Example of a cost functional:

$$J(y,u) = |y - y_d|^2 + \lambda |u|^2 \quad , \quad |\cdot|: \text{ Euclidean norm}, \; y_d \in \mathbb{R}^n \text{ given}$$

Assuming that the matrix A is invertible, we can solve for y in the state equation, obtaining

$$y = A^{-1} B u ,$$

and for any $u \in \mathbb{R}^m$ there is a unique $y \in \mathbb{R}^n$.
Hence, we can consider $y = y(u)$ for a given control u.

Introduce the <u>solution matrix</u> $S := A^{-1}B$, $S: \mathbb{R}^m \to \mathbb{R}^n$, and we can eliminate y from J to obtain the <u>reduced cost functional</u>

$$f(u) := J(Su, u) = J(y, u).$$

E.g. using the cost functional example above, we get

$$f(u) = |Su - y_d|^2 + \lambda |u|^2$$

Problem (*) has become a nonlinear optimisation problem,

$$\min f(u), \quad u \in U_{ad}.$$

In this <u>reduced problem</u> only u appears as an unknown.

<u>Existence of optimal controls</u>

<u>Definition</u>: A vector $\bar{u} \in U_{ad}$ is called an <u>optimal control</u> for (*) if $f(\bar{u}) \le f(u)$ for all $u \in U_{ad}$; then $\bar{y} := S\bar{u}$ is called <u>optimal state</u> associated with $\bar{u}$.

<u>Note</u>: Optimal (or locally optimal) quantities will be indicated by overlining, e.g. $\bar{u}$ or $\bar{y}$.

<u>Theorem 1.1</u>: Suppose J is continuous on $\mathbb{R}^n \times U_{ad}$ and U_{ad} is non-empty, bounded and closed.
If A is invertible, then (*) has at least one solution.

<u>Proof</u>: By assumption, U_{ad} is non-empty, bounded and closed, hence compact; and f is continuous on U_{ad}.
By the Weierstrass ~~ex~~ extreme value theorem, f attains its minimum in U_{ad}. Hence, some $\bar{u} \in U_{ad}$ exists such that $f(\bar{u}) = \min_{u \in U_{ad}} f(u)$. $\square$

<u>Note</u>: In the case of optimal control for PDE, the equivalent statement will be harder to prove, since bounded and closed sets need not be compact in (infinite-dimensional) function spaces.

First-order necessary optimality conditions

<u>Aim</u>: Find conditions to characterise $\bar{u}$ and $\bar{y}$, to enable us to determine $\bar{u}, \bar{y}$.

<u>Notation</u>: For $f: \mathbb{R}^m \to \mathbb{R}$ let $\quad D_i = \dfrac{\partial}{\partial x_i}, \quad D_x = \dfrac{\partial}{\partial x}$,

$$f'(x) = (D_1 f(x), \dots, D_m f(x)), \quad \text{(derivative)} \Leftrightarrow \text{row vector}$$

$$\nabla f(u) = f'(u)^T, \quad \text{(gradient)} \Leftrightarrow \text{column vector}$$

$$(u, v)_{\mathbb{R}^m} = u \cdot v = \sum_{i=1}^{m} u_i v_i \quad \text{(scalar product)}$$

For $h \in \mathbb{R}^m$: $\quad f'(u)h = (\nabla f(u), h)_{\mathbb{R}^m} = \nabla f(u) \cdot h$

Theorem 1.2: Let $\bar{u}$ be optimal control for $(*)$ and f differentiable in $\bar{u}$. If U_{ad} is convex, then $\bar{u}$ satisfies the variational inequality

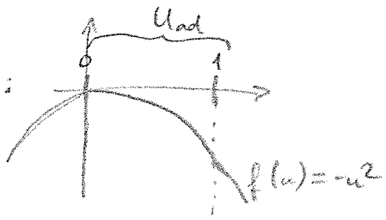
$$f'(\bar{u})(\tilde{u}-\bar{u}) \geq 0 \quad \text{f.a. } \tilde{u} \in U_{ad}$$

Proof: Let $\tilde{u} \in U_{ad}$ and define $u(t) = t\tilde{u} + (1-t)\bar{u} = \bar{u} + t(\tilde{u}-\bar{u})$.
Since U_{ad} is convex, the whole line $u(t)$ is in U_{ad}.
Optimality of $\bar{u}$ implies $f(u(t)) \geq f(\bar{u})$, hence

$$0 \leq \frac{1}{t}\left(f(u(t)) - f(\bar{u})\right) = \frac{1}{t}\left(f(\bar{u} + t(\tilde{u}-\bar{u})) - f(\bar{u})\right)$$

Since f is differentiable in $\bar{u}$, the limit $t \to 0$ implies the claim. $\square$

Note: In general, we cannot expect $f'(\bar{u}) = 0$:



"At a minimum $\bar{u}$, f cannot decrease in any direction in U_{ad}."

Interpretation of this necessary condition for J? (Recall: $f(u) = J(Su, u)$ and $S = A^{-1}B$)
By the chain rule we obtain

$$\begin{aligned} f'(\bar{u})h &= D_y J(S\bar{u}, \bar{u})Sh + D_u J(S\bar{u}, \bar{u})h \\ &= \left(\nabla_y J(\bar{y}, \bar{u}), A^{-1}Bh\right)_{\mathbb{R}^m} + \left(\nabla_u J(\bar{y}, \bar{u}), h\right)_{\mathbb{R}^m} \\ &= \left(B^T (A^{-1})^T \nabla_y J(\bar{y}, \bar{u}) + \nabla_u J(\bar{y}, \bar{u}),\ h\right)_{\mathbb{R}^m}. \end{aligned}$$

Hence, the variational inequality takes the form

$$\left(B^T (A^{-1})^T \nabla_y J(\bar{y}, \bar{u}) + \nabla_u J(\bar{y}, \bar{u}),\ \tilde{u} - \bar{u}\right)_{\mathbb{R}^m} \geq 0 \quad \text{for all } \tilde{u} \in U_{ad}.$$

It can be considerably simplified by introducing the so-called adjoint state.