Development of a Post-Optimality Analysis Algorithm for Optimal Control Problems

Von der Fakultät Luft- und Raumfahrttechnik und Geodäsie
der Universität Stuttgart zur Erlangung der Würde eines
Doktor-Ingenieur (Dr.-Ing.) genehmigte Abhandlung

Vorgelegt von

Sven Oliver Erb

aus Waldshut-Tiengen

Hauptberichter: Prof. Klaus Hinrich Well, Ph.D.
Mitberichter: Prof. Dr. rer. nat. Ernst W. Messerschmid

Tag der mündlichen Prüfung: 12. Januar 2007

Institut für Flugmechanik und Flugregelung
der Universität Stuttgart

2007

Contents

Zusammenfassung

Die Beschreibung von technischen Fragestellungen in Form von Optimalsteuerungsproblemen wird zunehmend zu einem populären Werkzeug. Dabei zeigt sich, dass die Aufbereitung und Auswertung der gefundenen optimalen Lösung eine Schlüsselkompetenz darstellt. Die vorgelegte Arbeit setzt hier an. Mittels post-optimaler Analyse wird eine zuvor ermittelte optimale Lösung so weiterverarbeitet, dass der Anwender auf effizientem Wege zusätzliche Daten über Optimalitätsverhalten, Einflussfaktoren und, in weiteren Schritten, über benachbarte Lösungsräume erhält.

Der entwicklte Algorithmus nutzt Verfahren aus der Nichtlinearen Optimierung zur Transkription des Optimalsteuerungsproblems. Dieser Vorgang ist insofern generisch gehalten, als verschiedene Parametrisierungsschemata der Kollokation, wie auch der Mehrzielverfahren implementiert sind. Eine darauf aufbauende parametrisierte Sensitivitätsanalyse dient als Grundlage für die eingehende post-optimale Analyse. Für die Untersuchungen erster und zweiter Ordnung wird zur Effizienzsteigerung und Erhöhung der Transparenz, soweit möglich, auf bereits vorhandene Basisdaten zurückgegriffen. Nicht verfügbare Daten, wie etwa Hesse-Matrix, werden neu berechnet.

Der entwickelte Algorithmus identifiziert Sensitivitäten der Kostenfunktion innerhalb der bestehenden Problembeschreibung und erlaubt darüber hinaus Aussagen zur Beeinflussung des Lösungsraumes durch die Variation nicht optimierbarer Parameter bzw. Gleichungs- oder auch Ungleichungsbeschränkungen.
Letztere Fähigkeit schließt die Prädiktion des Einflusses von finiten Variationen auf die Zusammensetzung des Set der aktiven Beschränkungen ein. Daraus wird dann unter Einschluß von Optimalitätsaspekten und Angabe über den Vertrauensbereich, die Lösung benachbarter Entwurfsräume vorhergesagt.
Damit ist die post-optimale Analyse nicht nur zur besseren Bewertung bereits ermittelter Lösungen geeignet, sondern bietet auch eine effiziente Alternative zur konsekutiven Optimierung, variierter Problemstellungen.

Um den Nutzen der post-optimalen Analyse für umfassende, moderne Anwendungen zu demonstrieren, wird der neue Algorithmus auf zwei typische Probleme aus der Raumfahrt angewendet. Es handelt sich um die optimale Wiedereintrittsbahn des wiederverwendbaren Transportfahrzeuges HOPPER, sowie um optimale Aufstiegsbahnen einer Ariane 5 bei unterschiedlichen Nutzlast-Konfigurationen.
Dabei werden Robustheit und Güte der bisherigen Entwürfe bewertet und das Verbesserungspotential quantifiziert. Außerdem wird demonstriert, welchen zusätzlichen Nutzen der neu entwickelte Algorithmus zur post-optimalen Analyse im Bereich von Optimalsteurungsproblemen hat.

Summary

The description of technical problems in the form of optimal control problems is becoming increasingly popular. And the assessment and evaluation of the obtained optimal solutions is developing into a key competence. This dissertation departs from such an optimal solution. By means of post-optimality analysis an earlier obtained result is processed in order to allow a user to efficiently produce information about optimality criteria, the main factors that influence the solution and, in consecutive steps also about neighboring solution spaces.

