Exercise 1 (Linear operators in multiple components, Jacobian, KKT-conditions)**.** Let $X_1, \dots, X_n$ and $Z_1, \dots, Z_m$ be Banach spaces and set $X := X_1 \times \cdots \times X_n$ as well as $Z := Z_1 \times \cdots \times Z_m$. Consider an operator $A \in \mathcal{L}(X; Z)$.

(a) Show that A uniquely corresponds to an $m \times n$-operator-matrix $\mathcal{A} = (A_{ij})$ of continuous linear operators $A_{ij} \in \mathcal{L}(X_j; Z_i)$ such that

$$Ax = \mathcal{A}\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad \text{for } x = (x_1, \dots, x_n) \quad \text{with } x_i \in X_i,$$

and that $A \mapsto \sum_{i=1}^m \max_{1 \le j \le n} \|A_{ij}\|_{\mathcal{L}(X_j; Z_i)}$ is an equivalent norm to $\|\cdot\|_{\mathcal{L}(X;Z)}$.

(b) Show that $X^* = X_1^* \times \cdots \times X_n^*$ and $Z^* = Z_1^* \times \cdots \times Z_m^*$ and determine the operator-matrix corresponding to $A^* \in \mathcal{L}(Z^*; X^*)$.

(c) Let $G\colon X \to Z$ be F-differentiable around $\bar{x} \in X$. Show that the operator-matrix $\mathcal{G}'(\bar{x})$ of $G'(\bar{x})$ is exactly a generalized Jacobian matrix of G in $\bar{x}$.

(d) Let $(\bar{y}, \bar{u})$ be a regular point of the control-constrained optimal control problem

$$\min_{(y,u) \in Y \times U} J(y, u) \quad \text{s.t.} \quad E(y, u) = 0, \quad u \in U_{\text{ad}},$$

where $J\colon Y \times U \to \mathbb{R}$ and $E\colon Y \times U \to Z$ are F-differentiable, Y, U, Z are Banach spaces, and U_{ad} is closed and convex. Apply the above results to the multiplier rule in the KKT-conditions of this problem for $(\bar{y}, \bar{u})$.

Remark: Recall (or verify) that every norm $\|\cdot\|_\alpha$ on $\mathbb{R}^n$ constructed in the form $\|\mathbf{x}\| = f(|\mathbf{x}_1|, \dots, |\mathbf{x}_n|)$ for $\mathbf{x} \in \mathbb{R}^n$ also gives rise to a norm $\|x\|_{\alpha,X} = f(\|x_1\|_{X_1}, \dots, \|x_n\|_{X_n})$ on X, for example $\|(x_1, \dots, x_n)\|_{1,X} := \sum_{i=1}^n \|x_i\|_{X_i}$, and all these norms are equivalent because the ones on $\mathbb{R}^n$ are; an analogous result of course holds for Z and $\mathbb{R}^m$. For convenience, we always choose the norm induced by the $\|\cdot\|_1$-norm on the finite-dimensional space.

Solution.

(a) We define the operator $A_{ij}\colon X_j \to Z_i$ by

$$A_{ij}x_j := \big(A(0, \dots, 0, x_j, 0, \dots, 0)\big)_i.$$

This operator is clearly linear and thanks to

$$\begin{aligned}\|A_{ij}x_j\|_{Z_i} &= \left\| \big(A(0,\dots,0,x_j,0,\dots,0)\big)_i \right\|_{Z_i} \\ &\le \|A\|_{\mathcal{L}(X_j;Z)} \|(0,\dots,x_j,\dots)\|_{X_j} = \|A\|_{\mathcal{L}(X_j;Z)} \|x_j\|_{X_j}\end{aligned}$$

continuous with $\|A_{ij}\|_{\mathcal{L}(X_j;Z_i)} \le \|A\|_{\mathcal{L}(X;Z)}$. (Recall that we use the $\|\cdot\|_1$-norms on X and Z.) Setting

$$\mathcal{A} = \begin{pmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{m1} & \cdots & A_{mn} \end{pmatrix},$$

we find, where $A_j \in \mathcal{L}(X_j;Z)$ are given by the columns of $\mathcal{A}$,

$$\mathcal{A}x := \begin{pmatrix} A_1 & \cdots & A_n \end{pmatrix} x := \sum_{j=1}^{n} A_j x_j = Ax,$$

which proves the unique correspondence between the operator-matrix $\mathcal{A}$ and A.

