ISSN 2186-7437

NII Shonan Meeting Report

No. 2018-9

Piecewise smooth system and optimization with piecewise linearization via algorithmic differentiation

Andreas Griewank
Andrea Walther
Siegfried Rump
Koichi Kubota

June 25–28, 2018

National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-Ku, Tokyo, Japan

Piecewise smooth system and optimization with piecewise linearization via algorithmic differentiation

Organizers:
Andreas Griewank (School of Mathematical Science and Information Technology, Yachay Tech, Ecuador)
Andrea Walther (Paderborn University, Germany)
Siegfried Rump (Hamburg University of Technology, Germany)
Koichi Kubota (Chuo University, Japan)

June 25–28, 2018

Functions with parameters that appear frequently as mathematical models in science, engineering and social phenomena often show a non-differentiable behavior, for example due to upper limits for temperatures, pressure, and the like. However, we usually simplify them to smooth and differentiable models in those fields because we analyze the models mathematically with well established methods suitable for smooth functions and need derivatives of the considered function with respect to the parameters to optimize the underlying system. Given the recent progress of algorithmic differentiation (AD), it is now possible to implement software systems that compute the sub-gradient or Clarke generalized Jacobian at non-differentiable points of a certain class of non-smooth function called frequently "piecewise" smooth (PS) functions. In the simplest case, these functions are even piecewise linear (PL). Using the extended AD technology, we could adopt more general models that are described as large programs with some non-differentiable intrinsic functions, e.g., $\min(x, y)$, $\max(x, y)$ and $\text{abs}(x)$, in nonlinear equation solving, integration of differential equations and optimization tasks.

The meeting provides a unique possibility to discuss, between worldwide leading researchers, the recent progress within the above mentioned related fields from the view point of practical non-differentiable techniques. This discussion should initiate the development of new methods to optimize PS and PL functions, to solve systems of PL and PS nonlinear equations, and to integrate differential equations with PS right-hand sides. Moreover, since the recent revivals of artificial intelligence and machine learning are remarkable, it is widely re-recognized that their fundamental techniques are mathematical optimization of PS and PL functions as nonlinear functions. Furthermore, it was also well known that the back-propagation algorithm by D.E. Rumelhart was an instance of the reverse mode of algorithmic differentiation. Hence, we also expect influence of the meeting in this direction of science, since also there non-smoothness plays an important role.

The starting point for the targeted non-smooth nonlinear optimization is the local linearization by differentiation as the mainstay of mathematical analysis and scientific computing. The local linearization of non-smooth continuous functions corresponds to the piecewise linearization by sub-gradients, the Clarke generalized Jacobian, and similar concepts. Using an adapted algorithmic differentiation, the piecewise linearizations of PS functions can be derived algorithmically, i.e., without additional effort. For their analysis and further usage, the abs-normal form (ANF) plays an important role. It was developed by A. Griewank and A. Walther, who are two of the organizers. This formalization of the local piecewise linear model will form a common basis throughout the meeting for the discussion of further developments. For example, it has been used to derive the first and the second order optimality conditions for piecewise smooth objective functions, assuring the use of the piecewise smooth models for describing more complex phenomena.

Overview of Talks

Abs-Linearization for Piecewise Smooth Optimization

Prof. Andrea Walther, Universität Paderborn, Germany

The main theme of this seminar, piecewise smooth systems and optimization is explained as well as the technique called "Abs-Linearization". Giving the definition of piecewise functions, their examples that appear in the real world problems are explained: (i) Exact L_1 penalty function, (ii) Fuzzy pattern tree, (iii) optimality conditions.

Comparing to current black-box approach, a new approach called a gray-box approach is introduced for their optimization. Abs-Linearization that is a key technique and plays important role in the new approach gives a smooth representation of Lipschitz-function defined by abs, max, min operations with algorithmic differentiation.

The new approach named SALOP (Successive Abs-Linear OPtimization with a proximal term) is introduced and its performance of convergence has been proved as well as numerical experiments.

Dealing with Optimization Problems Constrained by Non-linear Piecewise Smooth PDEs

Ms. Olga Ebel, Paderborn University, Germany

We investigate the problem of non-smooth PDE constrained optimization, where all non-differentiabilities are assumed to be given by non-smooth Lipschitz continuous functions such as abs(), min() and max(). These kinds of optimization problems arise in many applications and their efficient as well as robust solution requires numerical simulation combined with specific optimization algorithms. We formally define a new concept for solving optimal control problems constrained by non-smooth semilinear elliptic PDEs where standard optimal control techniques for obtaining first-order optimal points cannot be applied.