The developed algorithm exploits methods from the domain of nonlinear optimization for the transcription of the optimal control problem. This procedure is made generic in the sense that different parameterization schemes are incorporated, ranging from collocation to multiple shooting. The basis of consecutive post-optimality analysis is a parameterized sensitivity analysis. First and second order evaluation takes efficiently advantage of data that has already been computed during the process of the prior optimization. Data that is not readily available, like the Hesse matrix, is newly computed.

The developed algorithm identifies sensitivities of the cost function within the existing problem description, but also permits to investigate the influence of variations in non-optimizable parameters, equality or inequality constraints on the solution space. The latter includes the prediction of changes in the active set of constraints due to finite variations. Under full consideration of optimality conditions and the trust radius, the solution of neighboring problems is predicted. Thus, post-optimality analysis is not only suitable for evaluation of already computed optimal solutions. It also provides an efficient alternative to consecutive optimization of varied problem descriptions.

In order to demonstrate the benefits of post-optimality analysis for modern comprehensive problems, the algorithm is applied to two typical aerospace problems. These are an optimal entry of the reusable launch vehicle HOPPER, and ascent trajectories of Ariane 5 with several payload configurations. Robustness and cost quality of the current design is evaluated and potential for improvement quantified. Further, the added value is demonstrated that the newly developed algorithm for post-optimality analysis provides in the area of optimal control problems.

Chapter 1

Introduction

He who has knowledge does not predict. He who predicts does not have knowledge.

Chinese teaching

This statement was supposedly made by ancient Chinese philosopher Lao-Tse more than two thousand years ago. It is a popular phrase, variously quoted whenever people are encouraged to question the professionalism and qualification of self-promoted leaders and specialists. It essentially distinguishes between knowledge and prediction and, thus, suggests that there is an either-or. A proper understanding of cause and effect guarantees knowledge while a lack of this understanding triggers prediction. The connotation of prediction is undoubtedly negative, since knowledge is generally considered to be a positive quality.

Today, in research and engineering it becomes more and more important to combine knowledge and prediction in order to improve model understanding and accelerate product development.

Post-optimality analysis (POA) is seen as such a hybrid. It is a technique to study the behavior of an earlier obtained problem solution and helps to gather information about its sensitivity. At the same time, it also suggests an interpretation of the state space around the solution providing stability information. In other words, it serves to predict the solution of perturbed problems without need for recomputation.

Following, a definition is given for optimal control problems to familiarize the reader with this particular class of problems, which is in the focus of our efforts

to develop an expertise in post-optimality analysis. The role and importance of Sensitivity Analysis (SA) is briefly addressed in sec. 1.2 together with a synopsis of common methods. Then, the basic aspects of post-optimality analysis and prior contributions by other authors are given.
Considerations on the requirements for post-optimality analysis and applicable methods conclude the introduction and define the strategy.

1.1 The Optimal Control Problem

Problem formulations of dynamic systems with optimizable parameters and controls have a long lasting tradition in aerospace engineering. This class of Optimal Control Problems (OCP), has found its entry into flight maneuver optimization [42], launcher and return vehicle trajectory optimization [35], [2], [16], [63], satellite transfer [23], [18] and interplanetary travel [24]. The community of practitioners has also spread into other branches, like the automotive sector [20] and medicine [26], and has created a growing interest in solution and analysis methods.
Model complexity and the need for efficiency have raised the interest in methods to compute optimal solutions. They are frequently the only chance to bring about improvements towards enhanced performance.

To fully immerse into the topic, the proper mathematical formulation of optimal control problems is given in the following section. Afterwards, the value of their solution is assessed with regard to practical usefulness.

