\ &= (1 - \mu(\mathcal{D}))^N. \end{aligned} \tag{C.13}$$

But from (C.11) we know that

$$1 - \mu(\mathcal{D}) \leq 1 - z, \tag{C.14}$$

which then gives

$$Pr\{\mathbf{P}_j \in \Omega \setminus \mathcal{D}, \forall j \in \{1,...,N\}\} \leq (1-z)^N. \tag{C.15}$$

Finally from (C.8) and (C.12), we obtain

$$F_Z(z) \geq \widehat{F}_Z(z), \tag{C.16}$$

where

$$\widehat{F}_Z(z) = 1 - (1 - z)^N. \tag{C.17}$$

Note that $F_Z(z) = \widehat{F}_Z(z)$ when $\mathcal{Z}(x)$ is continuous (which does not, of course, imply that $\mathcal{E}_Y(\mathbf{p})$ is continuous). Note also that $Pr\{Z < z\}$ with strict inequality satisfies

$$Pr\{Z < z\} \geq \widehat{F}_Z(z), \tag{C.18}$$

although $Pr\{Z < z\} \geq F_Z(z)$ is potentially false if $\mathcal{Z}(x)$ is not continuous. In the latter case, Z is not an absolutely continuous, but rather a mixed, random variable.

Lastly, we note that our proof is valid, in fact, for any positive function $\mathcal{E}_Y$, with no explicit reference to Y or $\widetilde{Y}$. We can thus directly obtain the result (C.18) for

$$\mathcal{E}_Y(\mathbf{p}) = \max\left(\frac{|\Theta^T(\mathbf{p}) - \widetilde{\Theta}(\mathbf{p})|}{g_\Theta(\mathbf{p})}, \frac{|\Pi^T(\mathbf{p}) - \widetilde{\Pi}(\mathbf{p})|}{g_\Pi(\mathbf{p})}\right), \tag{C.19}$$

or, indeed, for the maximum of any K scaled outputs.

Bibliography

[1] G. E. P. Box and N. R. Draper. *Empirical model building and response surfaces.* John Wiley & Sons, New York, 1987.

[2] J.-F. Barthelemy and R. T. Haftka. Approximation concepts for optimum structural design —a review. *Structural Optimization*, 5:129–144, 1993.

[3] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn. Design and analysis of computer experiments. *Statistical Science*, 4(4):409–435, 1989.

[4] S. Yeşilyurt and A. T. Patera. Surrogates for numerical simulations; optimization of eddy–promoter heat exchangers. *Comp. Methods Appl. Mech. Engr.*, 121:231–257, 1995.

[5] S. Yeşilyurt, C. Ghaddar, M. E. Cruz, and A. T. Patera. Bayesian–validated surrogates for noisy computer simulations; application to random media. *SIAM J. Sci. Comput.*, 17(4), 1996.

[6] J. C. Otto, M. Paraschivoiu, S. Yeşilyurt, and A. T. Patera. Bayesian–validated computer–simulation surrogates for optimization and design: error estimates and applications. *IMACS J. Appl. Numer. Math.* (to appear).

[7] S. Yeşilyurt, M. Paraschivoiu, J. C. Otto, and A. T. Patera. Computer–simulation response surfaces: a Bayesian–validated approach. In *Proc. Computational Methods in Water Resources XI*, Cancun, July 1996.

[8] N. Alexandrov, J. E. Dennis, Jr., R. M. Lewis, and V. Torczon. A trust region framework for managing the use of approximation models in optimization. In preparation.

[9] V. Pareto. *Manual of Political Economy.* The McMillan Press Ltd., London–Basingslohe, 1971. Translation of the French edition by A. S. Schwier.

[10] H. A. Eschenauer, J. Koski, and A. Osyczka. Multicriteria optimization — fundamentals and motivations. In H. A. Eschenauer, J. Koski, and A. Osyczka, editors, *Multicriteria Design Optimization: Procedures and Applications.* Springer–Verlag, Berlin, 1990.

[11] D. H. van Campen, R. Nagtegaal, and A. J. G. Schoofs. Approximation methods in structural optimization using experimental designs for multiple responses. In H. A. Eschenauer, J. Koski, and A. Osyczka, editors, *Multicriteria Design Optimization: Procedures and Applications.* Springer–Verlag, Berlin, 1990.

[12] G. E. Karniadakis, B. B. Mikic, and A. T. Patera. Minimum-dissipation transport enhancement by flow destabilization: Reynolds' analogy revisited. *J. Fluid Mech.*, 192:365–391, 1988.

[13] M. F. Schatz, R. P. Tagg, H. L. Swinney, P. Fischer, and A. T. Patera. Supercritical transition in plane channel flow with spatially periodic perturbations. *Phys. Rev. Lett.*, 66(12):1579–1582, 1991.

[14] Fluent Inc., Lebanon, New Hampshire. *Nekton Users Manual*, 1992.

[15] Y. Maday and A. T. Patera. Spectral element methods for the Navier–Stokes equations. In A. K. Noor, editor, *State–of–the–art surveys in computational mechanics.* ASME, New York, 1989.

[16] P. Fischer and A. T. Patera. Parallel spectral element solution of the Stokes problem. *J. Comput. Phys.*, 92(2):380–421, 1991.

[17] Y. Maday, A. T. Patera, and E. M. Ronquist. An operator-integration-factor splitting method for time dependent problems: Application to incompressible fluid flow. *J. Sci. Comput.*, 5(4):263–292, 1990.

