Applied Nonlinear Optimization in the DFG-Center

Werner Römisch[1] and Fredi Tröltzsch[2]

[1] Institut für Mathematik, Humboldt-Universität Berlin, 10099 Berlin

[2] Institut für Mathematik, Technische Universität Berlin, 10623 Berlin

1 Introduction

Modelling and simulation of complex processes in key technologies are the main issues of the DFG Research Center. Being able to efficiently deal with the complex models established and having the corresponding mathematical basis available, engineers and practitioners are interested in optimizing these models according to some objective. This is reflected by several scientific projects in the Center. They cover various fields of optimization ranging from discrete, nonlinear, or stochastic optimization to optimal control of ordinary or partial differential equations. In some of the projects, different fields interact and grow together.

Traditionally, Berlin is a good place to deal with optimization. All fields mentioned above are covered by well known mathematicians working at one of the scientific institutions in Berlin.

In this paper, we first introduce some basic ideas of nonlinear optimization, aiming at guiding the reader from the calculus of variations started in 1696 to necessary optimality conditions for nonlinear optimization problems in Banach spaces. Finally, we arrive at specific applied optimization problems which shall be the subject of study in the DFG Center. A detailed presentation of all aspects of optimization that shall be treated in the Center would go beyond the scope of this contribution. Therefore, we will only shed light on a few selected projects contributing to the fields of stochastic optimization and optimal control of partial differential equations, where the two authors are engaged.

Nonlinear optimization developed as an independent field of applied mathematics in the early fifties of the 20th century. It grew out of linear programming, which had posed new questions in the field of extremal problems during world war II, and had invented the simplex method. Over a long time this method was the most successful numerical technique to solve optimization problems. For a long period before, nonlinear optimization – in the sense of extremal problems with equality constraints – has been part of calculus. Each mathematician knows the well known Lagrange multipier rule for extrema of differentiable functions subject to finitely many equality constraints. This theorem is a compulsory part of each curriculum in calculus.

The situation becomes more difficult if more general constraints (e.g., inequalities) are given. Let us consider the nonlinear optimization problem in finite-dimensional spaces,

$$\text{(P1)} \qquad \text{minimize} \;\; f_0(x) \text{ subject to } x \in C,$$

where $f_0 : \mathbb{R}^m \to \mathbb{R}$ is differentiable and C is a closed subset of $\mathbb{R}^m$. Clearly, the properties of the boundary of C can be crucial when characterizing a solution to (P1). The boundary may have all kinds of curvilinear facets, edges and corners. Hence, an approach to geometry that can cope with such a lack of smoothness is needed. Such an approach, sometimes called *variational geometry*, was developed by associating certain cones with each point of C instead of the classical subspaces, [4, 37].

To characterize optimality for (P1) one makes use of the concepts of tangent and normal cones to a set and of the regularity of a set. The sets

$$T_C(\bar{x}) := \Big\{ w \in \mathbb{R}^m : w = \lim_{n\to\infty} \frac{x_n - \bar{x}}{\tau_n} \text{ for some sequences } x_n \xrightarrow{C} \bar{x}, \tau_n \searrow 0 \Big\}$$

$$N_C(\bar{x}) := \Big\{ v \in \mathbb{R}^m : \limsup_{\substack{x \xrightarrow{C} \bar{x} \\ x \neq \bar{x}}} \frac{\langle v, x - \bar{x} \rangle}{\|x - \bar{x}\|} \leq 0 \Big\}$$

are called the *tangent cone* and the *normal cone*, respectively, to C at some element $\bar{x} \in C$. Indeed, the sets $T_C(\bar{x})$ and $N_C(\bar{x})$ are closed cones at every $\bar{x}$ in C. In addition, the normal cone is convex and polar to the tangent cone, i.e., $N_C(\bar{x}) = \{v \in \mathbb{R}^m : \langle v, w \rangle \leq 0 \text{ for all } w \in T_C(\bar{x})\}$. The set C is said to be *regular* at $\bar{x} \in C$, if for every $v \in N_C(\bar{x})$, there exist sequences (x_n) in C and (v_n) with $v_n \in N_C(x_n)$ that converge to $\bar{x}$ and v, respectively. If C is regular at $\bar{x}$, the cones $N_C(\bar{x})$ and $T_C(\bar{x})$ are polar to each other and the tangent cone is convex, too. If C is convex, it is regular at any $\bar{x} \in C$ and the normal cone is of the form $N_C(\bar{x}) = \{v \in \mathbb{R}^m : \langle v, x - \bar{x} \rangle \leq 0 \text{ for all } x \in C\}$.

Now, a *necessary condition* for $\bar{x} \in C$ to be a local minimizer of (P1) reads

$$0 \in \nabla f_0(\bar{x}) + N_C(\bar{x})\,. \tag{1}$$

If $\bar{x}$ is an interior point of C, it holds $N_C(\bar{x}) = \{0\}$ and, hence, condition (1) reduces to the classical Fermat's rule $\nabla f_0(\bar{x}) = 0$. If C is convex, condition (1) is equivalent to the well known variational inequality

$$\langle \nabla f_0(\bar{x}), x - \bar{x} \rangle \geq 0 \;\text{ for all }\; x \in C.$$

