Optimization in Function Spaces Exercise sheet **2**
Dr. Hannes Meinlschmidt October 29, 2020

Exercise 1 (Differentiability in Banach spaces)**.** Verify the assertions from Example 3.3 in the lecture notes and some more:

(a) Show that every bounded linear operator $A \in \mathcal{L}(X;Y)$ is F-differentiable and its derivative in every point $x \in X$ is given by A itself.

(b) Let $a\colon X \times X \to \mathbb{R}$ be a symmetric continuous bilinear form. Prove that the quadratic form given by $X \ni u \mapsto \frac{1}{2}a(u,u) \in \mathbb{R}$ is F-differentiable and its derivative in $u \in X$ is given by $h \mapsto a(u,h)$. Apply this to the function $u \mapsto \frac{1}{2}\|u\|_X^2$ for a Hilbert space X.

Hint: The bilinear form a is continuous if and only if there exists a number $C \geq 0$ such that $|a(u,v)| \leq C\|u\|_X\|v\|_X$ for all $u, v \in X$.

(c) We consider the superposition operator Ψ induced by $\sin\colon \mathbb{R} \to \mathbb{R}$.

(i) Show that Ψ is F-differentiable as a mapping from $L^\infty(0,1)$ into itself with the derivative given by $h \mapsto \cos(y)h$ for every $y \in L^\infty(0,1)$ (can you generalize this assertion to other inducing functions?), ...

Hint: Calculate pointwisely and use exact first-order Taylor approximation in integrated form.

(ii) ...but Ψ is *not* F-differentiable as a mapping from $L^p(0,1)$ into itself for any $1 \leq p < \infty$.

Hint: Determine the residual of the first-order approximation exactly for suitably chosen step functions h. Choosing $y \equiv 0$ also helps to clear the fog.

(iii) Guess and prove the relation between p and q such that Ψ is F-differentiable as a mapping from $L^q(0,1)$ to $L^p(0,1)$.

Hint: You may use without proof that a superposition operator mapping $L^q(0,1)$ to $L^p(0,1)$ for $1 \leq p, q < \infty$ is always automatically continuous.

(d) Let $\Omega \subset \mathbb{R}^n$ be an open and bounded set. Prove that the superposition operator Ξ induced by the real function $\varphi(t) := t^3$ is F-differentiable when considered as a mapping from $L^6(\Omega)$ to $L^2(\Omega)$ and its derivative is given by $L^6(\Omega) \ni h \mapsto 3y^2h \in L^2(\Omega)$.

Solution.

(a) The assertion follows immediately from $A \in \mathcal{L}(X;Y)$ and

$$A(x+h) - Ax - Ah = 0.$$

(b) From the properties of a, we derive

$$a(u+h,u+h) = a(u,u) + 2a(u,h) + a(h,h),$$

which on the one hand suggests that $h \mapsto a(u,h)$ will be the derivative of the quadratic form, and on the other hand shows that

$$\left\| \frac{1}{2}a(u+h,u+h) - \frac{1}{2}a(u,u) - a(u,h) \right\| = \frac{1}{2}\left\| a(h,h) \right\| \le \frac{C}{2}\|h\|^2,$$

so

$$\left\| \frac{1}{2}a(u+h,u+h) - \frac{1}{2}a(u,u) - a(u,h) \right\| = o\big(\|h\|\big).$$

Moreover, we have

$$\left| a(u,h) \right| \le C\|u\|\|h\|,$$

so $h \mapsto a(u,h)$ is a continuous linear operator from X to X, as required.

For the squared norm $j\colon u \mapsto \frac{1}{2}\|u\|_X$ on a Hilbert space X, we have $a(u,v) := (u,v)_X$, the scalar product in X, so $j(u) = \frac{1}{2}(u,u)_X$. With the assertion from the general case, this yields

$$j'(u)h = (u,h)_X \quad \text{or} \quad \nabla j(u) = u.$$

(c) Let $y, h \in L^\infty(0,1)$. We use first order Taylor expansion with exact remainder in integrated form and calculate for almost every $x \in (0,1)$

$$\begin{aligned}
\Psi(y+h)(x) &= \sin\big(y(x) + h(x)\big) \\
&= \sin(y(x)) + \cos(y(x))h(x) \\
&\quad + \int_0^1 \Big[\cos\big(y(x) + sh(x)\big) - \cos\big(y(x)\big)\Big]\, h(x)\, \mathrm{d}s.
\end{aligned}$$