For the norm equivalence, we further observe that by construction

$$\begin{aligned}\|Ax\|_Z = \sum_{i=1}^{m} \|(Ax)_i\|_{Z_i} &= \sum_{i=1}^{m} \left\| \sum_{j=1}^{n} A_{ij}x_j \right\|_{Z_i} \\ &\le \sum_{i=1}^{m}\sum_{j=1}^{n} \|A_{ij}x_j\|_{Z_i} \le \sum_{i=1}^{m}\sum_{j=1}^{n} \|A_{ij}\|_{\mathcal{L}(X_j;Z_i)} \|x_j\|_{X_j}\end{aligned}$$

and thus

$$\|Ax\|_Z \le \sum_{i=1}^{m} \max_{1\le j\le n} \|A_{ij}\|_{\mathcal{L}(X_j;Z_i)} \sum_{j=1}^{n} \|x_j\|_{X_j} = \left(\sum_{i=1}^{m} \max_{1\le j\le n} \|A_{ij}\|_{\mathcal{L}(X_j;Z_i)} \right) \|x\|_X,$$

so $\|A\|_{\mathcal{L}(X;Z)} \le \sum_{i=1}^{m} \max_{1\le j\le n} \|A_{ij}\|_{\mathcal{L}(X_j;Z_i)}$. Since we also had $\|A_{ij}\|_{\mathcal{L}(X_j;Z_i)} \le \|A\|_{\mathcal{L}(X;Z)}$ as above, this shows that

$$\|A\|_{\mathcal{L}(X;Z)} \le \sum_{i=1}^{m} \max_{1\le j\le n} \|A_{ij}\|_{\mathcal{L}(X_j;Z_i)} \le m\|A\|_{\mathcal{L}(X;Z)}. \tag{1}$$

(b) Let $x' \in X^* = \mathcal{L}(X;\mathbb{R})$. We have already observed that we may uniquely identify x' with $(x'_1,\dots,x'_n)$, where $x'_j \in \mathcal{L}(X_i;\mathbb{R}) = X_j^*$, via

$$\langle x', x\rangle_{X^*,X} = \sum_{j=1}^{n} \langle x'_j, x_j\rangle_{X_j^*,X_j}.$$

Moreover, per (1), we have $\|x^*\|_{X^*} = \max_{1\le j\le n} \|x_j^*\|_{X_j^*}$, which is exactly $\|\cdot\|_{\infty,X_1^*\times\cdots\times X_n^*}$ and thus equivalent to the $\|\cdot\|_1$-norm on $X_1^* \times \cdots \times X_n^*$.

Now let us consider the operator $A^* \in \mathcal{L}(Z^*; X^*)$. It has the fundamental defining property that

$$\langle z', Ax \rangle_{Z^*,Z} = \langle A^* z', x \rangle_{X^*,X} \quad \text{for all } x \in X,\ z' \in Z^*. \tag{2}$$

Identifying the operators and spaces in their "matrix format", we find

$$\begin{aligned}\langle z', Ax \rangle_{Z^*,Z} = \sum_{j=1}^{n} \langle z', A_j x_j \rangle_{Z^*,Z} &= \sum_{j=1}^{n}\sum_{i=1}^{m} \langle z_i', (A_j x_j)_i \rangle_{Z_i^*,Z_i} \\ &= \sum_{j=1}^{n}\sum_{i=1}^{m} \langle z_i', A_{ij} x_j \rangle_{Z_i^*,Z:i} = \sum_{j=1}^{n}\sum_{i=1}^{m} \langle A_{ij}^* z_i', x_j \rangle_{X_j^*,X_j}.\end{aligned}$$

By (2), the latter is nothing else than $\langle A^* z', x \rangle_{X^*,X}$. Carefully reading off indices and comparing, we find that the matrix $\mathcal{A}^*$ corresponding to A^* is obtained from $\mathcal{A}$ by transposing $\mathcal{A}$ and taking adjoint operators, so

$$A \sim \mathcal{A} = \begin{pmatrix} A_{11} & \dots & A_{1n} \\ \vdots & \ddots & \vdots \\ A_{m1} & \dots & A_{mn} \end{pmatrix} \quad \rightsquigarrow \quad A^* \sim \mathcal{A}^* = \begin{pmatrix} A_{11}^* & \dots & A_{m1}^* \\ \vdots & \ddots & \vdots \\ A_{1n}^* & \dots & A_{mn}^* \end{pmatrix}.$$