The key idea of the optimization method under consideration is locating stationary points by appropriate decomposition of the original problem into smooth branch problems. Subsequent successive exploitation of the corresponding adjoint variables leads to the next branch and thus to successive reduction of the function value. Numerical results illustrate the proposed approach and the applicability of the method is demonstrated by considering several model problems.

Parametric and Non-parametric Piecewise Linear Models and their Optimization

Prof. Xiaolin Huang, Shanghai Jiao Tong University, China

Optimization algorithms are designed for specific formulations, which are also the basis of modeling piecewise linear systems. Since the proposal of compact representation, the parametric models of continuous piecewise linear functions have been insightfully studied and some models with full representation capability, including our contributions, have been established. Also, we proposed non-parametric piecewise linear models by designing specific kernels. The recent progress of machine learning makes both parametric and non-parametric piecewise linear models promising to describe complicated systems. Therefore, it becomes more important to investigate optimization methods based on the learned functions. It is also desirable to develop efficient optimization method to train those piecewise linear models for machine learning.

Applying PL optimization to a data science problem

Dr. Jean Utke, Allstate Insurance Company, USA

Generalized linear models with non-smootheness are adopted in a data science problem. The algorithm for finding optimal solution of the generalized linear models is explained with high density case and low density case.

Non-Smooth Geometric Inverse Problems

Dr. Stephan Schmidt, Universität Würzburg, Germany

The primary concern of the presentation is geometric inverse problems governed by hyperbolic partial differential equations, meaning we are interested in reconstructing geometric objects such that they reproduce a measured echo of a scanning wave. There are a wide applications for problems of this type, including CFD, computational acoustics, Electrodynamics and mathematical imaging and each will be considered in the presentation.

We also study approaches that are robust with respect to non-smoothness, which arises naturally when objects with kinks are to be reconstructed. To this end, we consider both primal and dual optimization strategies for Total Variation denoising of surfaces. Either formulation requires non-standard finite element spaces such as the Raviart-Thomas space on surfaces or the DG0 space on the mesh skeleton. The presentation concludes with numerical examples where FEM solvers are interfaced with 3D scanners to conduct denoising of real world objects and novel edge preserving mesh denoising techniques.

Non-smooth tangent differentiation experiments with Source-Transformation AD tools (ALF with source-transformation AD)

Dr. Laurent Hascoët, INRIA, France

The research of Andreas Griewank, Andrea Walther, and their students has shown that Algorithmic Differentiation can be used to derive tangent models that cope with a certain class of non-smoothness, through the use of the so-called Abs-linear form (ALF). These tangent models incorporate some knowledge of the nearby discontinuities of the derivatives. These models seem to bring some additional power to processes that use tangent approximations, such as simulation, optimization, or solution of differential equations. The mechanics to derive these special tangent models may seem at first sightly exotic and remote from standard tangent linear Algorithmic Differentiation. However, successful experiments with Adol-C have shown that tangent AD may be adapted to produce these ALF models. Together with Krishna Narayanan and following suggestion from Torsten Bosse, we recently tried a similar adaption on Source-Transformation AD tools. It appears that very little development is needed in the AD-tool, be it OpenAD or Tapenade. Specifically for Tapenade, it appears that no development at all is needed in the tool itself. Any end-user can already produce ALF tangent without needing any access to the tool source. We will detail the required work, which amounts essentially to hand-writing, once and for all, a customized derivative of the absolute-value function (ABS). This is currently less than 40 lines of code. Building the ALF of a given program introduces one new variable per run-time execution of the ABS function. As the number of rows and columns of the constructed extended Jacobian are both more or less equal to the number of variables, this extended Jacobian may end up using unreasonably large memory space for large codes. To overcome this limitation of the approach, we would like to discuss the possibility of finding at run-time the "important" ABS calls that deserve this treatment, and those that don't. We believe we can base this decision on a notion of distance to the kink induced by this ABS call, using ideas from the PhD thesis of Mauricio Araya.

CppAD's Abs-normal Representation

Prof. Bradley Bell, University of Washington, USA

The algorithmic differentiation tool named CppAD is introduced. Its features for non-smooth function were provided precisely as well as examples. An algorithm for minimization of functions that contain absolute operations is given by using the tool.