The latter condition is sufficient for $\bar{x}$ to be globally optimal, if also f_0 is convex. In general, optimality conditions represent generalized equations.
If C is a set with constraint structure, it is possible to obtain explicit forms for their normal cones at points where C is regular. This leads to Lagrange multiplier formulations of optimality conditions. Let C be of the form

$$C = \{x \in D : F(x) \in K\},$$

where the sets $D \subseteq \mathbb{R}^m$, $K \subseteq \mathbb{R}^d$ are closed and $F = (f_1, \ldots, f_d) : \mathbb{R}^m \to \mathbb{R}^d$ is continuously differentiable. Then one has

$$N_C(\bar{x}) \supseteq \Big\{ \sum_{i=1}^{d} y_i \nabla f_i(\bar{x}) + z : y = (y_1, \ldots, y_d) \in N_K(F(\bar{x})), z \in N_D(\bar{x}) \Big\} \tag{2}$$

at any $\bar{x} \in C$. The set C is regular at $\bar{x}$ and equality holds in (2) if the sets D and K are regular at $\bar{x}$ and $F(\bar{x})$, respectively, and the *constraint qualification*

$$\big[y \in N_K(F(\bar{x})), -\sum_{i=1}^{d} y_i \nabla f_i(\bar{x}) \in N_D(\bar{x}) \big] \Longrightarrow y = 0 \tag{3}$$

or, equivalently,

$$T_K(F(\bar{x})) + \nabla F(\bar{x}) T_D(\bar{x}) = \mathbb{R}^d \tag{4}$$

is satisfied, [37, 6.14 and 6.39]. In many applications the sets D and K are convex and, hence, regular at any of their elements. D is often polyhedral, containing simple constraints (e.g., bounds), and K is a polyhedral cone given by finitely many inequalities and equations. In such cases the tangent and normal cones are polyhedral and allow explicit representations.
If the above regularity conditions are satisfied at a local minimizer $\bar{x}$ of (P1), the optimality condition (1) implies the existence of $\bar{y} \in N_K(F(\bar{x}))$ such that

$$-\Big(\nabla f_0(\bar{x}) + \sum_{i=1}^{d} \bar{y}_i \nabla f_i(\bar{x}) \Big) \in N_D(\bar{x}). \tag{5}$$

If K is a convex cone, the condition $\bar{y} \in N_K(F(\bar{x}))$ is equivalent to $F(\bar{x}) \in K$, $\bar{y} \in K^-$ with K^- denoting the polar cone to K and $\langle \bar{y}, F(\bar{x}) \rangle = 0$.
The latter result is known as the *Karush-Kuhn-Tucker theorem* [21, 24] or as the *Lagrange multiplier rule*. The crucial conditions (3) and (4), respectively, are usually referred to as (generalized) *Mangasarian-Fromovitz constraint qualifications*.

Many real world or scientific extremal problems cannot be posed in finite dimensional spaces. Often, *functions* $x = x(t)$ of a certain variable t are unknown rather than *vectors* x. One of the oldest examples of this type is the famous Brachistochrone problem posed by Johann Bernoulli in 1696. It considers a bead sliding frictionless under gravity along a smooth curve joining two points A and B and asks what shape the wire should be such that the bead, when released from rest in A, should slide to B in minimum time. In this problem, now viewed as the starting point of the *calculus of variations*, an optimal *function* has to be found, see [1] or [16]. Although this first mathematical problem in function spaces was posed more than 300 years ago, nonlinear optimization in function spaces or – more general – in Banach spaces came up only after the beginning of nonlinear optimization. This theory investigates problems of the type

$$\text{(P2)} \qquad \text{minimize } f(x) \text{ subject to } x \in C, F(x) \in K$$

where $f : X \to \mathbb{R}$ are $F : X \to Y$ are Fréchet differentiable mappings, X, Y are Banach spaces, and C and K are closed convex subsets of X and Y, respectively. Today, the abstract theory of necessary optimality conditions for optimization problems of the type (P2) is well developed. In particular, the Lagrange multiplier rule

$$-\nabla_x L(x, \lambda) \in N_C(\bar{x}),\ \lambda \in N_K(F(\bar{x}))$$

holds at a regular local minimizer $\bar{x}$ of (P2) that is analogous to (5) for (P1), [26, 8]. Here, $L(x, \lambda) := f(x) + \langle \lambda, F(x) \rangle$, for $(x, \lambda) \in X \times Y^*$, with Y^* denoting the dual to Y, is the Lagrangian of (P2).

This general class of problems covers many important mathematical optimization problems. In particular, the fields of *optimal control* and *stochastic optimization* belong to this class. Nevertheless, people working in optimal control theory have experienced that a direct application of the abstract optimality conditions does often not provide satisfactory results. The reason is that the dual spaces of Lagrange multipliers are not useful in many applications. Therefore, each class of optimization problems needs a special theoretical treatment, guided by the abstract theory.

2 Nonlinear Optimization

2.1 Optimal Control

In *optimal control*, the unknown element x is a vector of two functions y and u, $x = (y, u)$. The function u denotes the *control* that must be chosen optimally. For instance, think of the engines of a space craft that must be fired such that the vehicle moves optimally. The function y stands for the state of the physical system that is influenced by u. In optimal control, except in discrete control systems, y is obtained as the solution of a system of differential or integral equations.

Let us consider a simple academic example to explain the situation: An oscillating pendulum should be stopped by a controllable force in minimum time, see [27]. In mathematical terms, the problem reads

(P3) $\min \; T$

$$\begin{aligned} y''(t) &= -c_1 \sin(y(t)) + c_2\, u(t) \\ y(0) &= y_0 \quad y(T) = 0 \\ y'(0) &= y_1 \quad y'(T) = 0 \\ |u(t)| &\leq u_{\max}. \end{aligned}$$

Here, y is the deflection angle of the pendulum, $t \in [0, T]$ denotes the time, c_1, c_2 are certain constants, and y_0, y_1 are the initial values for the angle and the angular velocity. This is a characteristic example of optimal control of *ordinary differential equations*, and it is formally clear how this problem is related to the optimization problem in Banach spaces defined above: The function f corresponds to T, the mapping F with the cone $K = \{0\}$ covers the differential equation together with the initial- and terminal conditions, while C is formed by the functions u with maximum absolute value $u_{\max}$.

Today, it is no challenge to solve this simple problem numerically. A linearized version, where $\sin(y)$ is approximated by y, can be solved even analytically. The associated theoretical basis is the *Pontryagin Maximum Principle*, a fundamental result belonging to the greatest achievements of applied mathematics in the 20th century, [30].