(c) We consider without loss of generality $m = 1$. Denoting $e_j = (0, \dots, 0, h_j, 0, \dots, 0)$ for $h_j \in X_j$, we have seen that $G'(\bar{x})_j h_j$ is then exactly given by $G'(\bar{x}) e_j$. The derivative $G'(\bar{x})$ of G in $\bar{x}$ is defined as the operator which satisfies

$$G(\bar{x} + h) = G(\bar{x}) + G'(\bar{x})h + o(\|h\|_X), \quad h \in X.$$

Inserting $h = e_j$, this shows that

$$G(\bar{x} + e_j) = G(\bar{x}) + G'(\bar{x})_j h_j + o(\|h_j\|_{X_j}).$$

On the other hand, the partial derivative $G'_{x_j}(\bar{x})$ of G in $\bar{x}$ in direction of the jth variable is given exactly as the derivative in 0 of the function $h_j \mapsto G^j(h_j) := G(\bar{x} + e_j)$ which yields

$$G(\bar{x}+e_j) = G^j(h_j) = G^j(0) + G'_{x_j}(\bar{x})h_j + o(\|h_j\|_{X_j}) = G(\bar{x}) + G'_{x_j}(\bar{x})h_j + o(\|h_j\|_{X_j}).$$

From the uniqueness of the derivative, this shows that indeed $G'(\bar{x})_j = G'_{x_j}(\bar{x})$ and hence

$$\mathcal{G}'(\bar{x}) = \begin{pmatrix} G'_{x_1}(\bar{x}) & \cdots & G'_{x_n}(\bar{x}) \end{pmatrix}. \tag{3}$$

(d) The multiplier rule in the KKT conditions for the given optimal control problem states that there exists a Lagrange multiplier $\lambda \in (W \times U)^*$ such that, where $G(y, u) = \binom{E(y,u)}{u} : Y \times U \to W \times U$,

$$J'(\bar{y}, \bar{u}) + G'(\bar{y}, \bar{u})^* \bar{\lambda} = 0 \quad \text{in } (Y \times U)^*.$$

Since $(Y \times U)^* = Y^* \times U^*$, the equation consists of two components. From the foregoing exercise and (3), $G'(\bar{y}, \bar{u})$ can be written in the form of a Jacobian:

$$G'(\bar{y}, \bar{u}) = \begin{pmatrix} E'_y(\bar{y}, \bar{u}) & E'_u(\bar{y}, \bar{u}) \\ 0 & \mathrm{id}_U \end{pmatrix}$$

We have also seen that the adjoint operator $G'(\bar{y}, \bar{u})^*$ is then given in matrix-form by

$$G'(\bar{y}, \bar{u})^* = \begin{pmatrix} E'_y(\bar{y}, \bar{u})^* & 0 \\ E'_u(\bar{y}, \bar{u})^* & \mathrm{id}_{U^*} \end{pmatrix}.$$

Writing $\bar{\lambda} = (\bar{p}, \bar{\mu}) \in (W \times U)^* = W^* \times U^*$, we thus find

$$J'(\bar{y}, \bar{u}) + G'(\bar{y}, \bar{u})^* \bar{\lambda} = \begin{pmatrix} J'_y(\bar{y}, \bar{u}) \\ J'_u(\bar{y}, \bar{u}) \end{pmatrix} + \begin{pmatrix} E'_y(\bar{y}, \bar{u})\bar{p} \\ E'_u(\bar{y}, \bar{u})\bar{p} + \bar{\mu} \end{pmatrix},$$

This is exactly the form given in the lecture notes.

Exercise 2 (Lax-Milgram lemma and divergence-gradient operators)**.** Let H be a Hilbert space and consider a continuous coercive bilinear form $a\colon H \times H \to \mathbb{R}$ on H, which means that there exist constants $C, \alpha > 0$ such that

$$|a(u,v)| \leq C\|u\|_H \|v\|_H \qquad \text{for all } u, v \in H \qquad \text{(continuity/boundedness)}$$

and

$$a(u,u) \geq \alpha \|u\|_H^2 \qquad \text{for all } u \in H \qquad \text{(coercivity).}$$

(a) Prove the world-famous *Lax-Milgram lemma*: For every $f \in H^*$, there exists a unique $u = u_f \in H$ such that

$$a(u,v) = \langle f, v \rangle_{H^*,H} \quad \text{for all } v \in H$$

and there holds $\|u_f\|_H \leq \alpha^{-1} \|f\|_{H^*}$.

Hints:

(i) Recall the also world-famous *Fréchet-Riesz representation theorem*: There is a continuous linear isometric isomorphism $T \in \mathcal{L}(H^*; H)$ such that, for all $g \in H^*$, we have $\langle g, v \rangle_{H^*,H} = (Tg, v)_H$ for all $v \in H$.

(ii) Let $M \subseteq H$. Then $(u,v)_H = 0$ for all $u \in M$ implies $v = 0$ if and only if M is dense in H. (Prove this if needed!)

