Lecture 27

Finite-Dimensional Optimal Control Problems – Optimality Conditions

Some basic concepts in optimal control theory can be illustrated very well in the context of finite-dimensional optimization. In particular, we do not have to deal with partial differential equations and several aspects from functional analysis.

27.1 Finite-dimensional optimal control problem

Let us consider the minimization problem

$$\min J(y,u) \quad \text{subject to (s.t.)} \quad Ay = Bu \quad \text{and} \quad u \in U_{ad} \tag{27.1}$$

where $J : \mathbb{R}^m \times \mathbb{R}^{n_u} \to \mathbb{R}$ denotes the cost functional, $A \in \mathbb{R}^{m\times m}$, $B \in \mathbb{R}^{m\times n_u}$ and $\emptyset \neq U_{ad} \subset \mathbb{R}^{n_u}$ is the set of admissible controls. Moreover, we set $n = m + n_u > m$.

We look for vectors $y \in \mathbb{R}^m$ and $u \in \mathbb{R}^{n_u}$ which solve (27.1).

Example 27.1.1. Often the cost functional is quadratic, e.g.,

$$J(y,u) = \frac{1}{2}\,\|y - y_d\|_2^2 + \frac{\lambda}{2}\,\|u\|_2^2,$$

where $y_d \in \mathbb{R}^m$ and $\lambda \geq 0$ hold. ♢

Problem (27.1) has the form of an optimization problem. Now we assume that A is an invertible matrix. Then we have

$$y = A^{-1}Bu. \tag{27.2}$$

In this case there exists a unique vector $y \in \mathbb{R}^m$ for any $u \in \mathbb{R}^{n_u}$. Hence, y is a dependent variable. We call u the control and y the state. In this way, (27.1) becomes a finite-dimensional optimal control problem.

We define the matrix $S \in \mathbb{R}^{m\times n_u}$ by $S = A^{-1}B$. Then, S is the solution matrix of our control system: $y = Su$. Utilizing the matrix S we introduce the so-called reduced cost functional

$$\hat{J}(u) = J(Su, u).$$

This leads to the reduced problem

$$\min \hat{J}(u) \quad \text{s.t.} \quad u \in U_{ad}. \tag{27.3}$$

In (27.3) the state variable is eliminated.

27.2 Existence of optimal controls

Definition 27.2.1. The vector $\bar{u} \in U_{ad}$ is called an *optimal control* for (27.1) provided

$$\hat{J}(\bar{u}) \leq \hat{J}(u) \qquad \text{for all } u \in U_{ad}.$$

The vector $\bar{y} = S\bar{u}$ is the associated *optimal state*.

Theorem 27.2.2. *Suppose that J is continuous on $\mathbb{R}^m \times U_{ad}$, that U_{ad} is nonempty, bounded, closed and that A is invertible. Then, there exists at least one optimal control for (27.1).*

Proof. Since the cost functional J is continuous on $\mathbb{R}^m \times U_{ad}$, the reduced cost $\hat{J}$ is continuous on U_{ad}. Furthermore, $U_{ad} \subset \mathbb{R}^{n_u}$ is bounded and closed. This implies that U_{ad} is compact. Due to the theorem of Weierstrass (cf. [11, Folgerung 10.9]) $\hat{J}$ has a minimum $\bar{u} \in U_{ad} \neq \emptyset$, i.e., $\hat{J}(\bar{u}) = \min_{u \in U_{ad}} \hat{J}(u)$. □

In the case of an infinite-dimensional problem the proof for the existence of optimal controls is more complicated. The reason for this fact is that bounded and closed sets in infinite-dimensional linear spaces need not to be compact.

27.3 First-order necessary optimality conditions

To compute solutions to optimal control problems we make use of optimality conditions. For that purpose we study first-order conditions for optimality.