In some sense it is a generalization of the Lagrange multiplier rule for optimization problems in Banach spaces and provides an *optimality system* containing the equations of (P3) for y, an *adjoint differential equation* for an adjoint state w, and the so-called *maximum condition* for u, all being mutually coupled. This optimality system is the basis of *direct methods* to solve

optimal control problems numerically. The optimal control of ordinary differential equations has many applications, for instance in space flight, aviation, robotics, chemical processes or in the generation of electrical power.

There were so many phenomena that could not be appropriately modelled by ordinary differential equations that the investigation of optimal control problems for partial differential equations soon came into play. The associated theory is more difficult, and the numerical solution of such problems is still a challenge for practitioners and mathematicians. One of the most important monographs on this subject is the well-known book [25]. It did not take long and interesting applications were discussed. We refer, for instance, to [10], where several industrial problems were considered.

Let us explain the situation by an academic problem again. Consider a 3-dimensional bounded domain $\Omega \subset \mathbb{R}^3$ standing for a body of metal that should be heated in an optimal way. We assume that the energy is produced by induction heating so that the controlled heat source u appears in the body Ω. The temperature at a point $x \in \Omega$ is denoted by $y(x)$, while $y_d = y_d(x)$ stands for a desired steady-state temperature in Ω to be generated. This problem is modelled by the following optimal control problem for the Poisson equation:

$$\text{(P4)} \qquad \min \int_\Omega (y(x) - y_d(x))^2 dx$$

subject to

$$\begin{aligned} -\Delta y(x) &= u(x) && \text{in } \Omega \\ y(x) &= y_\Gamma(x) && \text{on } \Gamma \end{aligned}$$

and

$$0 \le u(x) \le u_{\max}.$$

Here, y_Γ is the temperature at the boundary Γ of Ω, which is assumed to be known.

Although this problem is very simple and yet academic, from a mathematical point of view it covers main aspects of the problems posed in some projects of the DFG-Center. Among other states, in these projects the temperature plays the most significant role too, and it is obtained as the solution of a heat equation. The control appears as a source term in the right-hand side of the heat equation. Moreover, inequalities are given as additional constraints.

Real-world equations are more difficult to handle than the simple Poisson equation in (P4). In the projects of the DFG-Center they form nonlinear systems for different state functions, hence the theory of existence and

uniqueness is more difficult. Also, their numerical solution is a real challenge. Consequently, associated optimal control problems are not easy to handle. Their discretization leads to very large scale optimization problems which require special algorithms to be solved. Another specific difficulty arises from the presence of so-called pointwise state-constraints, i.e. of inequality constraints imposed on the state function y.

2.2 Stochastic Optimization

Stochastic optimization is concerned with models that require an optimal decision on the basis of given probabilistic information on random data. We refer to [31, 7] for introductory textbooks and to [45] for a recent state-of-the-art volume. To give an idea of such models, let $\{\xi_t\}_{t=1}^T$ be a discrete-time stochastic data process on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with $\xi_t : \Omega \to \mathbb{R}^{s_t}$. The stochastic decision $x_t : \Omega \to \mathbb{R}^{m_t}$ at period t is assumed to be *nonanticipative*, i.e., to depend only on $\xi^t := (\xi_1, \ldots, \xi_t)$. With $\mathcal{F}_t \subseteq \mathcal{F}$ denoting the σ-algebra generated by ξ^t, we have $\mathcal{F}_t \subseteq \mathcal{F}_{t+1}$ for $t = 1, \ldots, T-1$. Furthermore, we assume that $\mathcal{F}_1 = \{\emptyset, \Omega\}$, i.e., that the data ξ_1 at $t = 1$ is deterministic, and that $\mathcal{F}_T = \mathcal{F}$. We assume that $\xi_t \in L_1(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^{s_t})$ and $x_t \in L_\infty(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^{m_t})$ for each $t = 1, \ldots, T$. Then the nonanticipativity condition may be expressed by the subspace

$$\mathcal{N}_{na} = \{x \in \times_{t=1}^T L_\infty(\Omega, \mathcal{F}, P; \mathbb{R}^{m_t}) \,:\, x_t = \mathbb{E}[x_t|\mathcal{F}_t]\,,\, t = 1, \ldots, T\}$$

using the conditional expectation $\mathbb{E}[\cdot|\mathcal{F}_t]$ with respect to the σ-algebra $\mathcal{F}_t$. We consider the stochastic optimization model

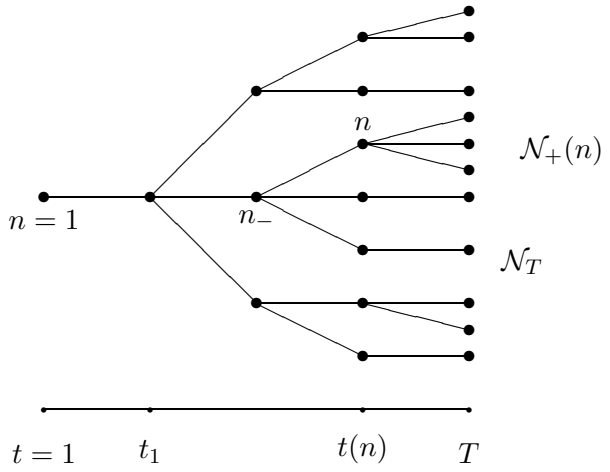
$$\text{(P5)}\quad \min\ \mathbb{E}\Big[\sum_{t=1}^T c_t(\xi_t, x_t)\Big] \text{ s.t. } \ x \in \mathcal{N}_{na},\ x_t \in X_t(\xi_t),\ g_t(\xi_t, x_t, x_{t-1}) \leq 0,$$

where the sets $X_t(\xi_t)$ are nonempty and compact, c_t and g_t are affine linear functions in each variable, and $\mathbb{E}$ denotes expectation w.r.t. $\mathbb{P}$.