[18] E. Ronquist. A domain decomposition method for elliptic boundary value problems: Application to unsteady incompressible fluid flow. In D. E. Keyes, T. F. Chan, G. A. Meurant, J. S. Scroggs, and R. G. Voigt, editors, *Proc. Int. Conf. on Domain Decomposition Methods for Partial Differential Equations, 5th*, pages 545–557, Philadelphia, 1992. SIAM.

[19] R. A. Nicolaides. Deflation of conjugate gradients with applications to boundary value problems. *SIAM J. Num. Anal.*, 51(2):355–365, 1987.

[20] J. F. M. Barthelemy and J. Sobieszczanski-Sobieski. Extrapolation of optimum design based on sensitivity derivatives. *AIAA Journal*, 21(5), 1983. Technical Note.

[21] J. F. Traub, G. W. Wasilkowski, and H. Wozniakowski. *Information–based complexity.* Academic Press, San Diego, 1988.

[22] A. M. Mood, F. A. Graybill, and D. C. Boes. *Introduction to the theory of statistics.* McGraw-Hill, New York, 1974.

[23] D. Haussler. Decision theoretic generalization of the PAC model for neural net and other learning applications. *Information and Computation*, (100):78–150, 1992.

[24] S. Yeşilyurt. PhD Thesis, 1995. Department of Nuclear Engineering, M.I.T.

[25] R. Y. Rubinstein. *Simulation and the Monte–Carlo Method.* John Wiley & Sons, New York, 1981.

[26] G. Strang and G. J. Fix. *An analysis of the finite element method.* Prentice–Hall, Englewood Cliffs, New Jersey, 1973.

[27] A. Ainsworth and J. T. Oden. *A posteriori* error estimation in finite element analysis. Technical Report 96-19, TICAM, 1996.

[28] J. C. Otto, M. Paraschivoiu, S. Yeşilyurt, and A. T. Patera. Bayesian–validated computer–simulation surrogates for optimization and design. In N. Alexandrov, editor, *Proc. ICASE Workshop on Multidisciplinary Design Optimization.* SIAM, Philadelphia, March 1995 (Hampton, Virginia).

[29] J. C. Otto, M. Paraschivoiu, S. Yeşilyurt, and A. T. Patera. Computer–simulation surrogates for optimization: application to trapezoidal ducts and axisymmetric bodies. In *Proc. of the Forum on CFD for Design and Optimization*, San Francisco, CA, November 1995. ASME International Mechanical Engineering Conference and Exposition.

[30] H. A. David. *Order statistics.* John Wiley & Sons, New York, 1981.

[31] J. Otto. PhD Thesis. Department of Aeronautics and Astronautics, M.I.T., in progress.

[32] F. Riesz and B. Sz.-Nagy. *Functional Analysis.* Frederick Ungar Publishing Co., 1955.

[33] M. Avriel. *Nonlinear Programming: Analysis and Methods.* Prentice-Hall, Englewood Cliffs, New Jersey, 1976.

[34] D. Lee and H. Wozniakowski. Testing nonlinear operators. *Numerical Algorithms*, 9(3-4):319–342, 1995.

[35] F.Y. Edgeworth. *Mathematical Psychics.* P. Keagan, London, England, 1881.

[36] J. Juran. *Quality control handbook.* McGraw-Hill, New York, 1962.

[37] E. Borel. La theorie du jeu et les equations integrals a noyeau symmetrique gauche. *Comptes Rendus de L'Academie des Sciences*, 173:1304–1308, 1921.

[38] W. Stadler. A survey of multicretiria optimization, or the vector maximum problem, 1776-1960, part i. *Journal of Optimization Theory and Applications*, 29:1–52, 1979.

[39] W. Stadler. *Preference Optimality and Application of Pareto Optimality.* Springer, Berlin, Heidelberg, New York, 1975. In: Marzollo/Leitman (eds.): Multicriterion Decision Making. CISM Courses and Lectures.

[40] W. Stadler. Natural structural shapes (the static case). *Quarterly J. Mech. Appl. Math.*, 31:169–217, 1975.

[41] W. Stadler. *Applications of Multicriteria Optimization in Engineering and the Sciences (A survey). MCDM–Past Decade and Future Trends.* JAI Press, Greenwich, Connecticut, 1984.

[42] W. Stadler. *Multicriteria Optimization in Engineering and in the Sciences.* Plenum Press, New York, 1988.

[43] H. M. Markowitz. Portfolio selection. *Journal of Finance*, 7:77–91, 1952.

[44] H. P. Benson. Optimization over the efficient set. *JOMAA*, 98(3):689–707, March 1996.

[45] A. Osyczka. *Multicriterion Optimization in Engineering with Fortran programs.* Ellis Horwood series in mechanical engineering. Halsted Press, New York, 1984.

[46] M. Kambourides, S. Yeşilyurt, and A. T. Patera. Nonparametric–validated computer–simulation surrogates: A pareto formulation. Submitted to *International Journal for Numerical Methods in Engineering.*

[47] L. Sirovich. Turbulence and the dynamics of coherent structures. *Quart. Appl. Math.*, 14(3):561–588, Oct. 1987.

[48] J. L. Lumley. Coherent structures in turbulence. In R. E. Meyer, editor, *Transition to Turbulence*, pages 215–242. Academic Press, New York, 1981.