1.1.1 Mathematical Description

The optimal control problem is detailed in a large number of publications [15], [31], [8]. In order to provide a concise nomenclature and to allow fundamental understanding of the later chapters, a basic description is given here.
The control problem for which an optimal solution shall be computed, is as follows: minimize the cost functional

$$min\ J(x,u,p,t) = \Phi(x_f,p,t_f) + \int_{t_0}^{t_f} L(x,u,p,t)dt. \tag{1.1.1}$$

The objective function is stated in Bolza format with Φ being the Mayer term and a Lagrange term with the integrand L. The vector $\mathbf{u} = \mathbf{u}(t)$ represents the time-variant control vector and t represents the independent variable. The state vector of the system is given as $\mathbf{x} = \mathbf{x}(t)$ and has the dynamics

$$\dot{\mathbf{x}} = f(x(t), u(t), p, t) \tag{1.1.2}$$

It is often convenient to assign time-invariant parameters $\mathbf{p}$, which describe certain system properties. Commonly, they are design parameters or related qualities of the model, and mission, respectively. Their value is optimizable, but does not change over time.

Additional conditions for the system are stated as path constraints. These can be defined as equality constraints

$$\mathbf{h}(x(t), u(t), p, t) = 0 \tag{1.1.3}$$

or as inequality constraints

$$\mathbf{g}(x(t), u(t), p, t) \geq 0. \tag{1.1.4}$$

The same holds for boundary constraints which can be equality

$$\boldsymbol{\Psi}_t = \boldsymbol{\Psi}_t \left(x(t), u(t), p, t\right) = 0 \tag{1.1.5}$$

or inequality constraints

$$\boldsymbol{\Psi}_t = \boldsymbol{\Psi}_t \left(x(t), u(t), p, t\right) \geq 0 \tag{1.1.6}$$

under the condition that either $t = t_0$ or $t = t_f$.

The classical way to solve such an optimal control problem are indirect methods. They are based on the calculus of variations.

The first formulation of first order necessary conditions for optimal control problems was published by Euler and Lagrange in 1744. The Euler–Lagrange equations compose the fundament for the solution of this kind of mathematical problem. In the time since, scientists have refined the formulations and have extended their use. The introduction of the Hamilton function in 1834/35, for instance, was a major contribution to improve the analytic structure of the conditions.

The first order necessary conditions for a problem with no path constraints, with terminal equality constraints and mixed initial constraints can be found in eqs. 1.1.7.

$$\begin{aligned}
&\text{Hamiltonian:} && H = L + \lambda^T f \\
&\text{Dynamics:} && \dot{\mathbf{x}} = f(x,u,t) = \tfrac{\partial H}{\partial \lambda} \\
&\text{Adjoint differential equations:} && \dot{\lambda} = -\tfrac{\partial L}{\partial \mathbf{x}} - \left(\tfrac{\partial f}{\partial \mathbf{x}}\right)^T \lambda = -\tfrac{\partial H}{\partial \mathbf{x}} \\
&\text{Optimality condition:} && 0 = \tfrac{\partial H}{\partial \mathbf{u}} = \tfrac{\partial L}{\partial \mathbf{u}} + \left(\tfrac{\partial f}{\partial \mathbf{u}}\right)^T \lambda \\
&\text{Initial conditions:} && \mathbf{x}(t_o) \text{ given or } \lambda(t_0) = 0 \\
&\text{Terminal constraints:} && \mathbf{\Psi}_f = \mathbf{\Psi}_f(x(t_f), t_f) = 0 \\
&\text{Transversality conditions:} && \lambda_f = \left[\tfrac{\partial \Phi}{\partial \mathbf{x}} + \left(\tfrac{\partial \Psi_f}{\partial \mathbf{x}}\right)^T \nu\right]_{t=t_f} \\
&\text{Transvers. for optimizable } t_f\text{:} && \Omega = \left[\tfrac{\partial \Phi}{\partial t} + \nu^T \tfrac{\partial \Psi_f}{\partial t} + H\right]_{t=t_f} = 0
\end{aligned} \qquad (1.1.7)$$

The Euler–Lagrange equations are only of first order and do not formulate sufficient conditions. Therefore extensive research has been undertaken to complete and extend them. The Legendre–Clebsch condition, demanding

$$\frac{\partial^2 H}{\partial \mathbf{u}^2} \geq 0 \qquad (1.1.8)$$

is an example for a necessary condition of second order.