(P5) represents an infinite-dimensional optimization problem. It becomes finite-dimensional if Ω is finite, i.e., $\Omega = \{\omega_s\}_{s=1}^S$ and $\mathcal{F}$ is the power set of Ω. Let $\xi_t^s := \xi_t(\omega_s)$ be the data scenario s at time t, x_t^s the decision scenario s at t, and $p_s := \mathbb{P}(\{\omega_s\})$ the probability of scenario s. Then the nonanticipativity condition $x \in \mathcal{N}_{na}$ can be expressed by a finite number of linear equality constraints for the decision variables x_t^s (e.g. using finite partitions $\mathcal{E}_t$ of Ω that generate $\mathcal{F}_t$). Then the stochastic program (P5) takes the *scenario* form

$$\min \sum_{s=1}^S \sum_{t=1}^T p_s c_t(\xi_t^s, x_t^s) \text{ s.t. } x \in \mathcal{N}_{na},\ x_t^s \in X_t(\xi_t^s),\ g_t(\xi_t^s, x_t^s, x_{t-1}^s) \leq 0.$$

Such optimization models contain ST vectorial decisions, about $3ST$ constraints and, hence, are large scale. The number S is typically large as it comes from an approximation or sampling procedure, and, thus, the model is extremely huge in most cases and requires specific solution approaches.
As $\mathcal{F}_t$ is contained in $\mathcal{F}_{t+1}$, each element of $\mathcal{E}_t$ can be represented as the union of certain elements of $\mathcal{E}_{t+1}$. Representing the elements of $\mathcal{E}_t$ by *nodes* and the above set relations by *arcs* leads to a tree that is called *scenario tree*. It is based on a finite set $\mathcal{N} \subset \mathbb{N}$ of nodes, where $n = 1$ stands for the root node at period $t = 1$. We denote by n_- the unique predecessor of node n, by path(n):=$\{1, \ldots, n_-, n\}$ the path from the root node to node n with length $t(n) := |\text{path}(n)|$, by $\mathcal{N}_t := \{n : t(n) = t\}$ the nodes at time t and by $\mathcal{N}_+(n)$ the set of successors to node n. By $\{\pi_n\}_{n\in\mathcal{N}_T} := \{p_s\}_{s=1}^S$ and $\pi_n := \sum_{n_+\in\mathcal{N}_+(n)} \pi_{n_+}$, $n \in \mathcal{N}$, we assign probabilities to each node.



Scenario tree with $t_1 = 2$, $T = 5$, $|\mathcal{N}| = 23$ and 11 leaves

Let $\{\xi^n\}_{n\in\mathcal{N}_t}$ and $\{x^n\}_{n\in\mathcal{N}_t}$ denote the realizations of ξ_t and x_t, respectively. Then the *scenario tree* form of the stochastic program (P5) reads

$$\min \sum_{n\in\mathcal{N}} \pi_n c_{t(n)}(\xi^n, x^n) \text{ s.t. } x^n \in X_{t(n)}(\xi^n),\ g_{t(n)}(\xi^n, x^n, x^{n_-}) \leq 0.$$

The dimensions of the scenario tree form are considerably smaller than in the previous formulation, as the number $|\mathcal{N}|$ of nodes is much smaller than TS and the nonanticipativity constraints are incorporated into the tree structure. The models exhibit special structures, but are still of enormous size.
The standard algorithmic approach in stochastic programming is, consequently, *decomposition*, [6, 43]. While primal decomposition approaches need

convexity to become efficient for reasonably large models, dualization techniques seem to work well in large scale nonconvex situations, too. Dual decomposition approaches circle around duality results for stochastic programs (see [36] for classical work in this direction). They are based on Lagrangian relaxation of certain groups of constraints and on solving the corresponding Lagrangian dual by subgradient-type methods. For instance, relaxing the nonanticipativity constraints leads to scenario decomposition, and relaxing component coupling constraints to component or geographic decomposition (see [41] for details). Hence, the specific decomposition strategy depends on the model structure and, in the nonconvex case, on the size of the relevant duality gaps, [11]. By means of the recursive *dynamic programming* construction

$$\begin{aligned} f_T(y_1,\ldots,y_T,\xi(\omega)) &:= \sum_{t=1}^{T} c_t(\xi_t(\omega),y_t), \\ \varphi_t(y_1,\ldots,y_t,\omega) &:= \mathbb{E}^r[f_t(y_1,\ldots,y_t,\xi(\cdot))|\mathcal{F}_t](\omega), \\ f_{t-1}(y_1,\ldots,y_{t-1},\xi(\omega)) &:= \inf_y \varphi_t(y_1,\ldots,y_{t-1},y,\omega), \end{aligned}$$

for $t = T,\ldots,2$ and for each $\omega \in \Omega$ and feasible y_τ, $\tau = 1,\ldots,T$, the original stochastic program can be shown to be equivalent to the nonlinear program

$$\text{(P6)} \qquad \min \ \mathbb{E}[f_1(x_1,\xi)] \ \text{ s.t. } \ x_1 \in X_1(\xi_1)$$

for determining the (deterministic) decision $x_1 \in \mathbb{R}^{m_1}$ at $t = 1$. Here, $\mathbb{E}^r[\cdot|\mathcal{F}_t]$ denotes the regular conditional expectation, which represents a version of $\mathbb{E}[\cdot|\mathcal{F}_t]$ satisfying special measurability properties. This procedure works under quite weak measurability and boundedness assumptions, [14].

