Bicriterial Optimal Control by the Reference Point Method

Stefan Volkwein *

* *Lecture Optimization III, Department of Mathematics and Statistics, University of Konstanz, 18 November 2022*

1. PROBLEM FORMULATION

1.1 The state equation

For time $T > 0$ the state equation is given by

$$
\begin{aligned}
y_t(t,\boldsymbol{x}) - \Delta y(t,\boldsymbol{x}) &= \sum_{i=1}^m u_i\chi_i(t,\boldsymbol{x}) && \text{for } (t,\boldsymbol{x}) \in Q,\\
\frac{\partial y}{\partial \boldsymbol{n}}(t,\boldsymbol{x}) &= 0 && \text{for } (t,\boldsymbol{x}) \in \Sigma,\\
y(0,\boldsymbol{x}) &= y_\circ(\boldsymbol{x}) && \text{for } \boldsymbol{x} \in \Omega,
\end{aligned} \tag{1}
$$

where $\Omega \subset \mathbb{R}^d$, $d \in \{2,3\}$, is a bounded domain with Lipschitz-continuous boundary $\Gamma = \partial\Omega$ and $\boldsymbol{n}$ stands for the outward normal vector. We set $Q = (0,T) \times \Omega$ and $\Sigma = (0,T) \times \Gamma$. Let $H = L^2(\Omega)$ and $V = H^1(\Omega)$ be endowed by the canonical inner products given as

$$
\begin{aligned}
\langle \varphi, \phi\rangle_H &= \int_\Omega \varphi(\boldsymbol{x})\phi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} && \text{for } \varphi,\phi \in H,\\
\langle \varphi, \phi\rangle_V &= \langle \varphi, \phi\rangle_H + \int_\Omega \nabla\varphi(\boldsymbol{x})\cdot\nabla\phi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} && \text{for } \varphi,\phi \in V.
\end{aligned}
$$

The variable $u = (u_1,\ldots,u_m) \in \mathcal{U} = \mathbb{R}^m$ denotes the *control* and $\chi_i \in L^\infty(Q)$, $1 = 1,\ldots,m$, are given control shape functions. Furthermore, $y_\circ \in L^\infty(\Omega)$ denotes a given initial heat distribution. We write $y(t)$ when y is considered as a function in $\boldsymbol{x}$ only for fixed $t \in [0,T]$. Recall that

$$
W(0,T) = \big\{\varphi \in L^2(0,T;V)\,\big|\,\varphi_t \in L^2(0,T;V')\big\}
$$

is a Hilbert space endowed with the common inner product

$$
\langle \varphi, \phi\rangle_{W(0,T)} = \int_0^T \langle \varphi_t(t), \phi_t(t)\rangle_{V'} + \langle \varphi(t), \phi\rangle_V\,\mathrm{d}t
$$

for $\varphi,\phi \in W(0,T)$; see, e.g., Dautray and Lions (2000). A weak solution $y \in \mathcal{Y} = W(0,T)$ to (1) is called a *state* and has to satisfy for all test functions $\varphi \in V$:

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\langle y(t),\varphi\rangle_H + \int_\Omega \nabla y(t)\cdot\nabla\varphi\,\mathrm{d}\boldsymbol{x} &= \sum_{i=1}^m u_i\langle\chi_i(t),\varphi\rangle_H,\\
\langle y(0),\varphi\rangle_H &= \langle y_\circ,\varphi\rangle_H.
\end{aligned} \tag{2}
$$

It is shown in Dautray and Lions (2000) that (2) admits a unique solution y and

$$
\|y\|_{\mathcal{Y}} \le C\big(\|y_\circ\|_H + \|u\|_{\mathcal{U}}\big) \tag{3}
$$

for a contant $C \geq 0$. We introduce the linear operator $\mathcal{S} : \mathcal{U} \to \mathcal{Y}$, where $y = \mathcal{S}u$ is the solution to (2) for given $u \in \mathcal{U}$ with $y_\circ = 0$. From (3) it follows that $\mathcal{S}$ is bounded. Moreover, let $\hat y \in \mathcal{Y}$ be the solution to (2) for $u = 0$. Then, the affine linear mapping $\mathcal{U} \ni u \mapsto y(u) = \hat y + \mathcal{S}u \in \mathcal{Y}$ is affine linear, and $y(u)$ is the weak solution to (1).