(b) Let $\Omega \subset \mathbb{R}^n$ be a bounded domain and let $\mu \in L^\infty(\Omega; \mathbb{S}_n)$, where $\mathbb{S}_n$ is the set of symmetric real $n \times n$-matrices equipped with the operator-norm inherited from $\|\cdot\|_2$ on $\mathbb{R}^n$.

(i) Show that the *weak divergence-gradient operator* A_μ given by

$$\langle A_\mu u, v\rangle := \int_\Omega (\mu \nabla u) \cdot \nabla v \, \mathrm{d}x \quad \text{for all } v \in H_0^1(\Omega)$$

for $u \in H_0^1(\Omega)$ is a linear continuous operator $H_0^1(\Omega) \to H^{-1}(\Omega) = H_0^1(\Omega)^*$.

(ii) Suppose that there is $\mu_0 > 0$ such that μ additionally satisfies

$$v^T \mu v \geq \mu_0 \|v\|_2^2 \quad \text{for all } v \in \mathbb{R}^n \quad \text{for almost all } x \in \Omega.$$

Show that then for every $f \in H^{-1}(\Omega)$ there is a unique solution $u = u_f \in H_0^1(\Omega)$ of the weak formulation

$$\int_\Omega (\mu \nabla u) \cdot \nabla v \, \mathrm{d}x = \langle f, v\rangle_{H^{-1}(\Omega), H_0^1(\Omega)} \quad \text{for all } v \in H_0^1(\Omega)$$

of the elliptic second-order partial differential equation

$$\begin{aligned} -\operatorname{div}(\mu \nabla u) &= f && \text{in } \Omega, \\ u &= 0 && \text{on } \partial\Omega. \end{aligned}$$

(This equation is to be seen formally, because μ and f are too general for the equation to be interpreted in a classic sense.) The function $u = u_f$ moreover satisfies $\|u_f\|_{H_0^1(\Omega)} \leq \mu_0^{-1} \|f\|_{H^{-1}(\Omega)}$, so $A_\mu^{-1} \in \mathcal{L}(H^{-1}(\Omega); H_0^1(\Omega))$, and it is also the unique solution of the minimization problem

$$\min_{w \in H_0^1(\Omega)} \frac{1}{2} \int_\Omega (\mu \nabla w) \cdot \nabla w \, \mathrm{d}x - \langle f, w\rangle_{H^{-1}(\Omega), H_0^1(\Omega)}. \tag{4}$$

Hint: Recall that $u \mapsto \|\nabla u\|_{L^2(\Omega)}$ is an equivalent norm on $H_0^1(\Omega)$.

Solution.

(a) First the second hint: The mapping $u \mapsto (u, v)_H$ defines a continuous linear functional on H with norm $\|v\|_H$: Clearly, its norm is less or equal to $\|v\|_H$ due to Cauchy-Schwarz, but inserting v itself shows that it is indeed exactly $\|v\|_H$. From $(u, v)_H = 0$ for all $u \in M$ it then follows that this functional is the zero functional and thus $\|v\|_H = 0$ if M is dense in H. For the reverse implication, assume that M is not dense in H. Then the Hahn-Banach theorem implies the existence of a functional $0 \neq \varphi \in H^*$ such that $\langle \varphi, u\rangle_{H^*,H} = 0$ for all $u \in \overline{M}$. From the first hint, we know that there exists an operator $T \in \mathcal{L}(H^*; H)$ such that $\langle \varphi, u\rangle_{H^*,H} = (T\varphi, u)_H$. But then $v = T\varphi \neq 0$ satisfies $(u, v)_H = 0$ for all $u \in M$, which is a contradiction.

Now the Lax-Milgram lemma: For every $u \in H$, boundedness and bilinearity of a implies that $v \mapsto a(u, v)$ is a continuous linear functional on H whose norm is bounded by $C\|u\|_H$ and that $u \mapsto [v \mapsto a(u, v)]$ is a continuous linear mapping from H to H^* whose norm is bounded by C. We denote this mapping by $B \in \mathcal{L}(H;H^*)$ and set $A := TB \in \mathcal{L}(H)$. Then we have, for all $u, v \in H$,

$$(Au, v)_H = (TBu, v)_H = \langle Bu, v\rangle_{H^*,H} = a(u, v)$$

and

$$\langle f, v\rangle_{H^*,H} = (Tf, v)_H.$$

Hence, $u = u_f \in H$ is the unique solution to

$$a(u, v) = \langle f, v\rangle_{H^*,H} \quad \text{for all } v \in H$$

with $\|u_f\|_H \leq \alpha^{-1}\|f\|_{H^*}$ if and only if $u = A^{-1}Tf = B^{-1}f$ and $\|B^{-1}\|_{\mathcal{L}(H^*;H)} \leq \alpha^{-1}$. So, we need to show that B—or equivalently A, because T is an isometric isomorphism—is bijective and its inverse is continuous.