We use the following notation for a function $\hat{J} : \mathbb{R}^{n_u} \to \mathbb{R}$:

$$\hat{J}'(u) = \nabla \hat{J}(u)^\top \quad \text{for } u \in \mathbb{R}^{n_u}.$$

For the directional derivative in direction $h \in \mathbb{R}^{n_u}$ we have

$$\hat{J}'(u)h = \langle \nabla \hat{J}(u), h \rangle_2 = \nabla \hat{J}(u)^\top h.$$

Throughout we assume that all partial derivatives of J exist and are continuous. From the chain rule it follows that $\hat{J}(u) = J(Su, u)$ is continuously differentiable.

Example 27.3.1. Let us consider the cost functional

$$\hat{J}(u) = \frac{1}{2} \|Su - y_d\|_2^2 + \frac{\lambda}{2} \|u\|_2^2,$$

see Example 27.1.1. We obtain $\nabla \hat{J}(u) = S^T(Su - y_d) + \lambda u$, $\hat{J}'(u) = (S^T(Su - y_d) + \lambda u)^T$ and $\hat{J}'(u)h = \langle S^T(Su - y_d) + \lambda u, h \rangle_{\mathbb{R}^{n_u}}$ at $u \in \mathbb{R}^{n_u}$ and for $h \in \mathbb{R}^{n_u}$. ◊

The next result is proved in [8, Theorem 22.1.2].

Theorem 27.3.2. *Suppose that $\bar{u}$ is an optimal control for (27.1) and U_{ad} convex. Then the variational inequality*

$$\hat{J}'(\bar{u})(u - \bar{u}) = \nabla \hat{J}(\bar{u})^\top (u - \bar{u}) \geq 0 \qquad \text{for all } u \in U_{ad} \tag{27.4}$$

holds.

It follows from Theorem 27.3.2 that at $\bar{u}$ the cost functional $\hat{J}$ can not decrease in any feasible direction. The proof follows from a more general result (see [30, page 63]).

From the chain rule we derive

$$\begin{aligned} \hat{J}'(\bar{u})h &= J_y(S\bar{u}, \bar{u})Sh + J_u(S\bar{u}, \bar{u})h = \langle \nabla_y J(\bar{y}, \bar{u}), A^{-1}Bh \rangle_2 + \langle \nabla_u J(\bar{y}, \bar{u}), h \rangle_2 \\ &= \langle B^\top A^{-\top} \nabla_y J(\bar{y}, \bar{u}) + \nabla_u J(\bar{y}, \bar{u}), h \rangle_2, \end{aligned} \tag{27.5}$$

where $(A^\top)^{-1} = (A^{-1})^\top := A^{-\top}$ holds and, e.g., J_y stands for the partial derivative of J with respect to the argument y. Thus, we derive from (27.4)

$$\langle B^\top A^{-\top} \nabla_y J(\bar{y}, \bar{u}) + \nabla_u J(\bar{y}, \bar{u}), u - \bar{u} \rangle_2 \geq 0 \tag{27.6}$$

for all $u \in U_{ad}$. In the following subsection we will introduce the so-called adjoint or dual variable. Then, we can express (27.6) in a simpler way.

27.4 Adjoint variable and reduced gradient

In a numerical realization the computation of A^{-1} is avoided. The same holds for the matrix $A^\top$. Thus, we replace the term $A^{-\top}\nabla_y J(\bar y,\bar u)$ by $\bar p = -A^{-\top}\nabla_y J(\bar y,\bar u)$, which is equivalent with

$$A^\top \bar p = -\nabla_y J(\bar y,\bar u). \tag{27.7}$$

Definition 27.4.1. Equation (27.7) is called the *adjoint* or *dual equation*. Its solution $\bar p$ is the *adjoint* or *dual variable* associated with $(\bar y,\bar u)$.

Example 27.4.2. For the quadratic cost functional $J(y,u) = (\|y-y_d\|_2^2 + \lambda\,\|u\|_2^2)/2$ with $y, y_d \in \mathbb{R}^m$ and $\lambda \ge 0$ we derive the adjoint equation

$$A^\top \bar p = y_d - \bar y.$$

Here we have used $\nabla_y J(y,u) = y - y_d$. ◊

The introduction of the dual variable yields two advantages:

1) We obtain an expression for (27.6) without the matrix $A^{-\top}$.