(i) From the preceding formula it follows that

$$\begin{aligned}
&\big\|\Psi(y+h) - \Psi(y) - \cos(y)h\big\|_{L^\infty(0,1)} = \big\|\sin(y+h) - \sin(y) - \cos(y)h\big\|_{L^\infty(0,1)} \\
&= \operatorname{esssup}_{x\in(0,1)} \left| \int_0^1 \Big[\cos\big(y(x) + sh(x)\big) - \cos\big(y(x)\big)\Big]\, h(x)\, \mathrm{d}s \right| \\
&\le \operatorname{esssup}_{x\in(0,1)} \left|h(x)^2\right| \cdot \left(\int_0^1 s\, \mathrm{d}s\right) \\
&= \frac{1}{2}\|h\|^2_{L^\infty(0,1)} = O\big(\|h\|^2_{L^\infty(0,1)}\big) = o\big(\|h\|_{L^\infty(0,1)}\big),
\end{aligned}$$

where we have used that cos is globally Lipschitz continuous with Lipschitz constant 1. Since $h \mapsto \cos(y)h$ clearly maps $L^\infty(0,1)$ to $L^\infty(0,1)$ in a continuous and linear fashion, the claim for $L^\infty(0,1)$-differentiability of Ψ follows.

Here we have essentially only used the Lipschitz continuity of cos, so of the derivative of the function inducing the superposition operator Ψ. Indeed, one can show that if the functions f and f' are Lipschitz continuous, then the superposition operator induced by f is F-differentiable from L^∞ to L^∞.

(ii) To show non-differentiability of Ψ as an operator from $L^p(0,1)$ to $L^p(0,1)$ for $1 \le p < \infty$, we, as the hint suggests, take the zero function $y \equiv 0$ as well as $h \in L^p(0,1)$ to obtain

$$\begin{aligned}\sin\big(y(x)+h(x)\big) &= \sin(0)+\cos(0)h(x) \\ &\qquad + \int_0^1 \Big[\cos\big(0+sh(x)\big)-\cos(0)\Big]\, h(x)\,\mathrm{d}s \\ &:= h(x)+r(x).\end{aligned}$$

Choosing

$$h_\varepsilon(x) := \begin{cases} 1 & \text{if } 0 \le x \le \varepsilon, \\ 0 & \text{if } \varepsilon < x \le 1, \end{cases}$$

we can calculate r_ε explicitly: For $0 \le x \le \varepsilon$, we have

$$r_\varepsilon(x) = \int_0^1 \big[\cos(s)-1\big]\,\mathrm{d}s = \big[\sin(s)-s\big]_{s=0}^{s=1} = \sin(1)-1 \neq 0,$$

whereas $r_\varepsilon(x) = 0$ for for $\varepsilon < x \le 1$. For Ψ to be continuously differentiable between $L^p(0,1)$, we would need $\|r_\varepsilon\|_{L^p(0,1)} = o(\|h_\varepsilon\|_{L^p(0,1)})$ as $\varepsilon \to 0$, that is,

$$\lim_{\varepsilon\to 0} \frac{\|r_\varepsilon\|_{L^p(0,1)}}{\|h_\varepsilon\|_{L^p(0,1)}} = 0.$$

But

$$\frac{\|r_\varepsilon\|_{L^p(0,1)}}{\|h_\varepsilon\|_{L^p(0,1)}} = \frac{\varepsilon^{\frac{1}{p}}\big(\sin(1)-1\big)}{\varepsilon^{\frac{1}{p}}} = \sin(1)-1 \neq 0$$

for *all* $\varepsilon > 0$, which shows that F-differentiability for Ψ fails to hold here.

(iii) Observing the preceding argument for non-differentiability of Ψ between $L^p(0,1)$ and itself for $1 \le p < \infty$, one notes that

$$\frac{\|r_\varepsilon\|_{L^p(0,1)}}{\|h_\varepsilon\|_{L^q(0,1)}} = \frac{\varepsilon^{\frac{1}{p}}\big(\sin(1)-1\big)}{\varepsilon^{\frac{1}{q}}} = \varepsilon^{\frac{1}{p}-\frac{1}{q}}\big(\sin(1)-1\big) \longrightarrow 0 \quad \text{as} \quad \varepsilon \to 0$$

if $1 \le p < q < \infty$ (in fact, this is true also for $q = \infty$ with $\varepsilon^{\frac{1}{\infty}} = 1$). This suggests that this relation between p and q could work for F-differentiability from $L^q(0,1)$ to $L^p(0,1)$.