1.2 The multiobjective optimal control problem

For given $u_a, u_b \in \mathcal{U}$ with $u_a \leq u_b$ in $\mathcal{U}$, the set of admissible controls is given as

$$
\mathcal{U}_{\mathsf{ad}} = \big\{u \in \mathcal{U}\,\big|\,u_a \le u \le u_b \text{ in } \mathbb{R}^m\big\}.
$$

Introducing the bicriterial cost functional

$$
J : \mathcal{Y}\times\mathcal{U} \to \mathbb{R}^2, \quad J(y,u) = \frac{1}{2}\begin{pmatrix} \|y(T) - y_\Omega\|_H^2 \\ \|u\|_{\mathcal{U}}^2 \end{pmatrix}
$$

the multiobjective optimal control problem (MOCP) reads

$$
(\mathbf{P}) \qquad \min J(y,u) \quad \text{subject to (s.t.)} \quad (y,u) \in \mathcal{F}(\mathbf{P})
$$

with the feasible set

$$
\mathcal{F}(\mathbf{P}) = \big\{(y,u) \in \mathcal{Y}\times\mathcal{U}_{\mathsf{ad}}\,\big|\,y \text{ solves (2)}\big\}.
$$

Next we define the reduced cost function $\hat J = (\hat J_1, \hat J_2) : \mathcal{U} \to \mathbb{R}^2$ by $\hat J(u) = J(\hat y + \mathcal{S}u, u)$ for $u \in \mathcal{U}$. Then, $(\mathbf{P})$ can be equivalently formulated as

$$
(\hat{\mathbf{P}}) \qquad \min \hat J(u) \quad \text{s.t.} \quad u \in \mathcal{U}_{\mathsf{ad}}.
$$

Problem $(\hat{\mathbf{P}})$ involves the minimization of a vector-valued objective. This is done by using the concepts of *order relation* and *Pareto optimality*; see, e.g., Ehrgott (2005). In $\mathbb{R}^2$ we make use of the following order relation: For all $z^1, z^2 \in \mathbb{R}^2$ we have

$$
z^1 \le z^2 \Leftrightarrow z^2 - z^1 \in \mathbb{R}^2_+ = \big\{z \in \mathbb{R}^2\,\big|\,z_i \ge 0 \text{ for } i = 1,2\big\}.
$$

Definition 1. The point $\bar u \in \mathcal{U}_{\mathsf{ad}}$ is called *Pareto optimal* for $(\hat{\mathbf{P}})$ if there is no other control $u \in \mathcal{U}_{\mathsf{ad}} \setminus \{\bar u\}$ with $\hat J_i(u) \le \hat J_i(\bar u)$, $i = 1,2$, and $\hat J_j(u) < \hat J_j(\bar u)$ for at least one $j \in \{1,2\}$.

2. THE REFERENCE POINT METHOD

2.1 The reference point problem

The theoretical and numerical challenge is to present the decision maker with an approximation of the *Pareto front*

$$
\mathcal{P} = \big\{\hat J(u)\,\big|\,u \in \mathcal{U}_{\mathsf{ad}} \text{ is Pareto optimal}\big\} \subset \mathbb{R}^2
$$

In order to do so, we follow the ideas laid out in Peitz et al. (2015) and make use of the *reference point method*: Given a reference point $z = (z_1, z_2) \in \mathbb{R}^2$ that satisfies

$$
z < \hat J(u) \quad \text{for all } u \in \mathcal{U}_{\mathsf{ad}} \tag{4}
$$

we introduce the *distance function* $F_z : \mathcal{U} \to \mathbb{R}$ by

$$
F_z(u) = \tfrac{1}{2}\,|\hat J(u) - z|^2 = \tfrac{1}{2}\big(\hat J_1(u) - z_1\big)^2 + \tfrac{1}{2}\big(\hat J_2(u) - z_2\big)^2.
$$

The mapping F_z measures the geometrical distance between $\hat J(u)$ and z.

Lemma 2. The mapping F_z is strictly convex.

Proof. The mapping F_z is of the form $F_z = \sum_{i=1}^2 g_i \circ \hat{J}_i$ where, because of (4), we have $g_i : (z_i, \infty) \to \mathbb{R}_0^+$ with $g_i(\xi) = (\xi - z_i)^2/2$. Because of the affine linearity of $u \mapsto y(u)$, $\hat{J}_1$ is convex and $\hat{J}_2$ strictly convex. Further, g_i is strictly convex and monotone increasing for $i = 1, 2$. Altogether, F_z itself is strictly convex. □

Suppose that z is componentwise strictly smaller than every objective value which we can achieve within $\mathcal{U}_{\mathsf{ad}}$. The goal is that – by approximating z as best as possible – we get a Pareto optimal point for $(\hat{\mathbf{P}})$. Therefore, we have to solve the *reference point problem*

$$(\hat{\mathbf{P}}_z) \qquad \min F_z(u) \quad \text{s.t.} \quad u \in \mathcal{U}_{\mathsf{ad}}$$

which is a scalar-valued minimization problem.