2) The expression (27.6) can be written in a more readable form.

Utilizing $\bar y = S\bar u$ in (27.5) we find that

$$\nabla \hat J(\bar u) = -B^\top \bar p + \nabla_u J(\bar y,\bar u).$$

The vector $\nabla \hat J(\bar u)$ is called the *reduced gradient*. The directional derivative of the reduced cost functional $\hat J$ at an arbitrary $u \in U_{ad}$ in direction h is given by

$$\hat J'(u)h = \langle -B^\top p + \nabla_u J(y,u), h\rangle_2,$$

where $y = Su$ and $p = -A^\top \nabla_y J(y,u)$ hold. From Theorem 27.3.2 and (27.6) we derive directly the following theorem.

Theorem 27.4.3. *Suppose that A is invertible, $\bar u$ is an optimal control for (27.1) and $\bar y = S\bar u$ the associated optimal state. Then, there exists a unique dual variable $\bar p$ satisfying (27.7). Moreover, the variational inequality*

$$\langle -B^\top \bar p + \nabla_u J(\bar y,\bar u), u - \bar u\rangle_2 \ge 0 \quad \text{for all } u \in U_{ad} \tag{27.8}$$

holds true.

We have derived an optimality system for the unknown variables $\bar y$, $\bar u$ and $\bar p$:

$$\begin{aligned} &A\bar y = B\bar u, \qquad \bar u \in U_{ad},\\ &A^\top \bar p = -\nabla_y J(\bar y,\bar u),\\ &\langle -B^\top \bar p + \nabla_u J(\bar y,\bar u), v - \bar u\rangle_2 \ge 0 \quad \text{for all } v \in U_{ad}. \end{aligned} \tag{27.9}$$

Every solution $(\bar y,\bar u)$ to (27.1) must satisfy, together with the dual variable $\bar p$, the necessary conditions (27.9).

If $U_{ad} = \mathbb{R}^{n_u}$ holds, then the term $u - \bar u$ can attain any value $h \in \mathbb{R}^{n_u}$. Therefore, the variational inequality (27.8) implies the equation

$$-B^\top \bar p + \nabla_u J(\bar y,\bar u) = 0.$$

Example 27.4.4. We consider the cost functional

$$J(y,u) = \frac{1}{2}\,\|Cy - y_d\|_2^2 + \frac{\lambda}{2}\,\|u\|_2^2$$

with $C \in \mathbb{R}^{\ell\times m}$, $y \in \mathbb{R}^m$, $y_d \in \mathbb{R}^\ell$, $\lambda \ge 0$ and $u \in \mathbb{R}^{n_u}$. Then,

$$\nabla_y J(y,u) = C^\top(Cy - y_d), \qquad \nabla_u J(y,u) = \lambda u.$$

Thus, we obtain the optimality system

$$\begin{aligned}
&A\bar{y} = B\bar{u}, \quad \bar{u} \in U_{ad},\\
&A^\top \bar{p} = C^\top (y_d - C\bar{y}),\\
&\langle -B^\top \bar{p} + \lambda\bar{u}, v - \bar{u}\rangle_2 \geq 0 \quad \text{for all } v \in U_{ad}.
\end{aligned}$$

If $U_{ad} = \mathbb{R}^{n_u}$ holds, we find $-B^\top \bar{p} + \lambda\bar{u} = 0$. For $\lambda > 0$ we have

$$\bar{u} = \frac{1}{\lambda} B^\top \bar{p}. \tag{27.10}$$

Inserting (27.10) into the state equation, we obtain a linear system in the state and dual variables:

$$A\bar{y} = \frac{1}{\lambda} BB^\top \bar{p}, \quad A^\top \bar{p} = C^\top (y_d - C\bar{y}).$$

If $(\bar{y}, \bar{p})$ is computed, $\bar{u}$ is given by (27.10). ◊

We have derived an optimality system for the unknown variables $\bar{y}$, $\bar{u}$ and $\bar{p}$:

$$\begin{aligned}
&A\bar{y} = B\bar{u}, \quad A^\top \bar{p} = -\nabla_y J(\bar{y}, \bar{u}), \quad \bar{u} \in U_{ad},\\
&\langle -B^\top \bar{p} + \nabla_u J(\bar{y}, \bar{u}), v - \bar{u}\rangle_2 \geq 0 \quad \text{for all } v \in U_{ad}.
\end{aligned} \tag{27.11}$$

Every solution $(\bar{y}, \bar{u})$ to (27.1) must satisfy, together with the dual variable $\bar{p}$, the necessary conditions (27.11).