So, we go back to the exact Taylor expansion and obtain via Hölder's inequality

$$\begin{aligned}\big\|\Psi(y+h)-\Psi(y)-\cos(y)h\big\|^p_{L^p(0,1)} &= \big\|\sin(y+h)-\sin(y)-\cos(y)h\big\|^p_{L^p(0,1)} \\ = \int_0^1 \bigg|\int_0^1 \Big[\cos\big(y(x)+sh(x)\big)-\cos\big(y(x)\big)\Big]\, h(x)\,\mathrm{d}s\bigg|^p\,\mathrm{d}x \\ \le \|h\|^p_{L^q(0,1)}\cdot\|\psi\|^p_{L^r(0,1)}\end{aligned}$$

with $r := \frac{pq}{q-p}$ and

$$\psi(x) := \int_0^1 \Big[\cos\big(y(x) + sh(x)\big) - \cos\big(y(x)\big)\Big] \,\mathrm{d}s.$$

To obtain F-differentiability for Ψ mapping $L^q(0,1)$ to $L^p(0,1)$, we need to show that $h \to 0$ in $L^q(0,1)$ implies $\psi \to 0$ in $L^r(0,1)$.

Writing $\|\psi\|_{L^r(0,1)}$ explicitly, we find

$$\begin{aligned}
\|\psi\|_{L^r(0,1)} &= \left(\int_0^1 \left|\int_0^1 \Big[\cos\big(y(x) + sh(x)\big) - \cos\big(y(x)\big)\Big] \,\mathrm{d}s\right|^r \mathrm{d}x\right)^{\frac{1}{r}} \\
&\leq \left(\int_0^1 \int_0^1 \Big|\cos\big(y(x) + sh(x)\big) - \cos\big(y(x)\big)\Big|^r \,\mathrm{d}s \,\mathrm{d}x\right)^{\frac{1}{r}} \\
&= \left(\int_0^1 \big\|\cos(y + sh) - \cos(y)\big\|^r_{L^r(0,1)} \,\mathrm{d}s\right)^{\frac{1}{r}},
\end{aligned}$$

where we have used Hölder's (or Jensen's) inequality and Fubini's theorem. Now, since the superposition operator Φ induced by cos clearly maps $L^q(0,1)$ to $L^\infty(0,1)$ and $(0,1)$ has finite measure, Φ maps $L^q(0,1)$ to $L^r(0,1)$ and is thereby, thanks to the hint, automatically continuous. But this means that for every $\varepsilon > 0$ there exists $\delta = \delta(\varepsilon) > 0$ such that $\|\cos(y+h) - \cos(y)\|_{L^r(0,1)} < \varepsilon$ whenever $\|h\|_{L^q(0,1)} < \delta$. We infer that

$$\|\psi\|_{L^r(0,1)} < \varepsilon \quad \text{whenever} \quad \|h\|_{L^q(0,1)} < \delta(\varepsilon),$$

or in other words, $\psi \to 0$ in $L^r(0,1)$ whenever $h \to 0$ in $L^q(0,1)$. From this, F-differentiability of Ψ from $L^q(0,1)$ to $L^p(0,1)$ follows.

Here we have only used that cos, so the derivative of the function sin inducing the superposition operator Ψ, maps $L^q(0,1)$ to $L^r(0,1)$ and this is indeed a sufficient condition for the general case (together with a condition ensuring that Ψ maps $L^q(0,1)$ to $L^p(0,1)$ in the first place, of course – but we have learned to know such a condition already on the first exercise sheet).

(d) We have already seen in the lecture notes and the last exercise sheet that Ξ maps $L^6(\Omega)$ to $L^2(\Omega)$ due to $t^3 = t^{\frac{6}{2}}$. Again by pointwise calculation, we have for every $y, h \in L^6(\Omega)$

$$\Xi(y+h)(x) = \big(y(x) + h(x)\big)^3 = y(x)^3 + 3y(x)^2h(x) + 3y(x)h(x)^2 + h(x)^3$$

for almost every $x \in \Omega$. This shows that

$$\big\|\Xi(y+h) - \Xi(y) - 3y^2h\big\|_{L^2(\Omega)} \leq 3\big\|yh^2\big\|_{L^2(\Omega)} + \big\|h^3\big\|_{L^2(\Omega)}.$$