It was the Russian mathematician Pontryagin in 1954, who extended the optimal control theory to cases with constrained variables. This was in so far an important contribution as it enabled the optimization of problems with path constraints, which are a common element in engineering problems. A detailed discussion of additional conditions can be found in [15].

Indirect methods have the potential to provide closed solutions for OCPs. They work with exact analytical terms and solve the problem via an intermediate elimination of the control and later back-calculation of the optimal control history. This technique has also been the name giver for the class of indirect methods.
The very attractive features of indirect methods are paid for by enormous mathematical overhead. Application of indirect methods requires a deep understanding

of the mathematical problem structure and extensive knowledge about its solution space. The structure of the equations describing the optimality conditions is very complex. And even small changes in the problem outline can trigger major modifications of, for instance, the Hamiltonian function.
The same holds for the adjoint, or costate, variables [11]. They do not have physical meaning which makes their estimation not at all intuitive and overly time consuming. However, an accurate estimate is essential, since the mathematical problem description is very sensitive to changes in the costate values. And convergence to an optimal solution requires a qualitatively good initial guess.

An alternative to indirect methods has appeared with the emergence of digital computers in the mid of the 20th century -the so called direct methods [46], [39], [9]. As the name suggests, they are straightforward techniques to calculate the optimal solution. The concept is to parameterize the solution space and to solve a constrained nonlinear programming problem. The algorithm ensures compliance with all constraints while in an iterative process closing in on the optimal solution.
The direct methods are numerical and, thus, not as elegant and analytically exact as indirect methods. But they exhibit a series of advantages, which make them the first choice more and more often in practical applications. It is the flexibility, adaptability and robustness that gives these methods a larger convergence radius and makes them well appropriate for users, who do not have an in-depth knowledge of optimization theory.

1.1.2 Understanding an Optimal Solution

It is one thing to describe an optimal control problem and obtain a solution. It is another to fully understand and exploit the found solution. There are numerous aspects that determine the usefulness of the results that an optimization algorithm delivers.

First of all, there is the question about the optimality of the solution with respect to a certain criterion. Is the solution strictly optimal, or is the cost function gradient only gently inclined?
Essentially, optimization algorithms terminate providing an optimal design point with a certain performance number, but lack a comprehensive survey of the design point sensitivity.
Unfortunately, the exact solution is purely theoretical in the eyes of practitioners, since the problem formulation generally describes a simplified model of reality. Hence, a transfer of the results is only reasonable when the behavior of the model is known and understood. Otherwise, reduction of main features of the problem result in erroneous solutions, from which faulty conclusions are drawn.

This motivates the assessment of perturbations in auxiliary design parameters, which could compensate for model short-comings. These parameters typically describe model properties that are held constant during the optimization, even though they are naturally not constrained to a particular value.
This leads to the identification of critical components, the evaluation of uncertainty sources and the question, what effect certain parameters have on the optimal solution.

The matter can be summarized in the term *Sensitivity*. The key to enhanced problem understanding is knowledge about the sensitivity of the optimum solution.

1.2 Sensitivity as a Means of Analysis

Sensitivity analysis has become a main competence for the modeling and analysis of complex systems in general. Blackwell [12] has congregated its meaning in the following definition:

> **Sensitivity analysis is defined as the study how variations in input parameters of a computational model cause variations in output. [...] A measure of this sensitivity is termed the sensitivity coefficient and is (mathematically) defined as a partial derivative of the output variable with respect to the parameter of interest.**

This gives us the mathematical expression

$$\mathcal{S}_{ij} = \frac{\partial y_i}{\partial x_j}. \tag{1.2.9}$$

for the sensitivity coefficient $\mathcal{S}_{ij}$. The scalar variable y_i represents an output of the investigated system. It commonly characterizes the objective and contains key performance properties of the dynamic system.
The scalar variable x_j constitutes an input to the system. Most often it is a design parameter, a model parameter, or a parameter defining a condition.