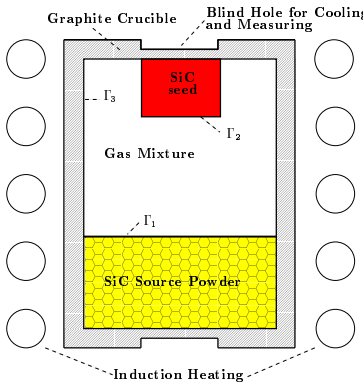
The nonlinear program (P6) serves as a basis for stability studies of the original stochastic program (P5). It is called *stable* at the underlying probability distribution P of ξ if the optimal values and solution sets of the latter program behave continuously for small perturbations of P in some spaces of measures equipped with some probability metric. The corresponding analysis is based on general perturbation results of optimization problems [5, 8, 37]. We refer to [40] for bibliographical notes on the stability of stochastic programs and to [32] for a systematic theory of probability metrics. Stability properties justify the approximation or estimation of P by simpler (e.g. finitely discrete) measures. In [33, 40] certain ideal probability metrics for stability are associated with specific classes of stochastic programs. They may be used directly or as a guide for constructing scenario tree approximations to P.

3 Optimal Control of PDEs and Stochastic Optimization in the DFG-Center

3.1 Production of silicon carbide bulk single crystals

The control of PDEs plays a role in a number of projects of the DFG-Center. In particular, this concerns the application of optimization methods in regional hyperthermia, methods of model reduction for various kinds of PDEs, and the optimization in crystal growth. To give the reader an impression, we concentrate here on the latter project.

Silicon carbide (SiC) bulk single crystals have important applications in key technologies (MESFETs, thyristors, LEDs, lasers, sensors, etc.), especially in high power, high frequency, high temperature, or intensive radiation environments. The main procedure for their production, the *Physical Vapor Transport* method, employs an induction-heated graphite crucible containing polycrystalline SiC source powder and an SiC single crystal seed cooled by means of a blind hole. The system is kept in a low pressure inert gas atmosphere and is heated up to temperatures between 2000 and 3000 K by an induction coil located around the crucible (see Figure below).



High temperature and low pressure let the SiC powder sublimate, adding species such as Si, Si_2C, and SiC_2 to the gas phase. The crystallization occurs at the cooled seed, which thereby grows into the reaction chamber. Quality and growth rate of the crystal strongly depend on the evolution of temperature distribution, mass transport, and species concentrations. The research group of Prof. S. Sprekels (WIAS) has a long standing cooperation in modelling and simulation of these processes with the Institute of Crystal Growth (IKZ) Adlershof .

It is a challenging technological aim to *optimize* quality and growth rate

of the crystals by adjusting the control parameters. In the DFG-Center, this issue will be considered in research groups headed by O. Klein, J. Sprekels (WIAS) and A. Rösch, F. Tröltzsch (TUB).

We explain the situation for a simplified model where the optimization aims at generating a radially constant temperature profile in the growth chamber $\Omega_{\rm g}$, i.e., $\partial\theta/\partial r = 0$ should hold in $\Omega_{\rm g}$. This gives rise to the problem

$$(P6) \qquad \min \int_{\Omega_{\rm g}} \int_0^T H(\theta(x,t) - \theta_{\min,\,\rm crystal}) \Big(\frac{\partial\theta}{\partial r}(x,t)\Big)^2 dx\,dt$$

(H: Heaviside function) subject to

$$\begin{aligned}
\rho\, c(\theta)\, \frac{\partial\theta}{\partial t} &= \text{div}\ (k(\theta)\,\text{grad}\ \theta) + u && \text{in } \Omega \times [0,T],\\
k(\theta)\frac{\partial\theta}{\partial n} &= \sigma\,(\theta_{\rm a}^4 - \theta^4) && \text{on } \Gamma_{\rm o} \times [0,T],\\
\big((k(\theta)\,\nabla\theta)_{\rm gas} - (k(\theta)\,\nabla\theta)_{\rm solid}\big)\cdot n &= R - J && \text{on } \Gamma_i \times [0,T],\\
\theta(x,0) &= \theta_0 && \text{in } \Omega
\end{aligned}$$

and

$$\begin{aligned}
&\theta(x,t) \le \theta(x',t) && \forall x \in \Gamma_1, \quad \forall x' \in \Gamma_2, \quad \forall t \in [0,T],\\
&\theta_{\min} \le \theta(x,t) \le \theta_{\max} && \forall x \in \Omega, \quad \forall t \in [0,T],\\
&\theta_{\min,\,\rm crystal} \le \theta(x,T) \le \theta_{\max,\,\rm crystal} && \forall x \in \Gamma_2,\\
&\delta + \theta(x,T) \le \theta(x',T) && \forall x \in \Omega_{\rm source}, \quad \forall x' \in \Gamma_2,\\
&0 \le P(t) \le P_{\max} && \forall t \in [0,T].
\end{aligned}$$

In this setting, the following quantities are used: The domain Ω (covering crucible and growth chamber $\Omega_{\rm g}$), the domain $\Omega_{\rm source}$ of the SiC source, the boundary $\Gamma_i = \Gamma_1 + \Gamma_2 + \Gamma_3$ of $\Omega_{\rm g}$ including the surface of the SiC source Γ_1 and the surface of the SiC seed crystal Γ_2 (Figure xx), the outer boundary $\Gamma_{\rm o}$ of the crucible, outer temperature $\theta_{\rm a}$, initial temperature θ_0, and the average power $P(t)$ of the induction coil. The number δ is a temperature difference to be established between SiC source and seed, and $[\theta_{\min,\,\rm crystal}, \theta_{\max,\,\rm crystal}]$ constitutes the temperature range leading to the growth of the desired SiC polytype. Moreover, the so-called radiosity R and the irradiation J occur. Both quantities depend on integrals containing θ and other quantities. Therefore, the heat equation is given with a non-local boundary condition.

