Operators

THE AUSTRALIAN NATIONAL UNIVERSITY

Linear Operator

Generalization of matrix. We map from one Banach space into another. Norm and eigenvalues/eigenvectors are defined as with matrices. We use $A : F \to G$.

A Matrix — Operator Dictionary

Transposed Matrix — Adjoint Operator

$$\langle f, Ag \rangle = \langle A^* f, g \rangle \text{ for all } f \in F, g \in G$$

Symmetric Matrix — Self Adjoint Operator

$$\langle f, Ag \rangle = \langle f, A^* g \rangle \text{ for all } f \in F, g \in G$$

Orthogonal Matrix — Isometry

$$\langle f, g \rangle = \langle Af, Ag \rangle$$

Note that the requirement $\|f\| = \|Af\|$ is sufficient (polarization trick).

Linear Operators: Examples

THE AUSTRALIAN NATIONAL UNIVERSITY

Input Transformation

Consider class of functions $f : \mathbb{R}^n \to \mathbb{R}$. A linear operator on such functions could be $A : f(\cdot) \to f(a(\cdot))$, i.e. the argument of f is transformed (e.g. we transform the images before feeding them into a classifier).

Differentiation

The differential operator $D : f \to \frac{d}{dx} f$ is linear.

Scaling

Transform $f \to \alpha f$.

Fourier Transformation

We map f into its Fourier transform. This leads to

$$f \to \tilde{f}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(x) e^{-i\omega x} dx$$

This map is an isometry, since

$$\|f\|^2 = \int_{\mathbb{R}} |f(x)|^2 dx = \int_{\mathbb{R}} |\tilde{f}(\omega)|^2 d\omega = \|\tilde{f}\|^2$$

Linear Forms

THE AUSTRALIAN NATIONAL UNIVERSITY

Linear Form

A specific kind of linear operator, namely one that maps into $\mathbb{R}$.

Dual Space

The space of all linear forms on a Banach space $\mathcal{X}$ is called the dual space and often denoted by $\mathcal{X}^*$.

We can write the linear forms $l(\mathbf{x})$ by $\langle \mathbf{x}^*, \mathbf{x} \rangle$ where $\mathbf{x}^* \in \mathcal{X}^*$ is chosen to correspond to $l(\mathbf{x})$.

Rank

All non-null linear forms have rank 1.

Example: Linear Functions on $\mathbb{R}^m$

$l(\mathbf{x}) = \sum_{i=1}^{m} x_i$. This corresponds to $\mathbf{x}^* = (1, \ldots, 1)$.

Useful Properties of Linear Form

THE AUSTRALIAN NATIONAL UNIVERSITY

Norm

We can induce a dual norm on $\mathcal{X}^*$ by

$$\mathbf{x}^* := \max_{\mathbf{x} \in \mathcal{X}} \frac{|\langle \mathbf{x}^*, \mathbf{x} \rangle|}{\|\mathbf{x}\|}$$

For ℓ_p spaces the dual space is an ℓ_q space with $\frac{1}{p} + \frac{1}{q} = 1$. We show this by Hölder's inequality

$$|\langle \mathbf{x}^*, \mathbf{x} \rangle| = \left| \sum_i x_i^* x_i \right| \leq \|\mathbf{x}^*\|_p \|\mathbf{x}\|_q$$

Since the inequality is tight for all $\frac{1}{p} + \frac{1}{q} = 1$ we are done. For ℓ_1 this is ℓ_∞.

Note that Hilbert spaces are dual to themselves.

Subspaces

We can define a linear form on a subspace spanned by $\mathbf{x}_i$ by requiring

$$\langle \mathbf{w}, \mathbf{x} \rangle \text{ with } \mathbf{w} = \sum_{i=1}^{m} \alpha_i \mathbf{x}_i$$

Derivatives in $\mathbb{R}^n$

THE AUSTRALIAN NATIONAL UNIVERSITY

Differentiability

A function $f : \mathcal{X} \to \mathbb{R}$ is **differentiable** if there exists a linear map $f'(\mathbf{x}) \in \mathcal{X}^*$ such that the following limit exists

$$\lim_{\mathbf{y} \to 0} \frac{f(\mathbf{x}+\mathbf{y}) - f(\mathbf{x}) - \langle f'(\mathbf{x}), \mathbf{y} \rangle}{\|\mathbf{y}\|} = 0.$$

The functions in this lecture are usually differentiable ...