Another expression which is commonly used for $\mathcal{S}_{ij}$ is sensitivity derivative. It has been introduced by Sobieski [65], [58], [64].

The dependencies illustrated by the sensitivity coefficients provide useful information about the behavior and character of the problem under investigation. Hence, the coefficients are valuable for analysis and can be processed for various tasks. Bose et al. [13] performs sensitivity analysis to identify uncertainty risk and categorizes as follows:

1. Structural Uncertainty
   All mathematical models work with assumptions and simplifications to represent physical phenomena. These pose the risk of not being realistic.

2. Parametric Uncertainty
   This type of uncertainty arises for uncertainties in the model parameter estimates. This happens frequently for parameters of dynamic systems which can not be measured explicitly, but have to be guessed.

3. Stochastic Uncertainty
   Natural fluctuations can cause this kind of uncertainty. A common example are atmospheric anomalies with stochastic behavior.

There is a multitude of scientific applications which go under the title of sensitivity analysis and document the very heterogeneous perception of its usefulness. Empirical methodologies can be found in environmental model analysis [50], labor market evaluation [62], or in chemistry [69]. It can be used to compute safety or probability margins [27] by widening the parameter range or provide gradient information for an optimization algorithm [55].

The rising interest in sensitivity computation in recent years has led to the development of a number of different methods. In general the selection of the most suitable method is dominated by the structure of the model and its accessibility. Therefore model properties and computational interests can be taken to broadly classify the various analysis methods.

If the model is completely unknown and the model equations are not accessible or if dependencies shall be scanned for a wide variational range, then sampling methods promise to show best performance. The relationship between input and output parameters in the state space is established empirically via model runs at sets of sampling points [51].
In this context, simulation campaigns are a widely used means for uncertainty and sensitivity analysis [48]. Telaar [68] has worked with them to identify sensitivities of a reentry vehicle. And Bose et al. [13] applied the concept to compute sensitivities in the thermochemical model for a Titan atmospheric entry.
The selection scheme for the sampling sets distinguishes the various sampling methods from each other and defines the computational expenses. This ranges from Monte Carlo-like approaches with almost random distribution, to Latin Hypercube

Sampling [66] with equal probability segments. Also Response Surface methods are in use. They process the obtained sampling information towards the definition of a secondary model with reduced parameter number [72]. The surrogate model is convenient for trend analysis and allows rapid, but also inaccurate sensitivity derivation.

Other methods are directly focused on producing derivatives. There are the analytic methods, which rely on the accessibility and differentiability of the model equations. One such technique compiles the Forward Sensitivity Equations. Another is the Reverse Adjoint Equations method. As the name suggests, it analyzes the origin of an anomaly by means of reverse signal flow [36] and, thus, allows an identification of perturbation sources. In [55] it is shown, how this method can be used to provide sensitivity data for aero-structural optimization.

All these analytic methods provide exact gradients, but require an enormous mathematical overhead. Particularly in engineering applications it is frequently the case that spreadsheets and switching functions are part of the model and make analytical techniques a prohibitively expensive or even impossible task.
Then, numerical methods become a convenient alternative. Finite differences are generally the prime choice for numerical differentiation. But Martins et al. [56] has also successfully tested the method of Complex Steps for sensitivity analysis purposes. The accuracy is superior. However, complex variation of the model requires a comprehensive complex algebra environment.
In any case, the sensitivity coefficients are computed for a certain reference point and therefore local.

Another method is Parameterized Sensitivity Analysis (PSA). The fundamental idea behind this is to reduce the model size by parameterization, while retaining its complexity. It is often possible to make reasonable assumptions for a model and establish parametric relationships as a substitute to function dependencies, for instance, through use of polynomial fitting. The infinite number of state propagators is approximated with a limited number of parameters, which subsequently constitute the parameter space of the new problem description.
The computation of sensitivities is turned into the computation of parameter dependencies. And the chain of such dependencies determines the impact of an input variation on a specific output. Each parameter which correlates to the input of interest potentially stimulates the output. Fiacco [28] and Sobieski [65] have provided valuable contributions for the advancement of PSA.