In a more realistic setting, the boundary conditions in the blind holes are different from those posed at the other parts of the crucible. Moreover, the heat equation is oversimplified. Being more precise, it admits different forms in $\Omega_{\rm g}$ and $\Omega \setminus \Omega_{\rm g}$.

The function $u = u(x, t; P)$ stands for the heat source generated by induction heating. It depends on a scalar magnetic potential Φ that is obtained from Maxwell's equations, where the average power $P(t)$ plays the role of the actual control. We do not specify all these equations and associated boundary conditions here. We only state that a mapping $P \mapsto \Phi \mapsto u$ is provided by the solution of Maxwell's equations. Under more realistic assumptions, the heat equation and Maxwell's equations are mutually coupled. In the simplified case of decoupling, all functions u form a set of admissible auxiliary controls which, in some sense, can be precomputed.

Compared with the optimization in regional hyperthermia, we see some similarities, but also additional difficulties. The state equations are quasilinear too, but now we have a coupled system of PDEs consisting of the *instationary* heat equation for θ and Maxwell's equations for induction heating. Moreover, certain diffusion equations modelling the transport of chemical substrates and equations for chemical reactions have to be considered additionally. Nonlocal boundary conditions are given. Altogether, the physical behaviour of the process is modelled by a highly nonlinear system of equations. Their analysis is a challenge in itself. The numerical analysis including all aspects of optimization is even more complicated, since also pointwise state constraints are given.

It is clear that the analysis of this problem must begin with simplified models. The same concerns the numerical optimization. Having solved the simplified problems, more difficulties can be included to finally approach the problem in its full generality.

The investigations do not have to start from scratch. We refer to [9, 22, 23, 29] (modelling and simulation of crystal growth), [35] (Pontryagin principle for the control of semilinear parabolic equations with state-contraints), [34] (associated second order sufficient optimality conditions), and [2, 17, 44, 42], (numerical analysis and application of the SQP Method).

4 Stochastic Optimization in the DFG-Center

4.1 Electricity portfolio management under risk

Traditional models in stochastic programming and in stochastic power management are based on minimizing costs or maximizing expected revenues. Typically, such models do not reflect the risk of decisions. In power utilities, portfolio and risk management are often considered separate tasks. Recently,

it was proposed to unify earlier approaches to risk modelling, to identify reasonable properties of risk (e.g. coherence, convexity), to develop a theory of risk measures, [3, 15], and to incorporate risk functionals into stochastic programming models, [38]. Incorporating risk into the objective or constraints of stochastic programs may change their structural and stability properties and algorithmic approaches, [39, 40]. Accordingly, the first challenge consists in identifying favourable properties of risk functionals for the stability and computations of/in mean-risk stochastic programming models. Here, *mean-risk* stands for models containing expectation and risk terms in the objective or constraints.

Let us consider a German power utility that owns a hydro-thermal generation system and acts in the liberalized electricity market. The utility aims at decisions on production and trading electricity such that its profit is maximized and all operational constraints are met. Corresponding (deterministic) optimization models were regularly solved for short- and mid-term time horizons in the past. During the last years the utilities were confronted with new challenges. The former mostly bilateral power contracts are now supplemented by a variety of electricity contracts at power exchanges. Due to the increasing role of competition and trading, the stochasticity of data (e.g. electricity prices and electrical load) becomes more and more important.

Thermal power plant

Let I be the set of power generation units and contracts (also called units in the following), respectively, and T be the number of time periods in the time horizon (day, week or year). Let $\xi = \{\xi_t\}_{t=1}^{T}$ be the multivariate stochastic data process, p_{it} the stochastic decision (vector) for unit i at time t and $S_i(\xi)$ the set of constraints for the decision of the single unit i. The vector p_{it} may contain a $\{0,1\}$-component to model an on/off-decision (e.g. in the case of thermal generation units). The sets $S_i(\xi)$ may depend on the data process e.g. in the case of a hydraulic unit with stochastic inflow. They typically contain bounds for each period and dynamic constraints (e.g. for the operation of reservoirs of hydraulic units and minimum up/down conditions for thermal units). In addition, there is a set $C(\xi)$ describing equilibrium, reserve and group constraints that couple the output of different power units. With $r_i(\xi_t, p_{it})$ denoting the (stochastic) revenue of the decision p_{it} at unit i and

time t, the electricity portfolio management model is of the form

$$\text{(P9)} \qquad \max \mathbb{E}\Big[\sum_{it} r_i(\xi_t, p_{it})\Big] \ \text{ subject to } \ p_i \in S_i(\xi), p \in C(\xi),$$

when the expected total revenue is to be maximized. As such an objective may lead to decisions with enormous risk, one might be led to replacing the objective function by some other risk measure $\mathbb{F}$ (instead of $\mathbb{E}$) of the stochastic total revenue $\sum_{it} r_i(\xi_t, p_{it})$ or to introducing additional risk constraints.
If the functions $r_i(\xi_t, \cdot)$ are piecewise linear concave (which can be assumed in most practical cases), the portfolio management model (P9) represents a large scale linear mixed-integer stochastic program. It can be solved by a Lagrangian relaxation strategy for the constraints in $C(\xi)$, which leads to a decomposition into single unit subproblems, and by a subsequent Lagrangian heuristic [18, 19]. If, however, risk terms enter the model, this decomposition technique has to be modified or may even fail.

Furthermore, the methods for constructing scenario trees for such mean-risk models that have to be developed must be adapted to the structural properties of the optimization model and of the risk term(s). Recent stability studies in [33, 40] offer certain ideal probability metrics that may be used for such constructions. Extending the work in [19], this methodology will be used to construct multivariate load, spot price and inflow trees.

The project on mean-risk models in electricity portfolio management in the DFG-Center will be directed by W. Römisch (HUB) and R. Henrion (WIAS). All algorithms mentioned above will be implemented and tested on real-life data of power exchanges and of the hydro-thermal generation system of the cooperation partner E.ON Sales & Trading GmbH.