We will derive this from the the coercivity of a, which means that

$$(Au, u)_H = a(u, u) \geq \alpha\|u\|_H^2 \quad \text{for all } u \in H,$$

by showing that A is injective and its range $\operatorname{Ran} A$ is dense and closed in H (and thus must be H). The first consequence of coercivity is injectivity: Indeed, assume that there is $u \in H$ such that $Au = 0$. Then $(Au, u) = 0$ which is a contradiction. Further, coercivity also implies that the image $\operatorname{Ran} A$ of A is dense in H: Let $v \in H$ be given and assume that $(Au, v)_H = 0$ for all $u \in H$. Then it follows that $v = 0$, since otherwise $(Av, v) > 0$ due to coercivity, which by the second hint implies that $\operatorname{Ran} A$ is dense in H. Finally, $\operatorname{Ran} A$ must also be *closed* in H: The coercivity property again implies that $\alpha\|u\|_H^2 \leq (Au, u)_H = \|Au\|_H\|u\|_H$, so $\|Au\|_H \geq \alpha\|u\|_H$, for all $u \in H$. Let (Au_k) be a sequence in $\operatorname{Ran} A$ which converges to some $v \in H$. We need to show that $v \in \operatorname{Ran} A$, i.e., there exists some $u \in H$ such that $v = Au$. Due to $\|Au_k - Au_\ell\|_H \geq \alpha\|u_k - u_\ell\|_H$, the sequence (u_k) is a Cauchy sequence and thus convergent to some $u \in H$. But then continuity of A implies $v = Au$, so $\operatorname{Ran} A$ is closed.

Altogether, A is bijective and thus, by the open mapping theorem, continuously invertible with $A^{-1} \in \mathcal{L}(H)$. The norm estimate for A^{-1} follows again from $\|Au\|_H \geq \alpha\|u\|_H$ for all $u \in H$, because using $v = A^{-1}u$ we have

$$\|A^{-1}u\|_H = \|v\|_H \leq \alpha^{-1}\|Av\|_H = \alpha^{-1}\|u\|_H \quad \text{for all } u \in H.$$

Note how we have derived continuous invertibility alone from the coercivity property of A. This argument is not limited to the operator A at hand, but works for every operator satisfying such a coercivity property.

(b) (i) Linearity is obvious and continuity follows quite immediately from by two applications of the Cauchy-Schwarz inequality, once in $\mathbb{R}^n$ and once in $L^2(\Omega)$:

$$\begin{aligned}\left|\int_\Omega (\mu\nabla u)\cdot\nabla v \,\mathrm{d}x\right| &\leq \|\mu\|_{L^\infty(\Omega;\mathbb{S}_n)}\int_\Omega \|\nabla u\|_2\|\nabla v\|_2 \,\mathrm{d}x \\ &\leq \|\mu\|_{L^\infty(\Omega;\mathbb{S}_n)}\|\nabla u\|_{L^2(\Omega)}\|\nabla v\|_{L^2(\Omega)}, \end{aligned} \tag{5}$$

and the observation that $\|\nabla u\|_{L^2(\Omega)}$ and $\|\nabla v\|_{L^2(\Omega)}$ are smaller than or equal to $\|u\|_{H_0^1(\Omega)}$ and $\|v\|_{H_0^1(\Omega)}$ (in the present case with zero boundary data, the expressions are in fact equivalent, see the hint for the second part of this exercise).

(ii) We show that the ellipticity assumption on μ implies that $(u, v) \mapsto \langle A_\mu u, v\rangle$ satisfies the assumptions in the Lax-Milgram lemma with $H = H_0^1(\Omega)$. Boundedness was already shown in (5), and coercivity can be seen as follows:

$$\langle A_\mu u, u\rangle = \int_\Omega (\mu \nabla u) \cdot \nabla u \, \mathrm{d}x \\ \geq \mu_0 \int_\Omega \|\nabla u\|_2^2 \, \mathrm{d}x = \mu_0 \|\nabla u\|_{L^2(\Omega)}^2 = \mu_0 \|u\|_{H_0^1(\Omega)}^2.$$

The Lax-Milgram lemma then yields the assertions about unique solvability and the norm stability estimate.