Lecture 28

Finite-Dimensional Optimal Control Problems – KKT System

28.1 The Lagrange function

We have derived an optimality system for the unknown variables $\bar{y}$, $\bar{u}$ and $\bar{p}$:

$$\begin{aligned} &A\bar{y} = B\bar{u}, \quad A^\top\bar{p} = -\nabla_y J(\bar{y},\bar{u}), \quad \bar{u} \in U_{ad}, \\ &\langle -B^\top\bar{p} + \nabla_u J(\bar{y},\bar{u}), v - \bar{u}\rangle_2 \geq 0 \quad \text{for all } v \in U_{ad}. \end{aligned} \tag{28.1}$$

Every solution $(\bar{y},\bar{u})$ to (27.1) must satisfy, together with the dual variable $\bar{p}$, the necessary conditions (27.9).

If $U_{ad} = \mathbb{R}^{n_u}$ holds, then the term $u - \bar{u}$ can attain any value $h \in \mathbb{R}^{n_u}$. Therefore, the variational inequality (27.8) implies the equation

$$-B^\top\bar{p} + \nabla_u J(\bar{y},\bar{u}) = 0.$$

Example 28.1.1. We consider the cost functional

$$J(y,u) = \frac{1}{2}\,\|Cy - y_d\|_2^2 + \frac{\lambda}{2}\,\|u\|_2^2$$

with $C \in \mathbb{R}^{\ell\times m}$, $y \in \mathbb{R}^m$, $y_d \in \mathbb{R}^\ell$, $\lambda \geq 0$ and $u \in \mathbb{R}^{n_u}$. Then,

$$\nabla_y J(y,u) = C^\top(Cy - y_d), \qquad \nabla_u J(y,u) = \lambda u.$$

Thus, we obtain the optimality system

$$\begin{aligned} &A\bar{y} = B\bar{u}, \quad \bar{u} \in U_{ad}, \\ &A^\top\bar{p} = C^\top(y_d - C\bar{y}), \\ &\langle -B^\top\bar{p} + \lambda\bar{u}, v - \bar{u}\rangle_2 \geq 0 \quad \text{for all } v \in U_{ad}. \end{aligned}$$

If $U_{ad} = \mathbb{R}^{n_u}$ holds, we find $-B^\top\bar{p} + \lambda\bar{u} = 0$. For $\lambda > 0$ we have

$$\bar{u} = \frac{1}{\lambda}B^T\bar{p}. \tag{28.2}$$

Inserting (28.2) into the state equation, we obtain a linear system in the state and dual variables:

$$A\bar{y} = \frac{1}{\lambda}\,BB^\top\bar{p}, \quad A^\top\bar{p} = C^\top(y_d - C\bar{y}).$$

If $(\bar{y},\bar{p})$ is computed, $\bar{u}$ is given by (28.2). $\Diamond$

The optimality condition can be expressed by utilizing the Lagrange function $\mathcal{L} : \mathbb{R}^{2m+n_u} \to \mathbb{R}$ which is defined as

$$\mathcal{L}(y,u,p) = J(y,u) + \langle Ay - Bu, p\rangle_2, \qquad (y,u,p) \in \mathbb{R}^m \times \mathbb{R}^{n_u} \times \mathbb{R}^m.$$

It follows that the second and third conditions of (27.11) can be expressed as

$$\nabla_y \mathcal{L}(\bar{y},\bar{u},\bar{p}) = 0, \quad \langle \nabla_u \mathcal{L}(\bar{y},\bar{u},\bar{p}), u - \bar{u}\rangle_{\mathbb{R}^{n_u}} \geq 0 \quad \text{for all } u \in U_{ad}.$$