Theorem 3. For any $z \in \mathbb{R}^2$ the reference point problem admits a unique solution $\bar{u}_z \in \mathcal{U}_{\mathsf{ad}}$.

Proof. By Lemma 2 the mapping F_z is strictly convex. Now, the proof follows by standard arguments utilizing that $\mathcal{U}_{\mathsf{ad}}$ is bounded and closed in $\mathcal{U}$. □

Theorem 4. Let (4) hold and $\bar{u}_z \in \mathcal{U}_{\mathsf{ad}}$ be an optimal solution to $(\hat{\mathbf{P}}_z)$ for a given $z \in \mathbb{R}^2$. Then $\bar{u}_z$ is Pareto optimal for $(\hat{\mathbf{P}})$.

Proof. We follow along the lines of Theorem 4.20 in Ehrgott (2005): Assume that $\bar{u}_z \in \mathcal{U}_{\mathsf{ad}}$ is not Pareto optimal, then there exists a point $u \in \mathcal{U}_{\mathsf{ad}}$ with $\hat{J}(u) \le \hat{J}(\bar{u}_z)$ and $\hat{J}_j(u) < \hat{J}_j(\bar{u}_z)$ for $j \in \{1, 2\}$. Using (4) we get

$$(5) \qquad 0 < \hat{J}_i(u) - z_i \le \hat{J}_i(\bar{u}_z) - z_i \quad \text{for } i = 1, 2$$

and strictly smaller for $i = j$. Together, this yields $F_z(u) < F_z(\bar{u}_z)$ which is a contradiction to the assumption that $\bar{u}_z$ is optimal for $(\hat{\mathbf{P}}_z)$. □

By solving $(\hat{\mathbf{P}}_z)$ consecutively with an adaptive variation of z, we are able to move along the Pareto front in a uniform manner. This way, we get a sequence $\{z^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^2$ of reference points along with optimal controls $\{u^k\}_{k\in\mathbb{N}} \subset \mathcal{U}_{\mathsf{ad}}$ that solve $(\hat{\mathbf{P}}_z)$ with $z = z^k$ as well as $\{\hat{J}^k\}_{k\in\mathbb{N}} \subset \mathbb{R}^2$ with $\hat{J}^k = \hat{J}(u^k)$. To be more precise, the next reference point z^{k+1} is chosen as

$$(6) \quad z^{k+1} = \hat{J}^k + h_J \frac{\hat{J}^k - \hat{J}^{k-1}}{|\hat{J}^k - \hat{J}^{k-1}|} + h_z \frac{\hat{J}^k - z^k}{|\hat{J}^k - z^k|} \text{ for } k \ge 2,$$

where $h_J, h_z \ge 0$ are chosen to control the coarseness of the approximation to the Pareto front. The algorithm is initialized by applying the weighted sum method to $(\hat{\mathbf{P}})$; Zadeh (1963). This yields the first iterates $\hat{J}^1, \hat{J}^2 \in \mathcal{P}$. We therefore do not require z^1, z^2 and compute z^3 by setting $h_z = 0$ in (6). Note that the algorithm only moves in one direction: If $\hat{J}_1^1 > \hat{J}_1^2$, then it turns to the upper left in the $\mathbb{R}^2$-plane. Therefore, we perform the algorithm twice, the second time with switched roles of $\hat{J}^1$, $\hat{J}^2$ to cover the other direction as well.

2.2 Optimality conditions

Applying the chain rule, we get for any $u \in \mathcal{U}$

$$\frac{\partial F_z}{\partial u_j}(u) = \sum_{k=1}^2 (\hat{J}_k(u) - z_k) \frac{\partial \hat{J}_k}{\partial u_j}(u), \quad \text{for } j = 1, \ldots, m$$

and

$$\nabla F_z(u) = \sum_{k=1}^2 (\hat{J}_k(u) - z_k) \nabla \hat{J}_k(u).$$

The *first-order necessary optimality condition* for an optimal $\bar{u}_z \in \mathcal{U}_{\mathsf{ad}}$ now reads as the variational inequality

$$(7) \qquad \begin{aligned} 0 &\le \langle \nabla F_z(\bar{u}_z), u - \bar{u}_z \rangle_{\mathcal{U}} \\ &= \nabla F_z(\bar{u}_z)^\top (u - \bar{u}_z) \quad \text{for all } u \in \mathcal{U}_{\mathsf{ad}}. \end{aligned}$$