Finally, define $F\colon H_0^1(\Omega) \to \mathbb{R}$ by

$$F(w) = \frac{1}{2} \int_\Omega (\mu \nabla w) \cdot \nabla w \, \mathrm{d}x - \langle f, w\rangle_{H^{-1}(\Omega), H_0^1(\Omega)} \quad \text{for } w \in H_0^1(\Omega).$$

The function $\bar{u} \in H_0^1(\Omega)$ is a solution of the minimization problem (4) if and only if $F'(\bar{u}) = 0$ in $H^{-1}(\Omega)$, or equivalently $F'(\bar{u})v = 0$ for all $v \in H_0^1(\Omega)$. Since the integral in the definition in F is exactly $\frac{1}{2}a(w, w)$, we can rely on Example 3.3 in the lecture notes which says that the derivative of this function in w is exactly $v \mapsto a(w, v)$, and thus find

$$F'(w)v = \int_\Omega (\mu \nabla w) \cdot \nabla v \, \mathrm{d}x - \langle f, v\rangle_{H^{-1}(\Omega), H_0^1(\Omega)} \quad \text{for all } v \in H_0^1(\Omega).$$

From this expression it is obvious that $u_f = \bar{u}$, the solution of (4), because then $F'(u_f)v = 0$ for all $v \in H_0^1(\Omega)$, and we have already seen that u_f is the unique function in $H_0^1(\Omega)$ with this property.

Exercise 3 (Projection formula for the optimal control)**.** Consider a bounded domain $\Omega \subset \mathbb{R}^n$ and the optimal control problem

$$\begin{aligned} \min_{(y,u) \in H_0^1(\Omega) \times L^2(\Omega)} \; & \frac{1}{2} \int_\Omega |y - y_d|^2 \, \mathrm{d}x + \frac{\beta}{2} \int_\Omega |u|^2 \, \mathrm{d}x \\ \text{s.t.} \quad & Ay = \mathcal{E}u \ \text{ in } H^{-1}(\Omega) \end{aligned} \tag{Ell-OCP}$$

with $A \in \mathcal{L}(H_0^1(\Omega); H^{-1}(\Omega))$ and $A^{-1} \in \mathcal{L}(H^{-1}(\Omega); H_0^1(\Omega))$; imagine the divergence-gradient operators from exercise 2. Moreover, $\mathcal{E} \in \mathcal{L}(L^2(\Omega); H^{-1}(\Omega))$ denotes the embedding $L^2(\Omega) \hookrightarrow H^{-1}(\Omega)$ and we have $y_d \in L^2(\Omega)$ and $\beta > 0$.

(a) Show that every feasible pair (y, u) is regular.

(b) Let $(\bar{y}, \bar{u})$ be a locally optimal solution of (Ell-OCP). Show that the optimal control $\bar{u}$ is given by

$$\bar{u}(x) = -\beta^{-1}\bar{p}(x) \quad \text{for almost all } x \in \Omega,$$

where $\bar{p} \in H_0^1(\Omega)$ satisfies $A^*\bar{p} = \bar{y} - y_d$. What does this imply for the regularity of $\bar{u}$? What if we can show higher H^2-regularity properties for A and/or A^* as in the example in the lecture notes?

(c) Now assume that there are also control constraints of the form

$$u \in U_{\text{ad}} = \{w \in L^2(\Omega) \colon a \leq w \leq b \text{ a.e. on } \Omega\}$$

in (Ell-OCP), with $L^2(\Omega)$-functions $a \leq b$. Show that the optimal control $\bar{u}$ then satisfies

$$\bar{u}(x) = \operatorname{proj}_{[a(x),b(x)]}\big(-\beta^{-1}\bar{p}(x)\big) \quad \text{for almost all } x \in \Omega.$$

Make an educated guess about the regularity of $\bar{u}$ in this case and how an analogous result to the (control-) unconstrained case could be achieved.

Solution.

(a) We set $X = H_0^1(\Omega) \times L^2(\Omega)$ and $Z = H^{-1}(\Omega)$ together with $G(x) = Ay - \mathcal{E}u$ and $K = \{0_{H^{-1}(\Omega)}\}$.

Then $G'(\bar{x}) = A - \mathcal{E}$ is surjective for every $\bar{x} \in X$ because $A \in \mathcal{L}(H_0^1(\Omega); H^{-1}(\Omega))$ is so: For every $f \in H^{-1}(\Omega)$ there exists $y \in H_0^1(\Omega)$ such that $Ay = f$ and thus $G'(\bar{x})(y, 0) = Ay = f$. By Proposition 3.18 in the lecture notes, surjectivity is a constraint qualification.