Remark 28.1.2. The adjoint equation (27.7) is equivalent to $\nabla_y \mathcal{L}(\bar{y}, \bar{u}, \bar{p}) = 0$. Thus, (27.7) can be derived from the derivative of the Lagrange functional with respect to the state variable y. Analogously, the variational inequality follows from the gradient $\nabla_u \mathcal{L}(\bar{y}, \bar{u}, \bar{p})$. ◊

It follows from Remark 28.1.2 that $(\bar{y}, \bar{u})$ satisfies the necessary optimality conditions of the minimization problem

$$\min \mathcal{L}(y, u, \bar{p}) \quad \text{s.t. } (y, u) \in \mathbb{R}^m \times U_{ad}. \tag{28.3}$$

Notice that (28.3) has no equality constraints (in contrast to (27.1)). In most applications $\bar{p}$ is not known a-priori. Thus, $(\bar{y}, \bar{u})$ can not be computed from (28.3).

28.2 Discussion of the variational inequality

In many applications the set of admissible controls has the form

$$U_{ad} = \big\{ u \in \mathbb{R}^{n_u} \,|\, u_a \le u \le u_b \big\}, \tag{28.4}$$

where $u_a \le u_b$ are given vectors in $\mathbb{R}^{n_u}$ and "$\le$" means less or equal in each component: $u_{a,i} \le u_i \le u_{b,i}$ for $i = 1, \ldots, n_u$. From (27.8) it follows that

$$\langle -B^\top \bar{p} + \nabla_u J(\bar{y}, \bar{u}), \bar{u} \rangle_2 \le \langle -B^\top \bar{p} + \nabla_u J(\bar{y}, \bar{u}), u \rangle_2$$

for all $u \in U_{ad}$. This implies that $\bar{u}$ solves the minimization problem

$$\min_{u \in U_{ad}} \langle -B^\top \bar{p} + \nabla_u J(\bar{y}, \bar{u}), u \rangle_2 = \min_{u \in U_{ad}} \sum_{i=1}^{n_u} \big(-B^\top \bar{p} + \nabla_u J(\bar{y}, \bar{u}) \big)_i u_i.$$

If U_{ad} is of the form (28.4), then the minimization of a component u_i is independent of u_j, $i \ne j$:

$$\big(-B^\top \bar{p} + \nabla_u J(\bar{y}, \bar{u}) \big)_i \bar{u}_i = \min_{u_{a,i} \le u_i \le u_{b,i}} \big(-B^\top \bar{p} + \nabla_u J(\bar{y}, \bar{u}) \big)_i u_i$$

for $1 \le i \le n_u$. Thus,

$$\bar{u}_i = \begin{cases} u_{b,i} & \text{if } (-B^\top \bar{p} + \nabla_u J(\bar{y}, \bar{u}))_i < 0, \\ u_{a,i} & \text{if } (-B^\top \bar{p} + \nabla_u J(\bar{y}, \bar{u}))_i > 0. \end{cases} \tag{28.5}$$

If $(-B^\top \bar{p} + \nabla_u J(\bar{y}, \bar{u}))_i = 0$ holds, we have no information from the variational inequality. In many cases we can use the equation $(-B^\top \bar{p} + \nabla_u J(\bar{y}, \bar{u}))_i = 0$ to obtain an explicit equation for one of the components of $\bar{u}$.