We use Hölder's inequality with $\frac{1}{3} + \frac{2}{3} = 1$ for the first term and rewrite the second one to an $L^6(\Omega)$ norm, from which we obtain

$$\begin{aligned}\big\|\Xi(y+h) - \Xi(y) - 3yh\big\|_{L^2(\Omega)} &\leq 3\|y\|_{L^6(\Omega)}\|h\|_{L^6(\Omega)}^2 + \|h\|_{L^6(\Omega)}^3 \\ &= O\big(\|h\|_{L^6(\Omega)}^2\big) = o\big(\|h\|_{L^6(\Omega)}\big).\end{aligned}$$

Since the derivative $h \mapsto 3y^2h$ clearly maps $L^6(\Omega)$ to $L^2(\Omega)$ in a linear and bounded fashion (use again Hölder's inequality with $\frac{2}{3} + \frac{1}{3} = 1$), this shows F-differentiability of Ξ.

Exercise 2 (Closedness of the tangential cone)**.** Let X be a Banach space and let $x \in M \subseteq X$. Show that the contingent cone $T(M, x)$ is closed.

Solution. Let $(d_k) \subseteq T(M, x)$ be a convergent sequence with limit $d \in X$. We need to show that $d \in T(M, x)$. By definition, for every k there exist sequences $(x_j^k) \subseteq M$ and $(\eta_j^k) > 0$ such that $x_j^k \to x$ and $\eta_j^k(x_j^k - x) \to d^k$, each as $j \to \infty$. We use a diagonal sequence to show that $d \in T(M, x)$. Depending on k, there exist numbers $j(k)$ such that

$$\big\|x_{j(k)}^k - x\big\| < \frac{1}{k} \quad \text{and} \quad \Big\|\eta_{j(k)}^k\big(x_{j(k)}^k - x\big) - d^k\Big\| < \frac{1}{k}$$

But then $y_k := x_{j(k)}^k$ converges to x and with $\mu_k := \eta_{j(k)}^k$ we have

$$\big\|\mu_k(y_k - x) - d\big\| \leq \big\|\mu_k(y_k - x) - d^k\big\| + \big\|d^k - d\big\| \longrightarrow 0 \quad \text{as } k \to \infty,$$

hence $(y_k) \subseteq M$ and $(\mu_k) > 0$ are the sequences for which $d \in T(M, x)$ by definition.

Exercise 3 (Linearizing cone)**.** Verify Remark 3.11 in the lecture notes, that is: The linearizing cone for the NLP

$$\min f(x) \quad \text{s.t.} \quad g(x) \leq 0, \quad h(x) = 0,$$

with $f\colon \mathbb{R}^n \to \mathbb{R}$, $g\colon \mathbb{R}^n \to \mathbb{R}^m$ and $h\colon \mathbb{R}^n \to \mathbb{R}^p$, is given by

$$T_\ell(G, K, x) = \Big\{d \in \mathbb{R}^n\colon \nabla h(x)^T d = 0,\ \nabla g_i(x)^T d \leq 0 \text{ for } i \in \mathcal{A}(x)\Big\}$$

for a feasible point x.

Solution. As in the lecture notes, we identify $X = \mathbb{R}^n$, $Z = \mathbb{R}^m \times \mathbb{R}^p = \mathbb{R}^{m+p}$, $G(x) = \binom{g(x)}{h(x)}$ and $K = (-\infty, 0]^m \times \{0\}^p$. The linearizing cone as in the lecture notes is given by

$$T_\ell(G, K, x) = \Big\{d \in \mathbb{R}^n\colon G'(x)d \in \overline{\operatorname{cone}(K, G(x))}\Big\},$$

for which we rewrite

$$d \in T_\ell(G,K,x) \quad \iff \quad \begin{pmatrix} \nabla g(x)^T d \\ \nabla h(x)^T d \end{pmatrix} \in \overline{\left\{ \lambda \begin{pmatrix} y - g(x) \\ 0 \end{pmatrix} : y \in (-\infty, 0]^m,\, \lambda > 0 \right\}}.$$

Set

$$\mathcal{C} := \left\{ d \in \mathbb{R}^n \colon \nabla h(x)^T d = 0,\, \nabla g_i(x)^T d \le 0 \text{ for } i \in \mathcal{A}(x) \right\}.$$

One immediately observes that a direction $d \in \mathbb{R}^n$ needs to satisfy $\nabla h(x)^T d = 0$ to be in either cone $T_\ell(G,K,x)$ or $\mathcal{C}$. In this sense, we concentrate on the inequality constraints in the following.