(b) Since $(\bar{y}, \bar{u})$ is regular by the foregoing exercise, we know that the KKT conditions must be satisfied in $(\bar{y}, \bar{u})$: There exists a Lagrange multiplier $\bar{\lambda} \in (H^{-1}(\Omega))^* = H_0^1(\Omega)$ such that

$$f'(\bar{x}) + G'(\bar{x})^*\bar{\lambda} = 0 \quad \text{in } H^{-1}(\Omega) \times L^2(\Omega),$$

where we have the Jacobian representation

$$G'(\bar{x}) = \begin{pmatrix} A & -\mathcal{E} \end{pmatrix}, \quad \text{so} \quad G'(\bar{x})^* = \begin{pmatrix} A^* \\ -\mathcal{E}^* \end{pmatrix}.$$

Identifying

$$f \colon X \to \mathbb{R}, \quad f(x) = J(y, u) = \frac{1}{2}\int_\Omega |Ey - y_d|^2 \,\mathrm{d}x + \frac{\beta}{2}\int_\Omega |u|^2 \,\mathrm{d}x,$$

where $E \in \mathcal{L}(H_0^1(\Omega); L^2(\Omega))$ is the embedding $H_0^1(\Omega) \hookrightarrow L^2(\Omega)$, we find

$$f'(\bar{x}) = J'(\bar{y}, \bar{u}) = \begin{pmatrix} E(\bar{y} - y_d) \\ \beta\bar{u} \end{pmatrix},$$

thus the KKT condition translates to (component-wise)

$$E(\bar{y} - y_d) + A^*\bar{\lambda} = 0 \quad \iff \quad A^*\bar{p} = E(\bar{y} - y_d) \quad \text{in } H^{-1}(\Omega)$$

with $\bar{p} := -\bar{\lambda}$ and

$$\beta\bar{u} - \mathcal{E}^*\bar{\lambda} = \quad \iff \quad \bar{u} = -\beta^{-1}\mathcal{E}^*\bar{p} \quad \text{in } L^2(\Omega).$$

Here, $\mathcal{E}^* = E$ is exactly the embedding $H_0^1(\Omega) \hookrightarrow L^2(\Omega)$ (why?), hence

$$\bar{u}(x) = -\beta^{-1}\bar{p}(x) \quad \text{for almost all } x \in \Omega$$

follows. In particular, $\bar{u}$ is also an $H_0^1(\Omega)$ function!

Concerning regularity: Since the right-hand side in the equation for $\bar{p}$ is in fact from $L^2(\Omega)$, we would even get $\bar{p}$ and thus $\bar{u} \in H^2(\Omega) \cap H_0^1(\Omega)$ if we had higher regularity for A^*. Note that we could also "bootstrap" the regularity of $\bar{y}$ if A itself admitted higher regularity, since $\bar{y}$ satisfies $A\bar{y} = \bar{u} \in L^2(\Omega)$, from which we could get $\bar{y} \in H^2(\Omega) \cap H_0^1(\Omega)$.

If y_d was in fact *also* more regular than $L^2(\Omega)$, improved regularity for $\bar{y}$ would imply that the right-hand side in $A^*\bar{p} = \bar{y} - y_d$ is better, and then we could hope for more regularity of $\bar{p}$, implying even more for $\bar{u}$, thus for $\bar{y}$ and so on …This shows that higher-regularity results can be particularly useful in an optimal control setting because all occurring quantities are linked by differential operators which allows to bootstrap regularities.

(c) Coming from the foregoing exercise, we modify G to $G(x) = \binom{Ay - \mathcal{E}u}{u}$ and Z to $Z = H^{-1}(\Omega) \times L^2(\Omega)$ as well as $K = \{0_{H^{-1}(\Omega)}\} \times U_{\text{ad}}$. We have already seen in the lecture notes that every feasible point is still regular due to the surjectivity of the second component of G (and the properties of A); this was in Example 3.37. Note moreover that, using the Jacobian representation,

$$G'(\bar{x}) = \begin{pmatrix} A & -\mathcal{E} \\ 0 & \text{id}_{L^2(\Omega)} \end{pmatrix}, \quad \text{so} \quad G'(\bar{x})^* = \begin{pmatrix} A^* & 0 \\ -\mathcal{E}^* & \text{id}_{L^2(\Omega)} \end{pmatrix}.$$

For the new problem, there now exists a Lagrange multiplier pair $(\bar{p}, \bar{\mu}) \in H_0^1(\Omega) \times L^2(\Omega)$ such that, using already the transformation from the foregoing exercise,

$$A^*\bar{p} = E(\bar{y} - y_d) \quad \text{in } H^{-1}(\Omega)$$

and

$$\bar{u} + \bar{\mu} = -\beta^{-1}\mathcal{E}^*\bar{p} \quad \text{in } L^2(\Omega). \tag{6}$$