28.3 The Karush–Kuhn–Tucker system

Define the two vectors

$$\mu_a = \big[-B^\top \bar{p} + \nabla_u J(\bar{y}, \bar{u}) \big]_+, \quad \mu_b = \big[-B^\top \bar{p} + \nabla_u J(\bar{y}, \bar{u}) \big]_-, \tag{28.6}$$

where $\mu_{a,i} = (-B^\top \bar{p} + \nabla_u J(\bar{y}, \bar{u}))_i$ if the right-hand side is positive and $\mu_{a,i} = 0$ otherwise. Analogously, $\mu_{b,i} = |(-B^\top \bar{p} + \nabla_u J(\bar{y}, \bar{u}))_i|$ if the right-hand side is negative and $\mu_{b,i} = 0$ otherwise. Utilizing (28.5) we have

$$\mu_a \ge 0, \quad u_a - \bar{u} \le 0, \quad \langle u_a - \bar{u}, \mu_a \rangle_2 = 0 \qquad\qquad \mu_b \ge 0, \quad \bar{u} - u_b \le 0, \quad \langle \bar{u} - u_b, \mu_b \rangle_2 = 0.$$

These conditions are called *complementarity conditions*. The inequalities are clear. We prove $\langle u_a - \bar{u}, \mu_a \rangle_{\mathbb{R}^{n_u}} = 0$. Suppose that $u_{a,i} < \bar{u}_i$ holds. Due to (28.5) we have $(-B^\top \bar{p} + \nabla_u J(\bar{y}, \bar{u}))_i \le 0$. Thus, $\mu_{a,i} = 0$ which gives $(u_{a,i} - \bar{u}_i)\mu_{a,i} = 0$. Now we assume $\mu_{a,i} > 0$. Using (28.6) we derive $(-B^T \bar{p} + \nabla_u J(\bar{y}, \bar{u}))_i > 0$. It follows from (28.5) that $u_{a,i} = \bar{u}_i$ holds. Again, we have $(u_{a,i} - \bar{u}_i)\mu_{a,i} = 0$. Summation over $i = 1, \ldots, m$ yields $\langle u_a - \bar{u}, \mu_a \rangle_2 = 0$.

Notice that

$$\mu_a - \mu_b = -B^\top \bar{p} + \nabla_u J(\bar{y}, \bar{u}).$$

Hence,

$$\nabla_u J(\bar{y}, \bar{u}) - B^\top \bar{p} + \mu_b - \mu_a = 0. \tag{28.7}$$

Let us consider an augmented Lagrange functional

$$\tilde{\mathcal{L}}(y, u, p, \mu_a, \mu_b) = J(y, u) + \langle Ay - Bu, p \rangle_2 + \langle u_a - u, \mu_a \rangle_2 + \langle u - u_b, \mu_b \rangle_2$$

Then, (28.7) can be written as

$$\nabla_u \tilde{\mathcal{L}}(\bar{y}, \bar{u}, \bar{p}, \mu_a, \mu_b) = 0.$$

Moreover, the adjoint equation is equivalent with

$$\nabla_y \tilde{\mathcal{L}}(\bar{y}, \bar{u}, \bar{p}, \mu_a, \mu_b) = 0.$$

Here, we have used that $\nabla_y \mathcal{L} = \nabla_y \tilde{\mathcal{L}}$. The vectors μ_a and μ_b are the Lagrange multipliers for the inequality constraints $u_a - \bar{u} \leq 0$ and $\bar{u} - u_b \leq 0$.

Theorem 28.3.1. *Suppose that $\bar{u}$ is an optimal control for (27.1), A is invertible and U_{ad} has the form (28.4). Then, there exist Lagrange multipliers $\bar{p} \in \mathbb{R}^m$ and $\mu_a, \mu_b \in \mathbb{R}^{n_u}$ satisfying*

$$\begin{aligned} &\nabla_y \tilde{\mathcal{L}}(\bar{y}, \bar{u}, \bar{p}, \mu_a, \mu_b) = 0, \quad \nabla_u \tilde{\mathcal{L}}(\bar{y}, \bar{u}, \bar{p}, \mu_a, \mu_b) = 0, \quad A\bar{y} = B\bar{u}, \\ &u_a - u \leq 0, \quad \mu_a \geq 0, \quad u - u_b \leq 0, \quad \mu_b \geq 0, \quad \langle u_a - \bar{u}, \mu_a \rangle_2 = \langle \bar{u} - u_b, \mu_b \rangle_2 = 0. \end{aligned} \tag{28.8}$$

The optimality system (28.8) is called the Karush-Kuhn-Tucker (KKT) system.