Applying automatic functorial substitutions to simplify programming
(Building on predefined real tuple operations and automatic functorial substitutions to simplify programming)

Prof. Fritz Mayer-Lindenberg, Technische Universität Hamburg, Germany

The notion of functorial substitution is for the class of transformations of colored graphs that result from mapping the set of colors to a set of alternative colors and changing the color of each node accordingly. It is applied to algorithms modeled as data flow graphs colored by the operations to be performed. The substitution transforms an algorithm for some function to one for a different function even operating on different sets. Colored data flow graphs are derived by compilers as the intermediate code of programs, and applying a functorial substitution can be implemented as an automatic operation of the compiler e.g. for code generation or during interpretation of the intermediate code. Several kinds of substitutions adhere to this scheme particularly for algorithms on the reals. First, there is the substitution of operations on the real numbers to the encoded operations executed on some machine. It remains implicit for languages not distinguishing the operations from their encoded versions but becomes explicit in a language for mathematical algorithms. Similar substitutions are for high-level operations on real functions on a continuous domain once they are sampled on a finite set of points. The second example is for a language providing predefined real vector and polynomial operations. Replacing the real add and multiply operations by complex ones extends to the operations defined in terms of them and leads to predefined complex vector and polynomial operations as well. In the same way, basic real arithmetics can be replaced by arithmetics for pairs of real numbers representing intervals to deal with imprecision. Then substitution extends real algorithms to algorithms on intervals. Finally, mapping differentiable functions to functions also mapping a tangent vector to its image under the derivative is the well-known method called automatic differentiation. It is truly functorial and extends an algorithm for a function to one for its tangential map. It is remarked that the same functoriality holds for higher order approximations than the linear one and allows for an automatic higher order differentiation which can be used for numerically solving certain differential equations. Automatic functorial substitutions can be deeply integrated into a language such that e.g. after the definition of a function through a program calls to its tangential and higher order approximations become allowed without having to implement an extra data type first.

Introduction to Algebraic and Computational Geometry

Prof. Francesc Anton Castro, Universidad Yachay Tech, Ecuador

An overview of the algebraic and computational algebraic geometry is introduced, after magma, monoid, group, Abelian group, unitary ring, field, ideal and topological space are briefly reviewed. The main relationships between Algebra and Geometry are introduced together with the main historical result of Algebraic Geometry: Hilbert's Nullstellensatz. Indicating piecewise linear set as semi-algebraic sets and showing the algebraic independence of a variable and its absolute value, considering them as 2 different variables, a result was provided that the size of dimension of piecewise linear sets is $2n-2$ for the homogeneous case.

Solution Techniques for Constrained Shape Optimization Problems in Shape Spaces

Dr. Kathrin Welker, Universität Trier, Germany

Shape optimization problems arise frequently in technological processes which are modelled in the form of partial differential equations (PDEs) or variational inequalities (VIs). In many practical circumstances, the shape under investigation is parameterized by finitely many parameters, which on the one hand allows the application of standard optimization approaches, but on the other hand limits the space of reachable shapes unnecessarily. In this talk, the theory of shape optimization is connected to the differential-geometric structure of shape spaces. In particular, efficient algorithms in terms of shape spaces and the resulting framework from infinite dimensional Riemannian geometry are presented. Moreover, VI constrained shape optimization problems are treated from an analytical and numerical point of view in order to formulate approaches aiming at semi-smooth Newton methods on shape vector bundles. Shape optimization problems constrained by VIs are very challenging because of the necessity to operate in inherently non-linear and non-convex shape spaces. In classical VIs, there is no explicit dependence on the domain, which adds an unavoidable source of non-linearity and non-convexity due to the non-linear and non-convex nature of shape spaces.

Lexicographic Derivatives

Prof. Paul Barton, MIT, USA

Nesterov's lexicographic derivatives is introduced and explanation of its advantage that is automatic computability is provided with comparison to computation of Clarke generalized derivatives that are not easily computed. Three types of sensitivity analysis with lexicographic derivatives are introduced as follows: (1) Sensitivity analysis with lexicographic derivatives of nonsmooth implicit functions is explained with a theorem of L-smooth implicit function is provided; (2) Sensitivity analysis of nonsmooth differential-algebraic equations and nonsmooth ordinary differential equations is provided; (3) Sensitivity analysis of nonlinear optimization problems with lexicographic derivatives is explained with many systems as well as nonsmooth KKT equation system.