4.2 OD optimization in airline revenue management

Revenue management aims at controlling the sale of inventory under uncertainty such that the (expected) profit is maximized, [28, 46]. Since its foundation, airline revenue management has concentrated on optimizing booking control parameters on a single flight (leg) level. However, many characteristics are not leg-specific (e.g. the uncertain passenger demand), but depend on origin-destination itineraries. Hence, the performance of leg-based approaches is limited and optimization methods that work on the origin-destination (OD) level are needed. But, the presently available OD optimization methods [46] have some shortcomings in common (e.g. unrealistic assumptions, ignoring the integer nature of the model, separation of related

tasks).

A new stochastic programming approach for OD optimization that applies to general stochastic passenger demand processes, i.e., without restrictive assumptions on the underlying probability distribution, shall be developed.

Let us consider an airline network at a specific day consisting of sets I of origins, J of destinations and L of legs. For each leg $l \in L$ we have K_l booking classes. Let $f_{i,j,k}$ denote the fare associated with a passenger in the OD market from the origin i to destination j in booking class k. The airline wishes to select the booking limits (i.e., the maximum number of available seats) for all booking classes and legs such that the expected profit is maximized. Hence, stochastic programming models for static OD optimization are of the form

$$\text{(P10)} \quad \max \mathbb{E}\Big[\sum_{i,j,k} f_{i,j,k} x_{i,j,k}\Big] \text{ subject to } 0 \le x_{i,j,k} \le d_{i,j,k}\,, \sum_{k} x_{i,j,k} \le c_{l_{i,j}}\,,$$

where $x_{i,j,k}$ and $d_{i,j,k}$ are the booking limit and the demand, respectively, from i to j in booking class k. By c_l we denote the capacity of leg $l \in L$ and $l_{i,j}$ denotes the leg in L that is relevant for the OD market from i to j. Dynamic OD optimization models are of particular importance in the airline industry as they take into account that the customer's willingness to pay tends to increase if t gets closer to T and, thus, discount fares are made available for those who make reservations early. In such models the revenues, the demand processes and the booking limits also depend on the booking time t, starting at the initial time $t = 0$ and ending at $t = T$, i.e., the day of departure. As a result of dynamic models the airline obtains answers on the optimal booking limits at the initial time and on their change over time until $t = T$. Clearly, such models are large scale linear stochastic integer programs.

The mathematical challenges for solving OD optimization models consist in the design of solution algorithms and in the approximation of the multivariate stochastic booking demand process $\{d_{i,j,k,t}\}_{k,t}$ for each OD pair $(i,j) \in I \times J$. Due to the huge size of the models, algorithms have to be based on suitable decomposition strategies. The choice of a decomposition method depends on the model structure, the subproblem complexity, and also on duality gap estimates, [11, 41]. The construction of scenario tree approximations of the stochastic processes $\{d_{i,j,k,t}\}_{k,t}$ is based on historical

data of the OD network at relevant days. Many approaches for constructing scenario trees use ideas from cluster analysis [12]. The approach described in [19] combines the clustering of scenarios and scenario reduction techniques [13, 20], where both components are based on a probability metric measuring the distance of multivariate probability distributions. This approach leads to (almost) optimal tree constructions, i.e., to best approximations with respect to the probability metric. An important requirement for such constructions is, in particular, that the procedure works without assuming independence of the demand for different booking classes and time periods, respectively.

The project on OD optimization in airline revenue management will be directed by W. Römisch (HUB). The whole solution method will be implemented and tested on real-life data of the company Lufthansa Systems Berlin. The stochastic programming model will also be compared with earlier approaches in airline revenue management.

References

[1] N.I. Akhiezer. *The Calculus of Variations*, Ginn (Blaisdell), Boston, 1962.

[2] N. Arada, J.-P. Raymond, and F. Tröltzsch. On an augmented Lagrangian SQP method for a class of optimal control problems in Banach spaces, *Computational Optimization and Applications* 22 (2002), 369–398.

[3] P. Artzner, F. Delbaen, J.-M. Eber and D. Heath. Coherent measures of risk, *Mathematical Finance* 9 (1999), 203–228.

[4] J.-P. Aubin and H. Frankowska. *Set-Valued Analysis*, Birkhäuser, Boston 1990.

[5] B. Bank, J. Guddat, D. Klatte, B. Kummer and K. Tammer. *Non-Linear Parametric Optimization*, Akademie-Verlag, Berlin 1982.

[6] J.R. Birge. Stochastic programming computation and applications, *INFORMS Journal on Computing* 9 (1997), 111–133.

[7] J.R. Birge and F. Louveaux. *Introduction to Stochastic Programming*, Springer, New York 1997.

[8] J.F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems*, Springer, New York 2000.

[9] N. Bubner, O. Klein, P. Philip, J. Sprekels and K. Wilmański. A transient model for the sublimation growth of silicon carbide single crystals, *Journal of Crystal Growth* 205 (1999), 294–304.

[10] A.G. Butkovskii. *Methods for the Control of Distributed Parameter Systems (in Russian)*, Nauka, Moscow, 1975.

[11] D. Dentcheva and W. Römisch. Duality gaps in nonconvex stochastic optimization, Preprint 02-5, Institute of Mathematics, Humboldt-University Berlin, 2002 and submitted.

[12] J. Dupačová, G. Consigli and S.W. Wallace. Scenarios for multistage stochastic programs, *Annals of Operations Research* 100 (2000), 25–53.

[13] J. Dupačová, N. Gröwe-Kuska and W. Römisch. Scenario reduction in stochastic programming: An approach using probability metrics, *Mathematical Programming* (to appear).

[14] I.V. Evstigneev. Measurable selection and dynamic programming, *Mathematics of Operations Research* 1 (1976), 267–272.