1. Let first $d \in T_\ell(G,K,x)$, that is, there is a sequence $\lambda_k(y^k - g(x)) \subseteq \operatorname{cone}(K, G(x))$ such that $\lambda_k(y^k - g(x)) \to G'(x)d$ where $\lambda_k > 0$ and $y^k \le 0$ for every $k \in \mathbb{N}$. This means that for every $i = 1, \ldots, m$ we have

$$\nabla g_i(x)^T d = \lim_{k\to\infty} \lambda_k \big(y_i^k - g_i(x)\big) \quad \text{with } \lambda > 0 \text{ and } y_i^k \le 0.$$

   For $i \in \mathcal{A}(x)$, this implies that $\nabla g_i(x)^T d = \lim_{k\to\infty} \lambda_k y_i^k \le 0$, and thus $d \in \mathcal{C}$ as required.

2. Now assume that $d \in \mathcal{C}$, so $\nabla g_i(x)^T d \le 0$ for $i \in \mathcal{A}(x)$. For $i \in \mathcal{A}(x)$, we can choose the nonpositive $y_i = g_i(x) + \frac{1}{\lambda}\nabla g_i(x)^T d$ and any number $\lambda > 0$ to obtain a representation

$$\nabla g_i(x)^T d = \lambda\big(y_i - g_i(x)\big) \quad \text{with } \lambda > 0 \text{ and } y_i \le 0.$$

   If $i \notin \mathcal{A}(x)$, then we have $g_i(x) < 0$. An analogous ansatz as in the previous case (rearrange the previous equality) yields

$$g_i(x) + \frac{1}{\lambda}\nabla g_i(x)^T d = y_i \overset{!}{\le} 0$$

   for some number $\lambda > 0$. This requires

$$\lambda \ge \frac{-\nabla g_i(x)^T d}{g_i(x)} \quad \text{for all } i \notin \mathcal{A}(x),$$

   which we achieve by setting

$$\lambda := \max\left(1, \left\{ \frac{-\nabla g_i(x)^T d}{g_i(x)} : i \notin \mathcal{A}(x) \right\}\right).$$

   With this choice of λ and $y_i := g_i(x) + \frac{1}{\lambda}\nabla g_i(x)^T d$, we have

$$\nabla g(x)^T d \in \left\{ \lambda\big(y - g(x)\big) \colon y \in (-\infty, 0]^m,\, \lambda > 0 \right\},$$

   hence $d \in T_\ell(G,K,x)$, again without the closure.

Exercise 4 (Optimal control problem in reduced form)**.** Consider the basic optimal control problem

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

with Y, U and Z Banach spaces and $J\colon Y\times U\to\mathbb{R}$ and $E\colon Y\times U\to Z$ (continuously) F-differentiable. Assume that for every $u\in U$ there exists a unique $y=y(u)\in Y$ such that $E(y(u),u)=0$. Then we can reduce the above optimal control problem to the unrestricted optimization problem

$$\min_{u\in U} j(u) := J(y(u),u). \tag{ROCP}$$

In this exercise, we investigate the *control-to-state* operator $U\ni u\mapsto y(u)\in Y$ more in depth. We additionally assume that $E_y'(y(u),u)=(\partial_y E)(y(u),u)\in\mathcal{L}(Y;Z)$ is continuously invertible for every $u\in U$, so the inverse operator exists and is also linear and continuous.

(a) Use the *implicit function theorem* to show that y is (continuously) F-differentiable and determine an explicit formula for $y'(u)$ from $E(y(u),u)=0$ for all $u\in U$.

(b) Give an expression for the *sensitivity* $j'(u)h$ in direction $h\in U$ using the derived formula for $y'(u)$.

(c) Show that we can represent the total derivative $j'(u)$ by

$$j'(u)=y'(u)^*J_y'\big(y(u),u\big)+J_u'\big(y(u),u\big)=E_u'\big(y(u),u\big)^*p+J_u'\big(y(u),u\big),$$

where $p=p(u)\in Z^*$ is the *adjoint state* satisfying the *adjoint equation*

$$E_y'\big(y(u),u\big)^*p=-J_y'\big(y(u),u\big).$$

(d) Let lastly Y, U and Z be finite-dimensional. We imagine this to originate from a discretization of the infinite-dimensional problem, so the underlying space dimensions n_Y, n_u and n_Z may be very high and taking inverse matrices is not an option. Compare the effort needed to compute the total derivative $j'(u)$ (or $\nabla j(u)$ for that matter) using the sensitivity approach and the adjoint approach, respectively.

Solution.