Next, we investigate second-order derivatives: Note that for $1 \le i, j \le m$ we find

$$\begin{aligned} \frac{\partial^2 F_z}{\partial u_i \partial u_j}(u) &= \frac{\partial}{\partial u_i}\left(\frac{\partial F_z}{\partial u_j}(u)\right) \\ &= \frac{\partial}{\partial u_i}\left(\sum_{k=1}^2 (\hat{J}_k(u) - z_k) \frac{\partial \hat{J}_k}{\partial u_j}(u)\right) \\ &= \sum_{k=1}^2 \left((\hat{J}_k(u) - z_k) \frac{\partial^2 \hat{J}_k}{\partial u_i \partial u_j}(u) + \frac{\partial \hat{J}_k}{\partial u_i}(u) \frac{\partial \hat{J}_k}{\partial u_j}(u)\right). \end{aligned}$$

Now we choose an arbitrary vector $v = (v_i)_{1\le i\le m}$ in $\mathcal{U}$. Then, $w = \nabla^2 F_z(u) v$ is a vector in $\mathcal{U}$ and

$$\begin{aligned} \big(\nabla^2 F_z(u) v\big)_i &= \sum_{j=1}^m \left(\frac{\partial^2 F_z}{\partial u_i \partial u_j}(u) v_j\right) \\ &= \sum_{k=1}^2 \left((\hat{J}_k(u) - z_k) \sum_{j=1}^m \Big(\frac{\partial^2 \hat{J}_k}{\partial u_i \partial u_j}(u) v_j\Big)\right) \\ &\quad + \sum_{k=1}^2 \left(\frac{\partial \hat{J}_k}{\partial u_i}(u) \sum_{j=1}^m \Big(\frac{\partial \hat{J}_k}{\partial u_j}(u) v_j\Big)\right) \\ &= \sum_{k=1}^2 \Big((\hat{J}_k(u) - z_k)\big(\nabla^2 \hat{J}_k(u) v\big)_i\Big) \\ &\quad + \sum_{k=1}^2 \Big(\big(\nabla \hat{J}_k(u)\big)_i \big(\nabla \hat{J}_k(u)^\top v\big)\Big). \end{aligned}$$

Consequently, we have

$$\begin{aligned} \nabla^2 F_z(u) v &= \sum_{k=1}^2 \Big((\hat{J}_k(u) - z_1)\big(\nabla^2 \hat{J}_k(u) v\big)\Big) \\ &\quad + \sum_{k=1}^2 \Big(\langle \nabla \hat{J}_k(u), v \rangle_{\mathcal{U}} \nabla \hat{J}_k(u)\Big) \in \mathcal{U}. \end{aligned}$$

We are interested in whether the second derivative of F_z is coercive at the optimal solution $\bar{u}_z \in \mathcal{U}_{\mathsf{ad}}$. We set $\kappa = \min\{\hat{J}_1 - z_1, \hat{J}_2 - z_2\} > 0$; cf. (5). Let $v \in \mathcal{U}$ be chosen arbitrarily. Then we estimate

$$\begin{aligned} &\langle \nabla^2 F_z(u) v, v \rangle_{\mathcal{U}} \\ &= \sum_{k=1}^2 \Big((\hat{J}_k(u) - z_k) \langle \nabla^2 \hat{J}_k(u) v, v \rangle_{\mathcal{U}} + \underbrace{\big|\langle \nabla \hat{J}_k(u), v \rangle_{\mathcal{U}}\big|^2}_{\ge 0}\Big) \\ &\ge \kappa \sum_{i=1}^2 \langle \nabla^2 \hat{J}_k(u) v, v \rangle_{\mathcal{U}}. \end{aligned}$$

Thus, if for $k = 1, 2$ the Hessians $\nabla^2 \hat{J}_k(\bar{u}_z)$ are positive semidefinite and at least one of them postive definite, we obtain that $\nabla F_z(\bar{u}_z)$.

REFERENCES

R. Dautray and J.-L. Lions: *Mathematical Analysis and Numerical Methods for Science and Technology. Volume 5: Evolution Problems I.* Springer-Verlag, Berlin, 2000.

M. Ehrgott: *Multicriteria Optimization.* Springer, Berlin, 2005.

S. Peitz, S. Oder-Blöbaum and M. Dellnitz: Multiobjective optimal control methods for fluid flow using reduced order modeling. http://arxiv.org/pdf/1510.05819v2.pdf, 2015.

L. Zadeh. Optimality and non-scalar-valued performance criteria. *IEEE Transactions on Automatic Control*, 8, 1963.