[15] H. Föllmer and A. Schied. Convex risk measures and trading constraints, *Finance and Stochastics* 6 (2002), 429–447.

[16] P. Funk. *Variationsrechnung und ihre Anwendung in Physik und Technik*, Springer, Berlin, 1962.

[17] H. Goldberg and F. Tröltzsch. On a SQP–multigrid technique for nonlinear parabolic boundary control problems, in: *Optimal Control: Theory, Algorithms and Applications* (W.W. Hager and P.M. Pardalos eds.), Kluwer, Dordrecht, 1997, 154–177.

[18] N. Gröwe-Kuska, K.C. Kiwiel, M.P. Nowak, W. Römisch and I. Wegner. Power management in a hydro-thermal system under uncertainty by Lagrangian relaxation, in: *Decision Making under Uncertainty: Energy and Power* (C. Greengard and A. Ruszczyński eds.), IMA Volumes in Mathematics and its Applications Vol. 128, Springer, New York, 2002, 39-70.

[19] N. Gröwe-Kuska and W. Römisch. Stochastic unit commitment in hydrothermal power production planning, Preprint 02-3, Institut für Mathematik, Humboldt-Universität Berlin, 2002 and in: *Applications of Stochastic Programming* (Wallace, S.W.; Ziemba, W.T. eds.), MPS-SIAM Series in Optimization (accepted).

[20] H. Heitsch and W. Römisch. Scenario reduction algorithms in stochastic programming, *Computational Optimization and Applications* (to appear).

[21] W. Karush. Minima of functions of several variables with inequalities as side conditions, Master's thesis, University of Chicago, 1939.

[22] O. Klein, P. Philip. Transient numerical investigation of induction heating during sublimation growth of silicon carbide single crystals. *J. of Crystal Growth* (accepted).

[23] O. Klein, P. Philip, J. Sprekels, and K. Wilmański. Radiation- and convection-driven transient heat transfer during sublimation growth of silicon carbide single crystals, *Journal of Crystal Growth* 222 (2001), 832–851.

[24] H.W. Kuhn and A.W. Tucker. Nonlinear programming, in: Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability (Neyman, J. ed.), University of California Press, Berkeley, 1951, 481–492.

[25] J.L. Lions. *Contrôle optimal de systèmes gouvernès par des équations aux dérivées partielles*, Dunod, Gauthier–Villars, Paris, 1968.

[26] D.G. Luenberger. *Optimization by Vector Space Methods*, Wiley, New York, 1969.

[27] J. Macki and A. Strauss. *Introduction to Optimal Control Theory*, Springer–Verlag, Berlin, Heidelberg, New York, 1982.

[28] J. McGill and G. van Ryzin. Revenue management: Research overview and prospects, *Transportation Science* 33 (1999), 233–256.

[29] P. Philip. Transient Numerical Simulation of Sublimation Growth of SiC Bulk Single Crystals. Modeling, Finite Volume Method, Results. Doctoral thesis, submitted to the Department of Mathematics, Humboldt University of Berlin, 2002.

[30] L.S. Pontrjagin, V.G. Boltyanskii, R.V. Gamkrelidze and E.F. Mischenko. *The Mathematical Theory of Optimal Processes*, Wiley, New York, 1962.

[31] A. Prékopa. *Stochastic Programming*, Kluwer, Dordrecht 1995.

[32] S.T. Rachev.*Probability Metrics and the Stability of Stochastic Models*, Wiley, Chichester 1991.

[33] S.T. Rachev and W. Römisch. Quantitative stability in stochastic programming: The method of probability metrics, *Mathematics of Operations Research* (2002) (to appear).

[34] J.-P. Raymond and F. Tröltzsch. Second order sufficient optimality conditions for nonlinear parabolic control problems with state constraints, *Discrete and Continuous Dynamical Systems* 6 (2000), 431–450.

[35] J.-P. Raymond and H. Zidani. Hamiltonian Pontryagin's principles for control problems governed by semilinear parabolic equations, *Applied Mathematics and Optimization* 39 (1999), 143–177.

[36] R.T. Rockafellar and R.J-B Wets. The optimal recourse problem in discrete time: L^1-multipliers for inequality constraints, *SIAM Journal on Control and Optimization* 16 (1978), 16–36.

[37] R.T. Rockafellar and R.J-B Wets. *Variational Analysis*, Springer, Berlin 1998.

[38] R.T. Rockafellar and S. Uryasev. Conditional value-at-risk for general loss distributions, *Journal of Banking & Finance* 26 (2002), 1443–1471.

[39] W. Römisch. Optimierungsmethoden für die Energiewirtschaft: Stand und Entwicklungstendenzen, in: *Optimierung in der Energieversorgung*, VDI Berichte 1627, VDI-Verlag, Düsseldorf, 2001, 23–36.

[40] W. Römisch. Stability of Stochastic Programming Problems, Preprint submitted to Elsevier Science, 2002.

[41] W. Römisch and R. Schultz. Multistage stochastic integer programs: An introduction, in: *Online Optimization of Large Scale Systems* (M. Grötschel, S.O. Krumke and J. Rambau eds.), Springer, Berlin, 2001, 579–598.

[42] A. Rösch. Analysis schneller und stabiler Verfahren zur optimalen Steuerung partieller Differentialgleichungen. Habilitation thesis, Technical University of Berlin, 2001.

[43] A. Ruszczyński. Decomposition methods in stochastic programming, *Mathematical Programming* 79 (1997), 333–353.

[44] F. Tröltzsch. On the Lagrange-Newton-SQP method for the optimal control of semilinear parabolic equations, *SIAM Journal on Control and Optimization* 38 (1999), 294–312.

[45] R.J-B Wets and W.T. Ziemba. Stochastic programming. State of the Art 1998, *Annals Operations Research* 85 (1999).

[46] G. Yu (ed.). *Operations Research in the Airline Industry*, Kluwer, Boston 1998.