Additionally, there is the constraint $\bar{\mu} \in T(U_{\text{ad}}, \bar{u})^\circ$. In the lecture notes (Example 3.37), we have already identified this polar cone to be

$$T(U_{\text{ad}}, \bar{u})^\circ = \Big\{ s \in L^2(\Omega) \colon s|_{[\bar{u}=a]} \leq 0,\ s_{[\bar{u}=b]} \geq 0,\ s_{[a<\bar{u}<b]} = 0 \Big\}.$$

We use this together with (6) to derive the projection formula:

Suppose that $\bar{u}(x) < -\beta^{-1}\bar{p}(x)$ for some $x \in \Omega$. This is equivalent to $\bar{\mu}(x) > 0$ which implies $\bar{u}(x) = b(x)$. Conversely, $\bar{u}(x) > -\beta^{-1}\bar{p}(x)$ is equivalent to $\bar{\mu}(x) < 0$ and thus we must already have $\bar{u}(x) = a(x)$. If $\bar{u}(x) = -\beta^{-1}\bar{p}(x)$, then $\bar{\mu}(x) = 0$ and we cannot say anything more than $a(x) \leq \bar{u}(x) \leq b(x)$, which we already know by feasibility of $\bar{u}$. We collect these properties and rewrite them to obtain the projection formula:

$$\bar{u}(x) = \begin{cases} b(x) & \text{if } b(x) < -\beta^{-1}\bar{p}(x), \\ a(x) & \text{if } a(x) > -\beta^{-1}\bar{p}(x), \\ -\beta^{-1}\bar{p}(x) & \text{if } a(x) \leq -\beta^{-1}\bar{p}(x) \leq b(x) \end{cases} = \operatorname{proj}_{[a(x),b(x)]}\big(-\beta^{-1}\bar{p}(x)\big).$$

Now, $\bar{u}$ given by this projection formula will in general not be an $H_0^1(\Omega)$ function, since a and b are only $L^2(\Omega)$ functions; an instructive way to see this is to imagine that $-\beta^{-1}\bar{p} \geq b$ and thus $\bar{u} = b$ on Ω. On the other hand, this way of thinking shows that there is hope if $a, b \in H_0^1(\Omega)$. If a, b are in fact constant, then one can show that $\bar{u}$ inherits the $H_0^1(\Omega)$-regularity quite immediately, but the general case is also true.

On the other hand, there is no hope to obtain higher H^2-regularity because the projection is generally "only" Lipschitz-continuous, and it is known that the composition of Lipschitz- and H^2-functions does not preserve H^2-regularity.

Exercise 4 (Partial ordering induced by pointed cone)**.** Let X be a Banach space and let $K \subset X$ be a closed convex and pointed cone, that is, $K \cap (-K) = \{0\}$. Show that the relation $\leq_K$ given by

$$x_1 \leq_K x_2 \quad \iff \quad x_2 - x_1 \in -K$$

is a *partial ordering*, that is, it is reflexive, anti-symmetric and transitive. Convince yourself that you are allowed to cancel positive factors α on both sides.

Solution. First of all if $\alpha x_1 \leq_K \alpha x_2$ for some $\alpha > 0$, then $\alpha(x_2 - x_1) \in -K$, and since K (and thus also $-K$) is a cone, $\frac{1}{\alpha}\alpha(x_2 - x_1) \in -K$, hence $\alpha x_1 \leq_K \alpha x_2$.

For the partial ordering property, we collect the three properties:

- **Reflexivity**: Since $x - x = 0 \in -K$ since K is *closed*, we have $x \leq_K x$ for every $x \in X$.
- **Anti-symmetry**: Let $x_1, x_2 \in X$ be related by $x_1 \leq_K x_2$ and $x_2 \leq_K x_1$. Then $x_2 - x_1 \in -K$ by the first relation and $x_1 - x_2 \in -K$ by the second. The latter means that $x_2 - x_1 \in K$, so from the *pointed* property of K we infer that $x_1 - x_2 = x_2 - x_1 = 0$, hence $x_1 = x_2$.

- **Transitivity**: Let $x_1, x_2, x_3 \in X$ be related by $x_1 \leq_K x_2$ and $x_2 \leq_K x_3$. Then we know that $x_2 - x_1 \in -K$ as well as $x_3 - x_2 \in -K$. Since K is a *convex* set, we infer that
$$\frac{1}{2}(x_3 - x_1) = \frac{1}{2}(x_3 - x_2) + \frac{1}{2}(x_2 - x_1) \in -K,$$
and since K is a *cone*, this also implies $x_3 - x_1 \in -K$. Hence, $x_1 \leq_K x_3$.

Note how we have used each property of the set K in the proof: Closedness (and the cone definition, but $0 \in K$ would have been sufficient) for reflexivity, pointedness for anti-symmetry and convexity and being a cone for transitivity.