Introduction to INTLAB

Prof. Siegfried M. Rump, Technische UniversitÄ̈Hamburg, Germany

Summary of the features of INTLAB is explained by the original developer himself. INTLAB is a very popular tool for interval computation with Matlab and Octave. Many settled and new features are explained with adequate examples for indicating the performance and quality of the tool. At the meeting, Algorithmic Differentiation in INTLAB was extended by min and max to attack non-smooth functions. It will be available in the next version.

How Interval Measurement Uncertainty Affects the Results of Data Processing: A Calculus-Based Approach to Computing the Range of a Box

Prof. Vladik Kreinovich, University of Texas at El Paso, USA

In many practical applications, we are interested in the values of the quantities $y_1, \ldots, y_m$ which are difficult (or even impossible) to measure directly. A natural idea to estimate these values is to find easier-to-measure related quantities $x_1, \ldots, x_n$ and to use the known relation to estimate the desired values y_j. Measurements come with uncertainty, and often, the only thing we know about the actual value of each auxiliary quantity x_i is that it belongs to the interval $[\underline{x}_i, \overline{x}_i] = [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$, where $\widetilde{x}_i$ is the measurement result, and Δ_i is the upper bound on the absolute value of the measurement error $\widetilde{x}_i - x_i$. In such situations, instead of a single value of a tuple $y = (y_1, \ldots, y_m)$, we have a range of possible values. In this talk, we provide calculus-based algorithms for computing this range.

A Recursive Recomputation Approach to Smoothing in Nonlinear and Bayesian State-Space Modeling
Chubanovs new polynomial-time algorithm for linear programming and extensions

Prof. Takashi Tsuchiya, National Graduate Institute for Policy Studies, Japan

In this talk, we presented two topics. The first topic was an application of check-pointing to smoothing in particle filter and the second topic was extensions of Chubanov's new polynomial-time.

Check-pointing is a technique proposed by Griewank to save memory required in a generic forward-backward computational procedure. If we apply the technique to particle filter smoothing, we can reduce the required memory from $\mathrm{O}(TM)$ to $\mathrm{O}(M \log T)$ at the cost of $\mathrm{O}(\log T)$ times of filtering (instead of once), where T is the length of the time series and M is the number of particles. We explained an easy implementation of this procedure using binary number representation. We also showed a smoothing result by using 3,000,000 particles with this technique, where at most 150,000 particles would be the maximum number of particles with a standard implementation.

In the second part, we discussed extension of Chubanov's new polynomial-time algorithm for linear programming to symmetric cone programming. Recently, Chubanov proposed a third polynomial-time algorithm for linear programming. The algorithm deals with homogeneous feasibility problem of a linear program. The algorithm finds an interior feasible solution of a linear program by repeating projection and scaling. In the talk, we introduced Chubanov's algorithm and discussed two types of further extensions for symmetric cone programming and semi-infinite programming based on separation oracle.

Index Reduction for Nonlinear Differential-Algebraic Equations via Combinatorial Relaxation

Mr. Taihei Oki, University of Tokyo, Japan

Differential-algebraic equations (DAEs) are widely used for modeling of dynamical systems. The difficulty in numerically solving a DAE is measured by its differentiation index. For highly accurate simulation of dynamical systems, it is important to convert high index DAEs into low index DAEs. Most of existing simulation software packages for dynamical systems are equipped with an index reduction algorithm given by Mattsson and Söderlind. Unfortunately, this algorithm fails if there are unlucky numerical or symbolic cancellations. This talk gives a new index reduction algorithm for nonlinear DAEs. This algorithm modifies a DAE into another DAE to which Mattsson–Söderlind's index reduction algorithm is applicable by iteratively applying the implicit function theorem. Our approach is based on the combinatorial relaxation approach, which is a framework to solve a linear algebraic problem by iteratively relaxing it into an efficiently solvable combinatorial optimization problem. Though this algorithm heavily uses symbolic manipulations, we give implementation strategies to overcome the drawback.