(a) The implicit function theorem says, in the terminology of this exercise: If E is continuously F-differentiable in a point (y,u) and the partial derivative $E_y'(y,u)\in\mathcal{L}(Y;Z)$ is continuously invertible, then there exist neighborhoods $\mathcal{U}_y\subset Y$ of y and $\mathcal{U}_u\subset U$ of u together with an implicit function

$$\varphi\colon\mathcal{U}_y\to\mathcal{U}_u \quad \text{such that} \quad E\big(\varphi(u),u\big)=E(y,u) \quad \text{for all } u\in\mathcal{U}_u.$$

This implicit function φ is continuously F-differentiable.

If indeed (y,u) is of the form $(y(u),u)$, then we have $E\big(y(u),u\big) = 0$ and φ coincides exactly with the control-to-state operator $u \mapsto y(u)$ on $\mathcal{U}_u$. In this case, the implicit function theorem implies that the control-to-state operator is continuously F-differentiable. Since we have assumed the assumptions for the implicit function theorem to hold for every $u \in U$ and every pair $(y(u),u)$, we obtain global continuous F-differentiablity of the control-to-state operator.

Since $E(y(u),u) = 0$ for all $u \in U$, we know that its derivative in any direction will be zero, hence

$$E'_y\big(y(u),u\big)y'(u)h + E'_u\big(y(u),u\big)h = 0 \quad \text{for all } h \in U$$

and

$$y'(u) = -E'_y\big(y(u),u\big)^{-1}E'_u\big(y(u),u\big). \tag{1}$$

(b) From the chain rule and the foregoing expression for $y'(u)$ in 1, we obtain

$$\begin{aligned} j'(u)h &= J'_y\big(y(u),u\big)y'(u)h + J'_u\big(y(u),u\big)h \\ &= -J'_y\big(y(u),u\big)E'_y\big(y(u),u\big)^{-1}E'_u\big(y(u),u\big)h + J'_u\big(y(u),u\big)h \end{aligned}$$

for every direction $h \in U$.

(c) Since $J'_y\big(y(u),u)\big) \in Y^*$ and $y'(u) \in \mathcal{L}(U;Y)$, we can rewrite

$$J'_y\big(y(u),u\big)y'(u)h = \big\langle J'_y\big(y(u),u\big), y'(u)h\big\rangle_{Y^*,Y} = \big\langle y'(u)^*J'_y\big(y(u),u\big), h\big\rangle_{U^*,U},$$

where $y'(u)^* \in \mathcal{L}(Y^*;U^*)$ is the adjoint operator to $y'(u)$ given by

$$y'(u)^* = -E'_u\big(y(u),u\big)^*E'_y\big(y(u),u\big)^{-*}.$$

Setting

$$p := -E'_y\big(y(u),u\big)^{-*}J'_y\big(y(u),u\big) \qquad \Longleftrightarrow \qquad E'_y\big(y(u),u\big)^*p = -J'_y\big(y(u),u\big),$$

we obtain

$$J'_y\big(y(u),u\big)y'(u)h = \big\langle y'(u)^*J'_y\big(y(u),u\big), h\big\rangle_{U^*,U} = \big\langle E'_u\big(y(u),u\big)^*p, h\big\rangle_{U^*,U}.$$

Re-inserting into the formula for $j'(u)h$ from (b), this yields

$$j'(u)h = \big\langle E'_u\big(y(u),u\big)^*p + J'_u\big(y(u),u\big), h\big\rangle_{U^*,U} \quad \text{for all } h \in U$$

and thus the searched-for formula.

(d) Given $h \in U$, calculating $j'(u)h$ via the sensitivity formula as in (b) requires to solve the linear system of equations of size $n_Z \times n_Y$ (recall that taking inverse matrices is forbidden)

$$E'_y\big(y(u), u\big)v(u; h) = -E'_u\big(y(u), u\big)h$$

and then set

$$j'(u)h = J'_y\big(y(u), u\big)v(u; h) + J'_u\big(y(u), u\big)h.$$

This means that in order to fully determine the total derivative $j'(u)$, or the gradient $\nabla j(u)$, we need to solve n_U linear system of equations of size $n_Z \times n_Y$ – once for *each* basis vector e_i of U to rebuild $j'(u)$ from $v(u, e_i)$.

For the adjoint approach on the other hand, we only need to solve the linear system of equations of size $n_Y \times n_Z$

$$E'_y\big(y(u), u\big)^* p = -J'_y\big(y(u), u\big)$$

once to then calculate $j'(u) = E'_u\big(y(u), u\big)^* p + J'_u\big(y(u), u\big)$. This is an immense practical advantage of the adjoint method.