High Order Taylor-like Expansions of Certain Piecewise Smooth Functions and Their Application to DAEs with Flow Network Structure

Mr. Tom Streubel, Humboldt-University of Berlin, Germany

Computer evaluation procedures of piecewise smooth functions can be represented in a normalized fashion called the Abs-Normal Form (ANF). With slight modification of AD-tools (e.g. Algopy, CodiPack or Adol-C to name a few) high order piecewise polynomial approximations of ANFs can be propagated in a Taylor-like fashion efficiently. This generalized expansion-process for functions in ANF is an extension to the so called tangent mode of piecewise linearization which was introduced in [1]. We want to discuss the overall propagation process and discuss a proof of the higher order approximation qualities. Finally we will outline a framework for the construction of generalized integrators of ODE- and semi explicit DAE systems in ANF representation which treats occurring events of non differentiability without explicit detection.

[1] Griewank, Andreas: On stable piecewise linearization and generalized algorithmic differentiation; Optimization Methods and Software, 2013.

List of Participants

- Paul Barton, MIT
- Bradley Bell, University of Washington
- Francesc Anton Castro, Universidad Yachay Tech
- Olga Ebel, Universität Paderborn
- Andreas Griewank[†], School of Mathematical Sciences and Information Technology, Yachay Tech
- Laurent Hascoët, INRIA
- Xiaolin Huang, Shanghai Jiao Tong University
- Satoru Iwata, University of Tokyo
- Vladik Kreinovich, University of Texas at El Paso
- Koichi Kubota, Chuo University
- Fritz Mayer-Lindenberg, Technische Universität Hamburg
- Taihei Oki, University of Tokyo
- Siegfried Rump, Technische Universität Hamburg
- Stephan Schmidt, Universität Würzburg
- Florian Steinberg, INRIA
- Tom Streubel, Humboldt-Universität zu Berlin
- Mizuyo Takamatsu, Chuo University
- Takahito Tanabe, NTT Data Mathematical Systems inc.
- Takashi Tsuchiya, National Graduate Institute for Policy Studies
- Jean Utke, Allstate Insurance Company
- Andrea Walther, Universität Paderborn
- Kathrin Welker, Universität Trier

[†]He participated via Skype from Germany.

Meeting Schedule

Check-in Day: June 24th (Sunday)

- Welcome Banquet

Day1: June 25th (Monday)

- Session 1
  - Prof. Andrea Walther: "Abs-Linearization for Piecewise Smooth Optimization"
- Session 2
  - Ms. Olga Ebel: "Dealing with Optimization Problems Constrained by Nonlinear Piecewise Smooth PDEs"
- Group Photo Shooting
- Session 3
  - Prof. Xiaolin Huang: "Parametric and Non-parametric Piecewise Linear Models and their Optimization"
- Session 4
  - Dr. Jean Utke: "Applying Piecewise Linear optimization to a data science problem"
  - Dr. Stephan Schmidt: "Non-Smooth Geometric Inverse Problems"

Day2: June 26th (Tuesday)

- Session 1
  - Dr. Laurent Hascoët: "Non-smooth tangent differentiation experiments with Source-Transformation AD tools"
- Session 2
  - Prof. Bradley Bell: "CppAD's Abs-normal Representation"
- Session 3
  - Prof. Fritz Mayer-Lindenberg: "Applying automatic functorial substitutions to simplify programming"
  - Prof. Francesc Anton Castro: "Introduction to Algebraic and Computational Geometry"
- Session 4
  - Dr. Kathrin Welker: "Solution Techniques for Constrained Shape Optimization Problems in Shape Spaces"
  - Prof. Paul Barton: "Lexicographic Derivatives"

Day 3: June 27th (Wednesday)

- Session 1
  - Prof. Siegfried Rump: "Introduction to INTLAB"
  - Prof. Vladik Kreinovich: "How Interval Measurement Uncertainty Affects the Results of Data Processing: A Calculus-Based Approach to Computing the Range of a Box"
- Session 2
  - Prof. Takashi Tsuchiya: "Chubanov's new polynomial-time algorithm for linear programming and extensions" , and "A Recursive Recomputation Approach to Smoothing in Nonlinear and Bayesian State-Space Modeling"
- Excursion, Grate Buddha, Hase Temple
- Main Banquet

Day 4: June 28th (Thursday)

- Session 1
  - Mr. Taihei Oki: "Index Reduction for Nonlinear Differential-Algebraic Equations via Combinatorial Relaxation"
- Session 2
  - Mr. Tom Streubel: "High Order Taylor-like Expansions of PS functions and their Application to DAEs"