Partial Derivative

For a function $f : \mathbb{R}^m \to \mathbb{R}$ compute the partial derivative by treating $f(\mathbf{x})$ as a function of each x_i separately while keeping the rest fixed.

Gradient

This is the vector of all partial derivatives. See also the linear form above.

Criterion for Differentiability

- Existence of all partial derivatives in $f(\mathbf{x})$

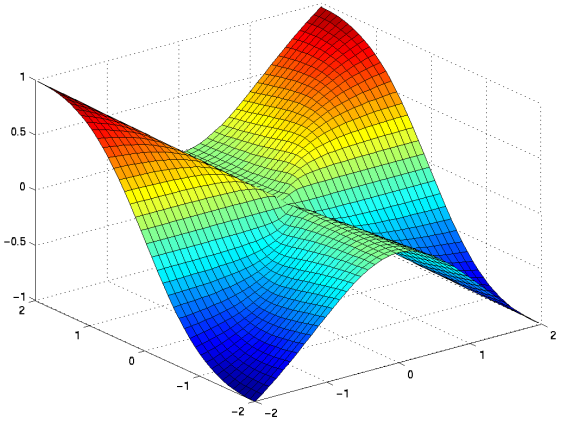
- Continuity of the partial derivatives in $f(\mathbf{x})$.

When things go wrong

THE AUSTRALIAN NATIONAL UNIVERSITY

$$f(\mathbf{x}) = \frac{x_1^2 x_2}{x_1^2 + x_2^2}$$

All partial derivatives exist in $(0,0)$ but $f(\mathbf{x})$ is not differentiable in $(0,0)$.



Taylor Expansion

THE AUSTRALIAN NATIONAL UNIVERSITY

Basic Idea

Extend the approximation by linear functionals as in the definition of the derivative.

Taylor Expansion in $\mathbb{R}$

For an $n+1$-times differentiable function we have

$$f(x+\varepsilon) = f(x) + \sum_{i=1}^{n} \frac{1}{i!} \varepsilon^i f^{(n)}(x) + o(\varepsilon^{n+1})$$

Taylor Expansion in $\mathbb{R}^n$

Usually we only need first and second order expansions. The second order information of f at $\mathbf{x}$ is called the **Hessian**. We have

$$f(\mathbf{x}+\boldsymbol{\epsilon}) = f(\mathbf{x}) + \boldsymbol{\epsilon}^\top f'(\mathbf{x}) + \frac{1}{2} \boldsymbol{\epsilon}^\top f''(\mathbf{x}) \boldsymbol{\epsilon} + o(\|\boldsymbol{\epsilon}\|^3)$$

Admissibility of Expansion

Expansion is OK if the $n+1$st derivative exists. Then the error term is bounded by the size of the latter and $\|\boldsymbol{\epsilon}\|^{n+1}$.

Newton Method

THE AUSTRALIAN NATIONAL UNIVERSITY

Basic Idea

We want to minimize $f(\mathbf{x})$. Use quadratic approximation and solve at each step for the minimum of the latter explicitly. We get $f'(\mathbf{x}) + f''(\mathbf{x})\boldsymbol{\epsilon} = 0$ which yields the following algorithm:

Require: x_0, Precision ϵ

Set $x = x_0$

repeat

$x = x - \frac{f'(x)}{f''(x)}$

until $|f'(x)| \leq \epsilon$

Output: x

Convergence of Newton Method Let $g : \mathbb{R} \to \mathbb{R}$ be a twice continuously differentiable function and denote by $x^* \in \mathbb{R}$ a point with $f''(x^*) \neq 0$ and $f'(x^*) = 0$. Then, provided x_0 is sufficiently close to x^*, the sequence generated by the Newton method will converge to x^* at least quadratically.

Derivatives in Banach Spaces

THE AUSTRALIAN NATIONAL UNIVERSITY

Basic Idea

Extend the definitions from derivatives in $\mathbb{R}^n$. Linear approximation

$$\lim_{\mathbf{y} \to 0} \frac{f(\mathbf{x}+\mathbf{y}) - f(\mathbf{x}) - \langle f'(\mathbf{x}), \mathbf{y} \rangle}{\|\mathbf{y}\|}$$

is useful if we can compute it. Otherwise, also the Gateaux derivative is convenient. We define $f'(\mathbf{x})$ by

$$\langle f'(\mathbf{x}), \mathbf{y} \rangle = \frac{d}{d\alpha} f(\mathbf{x} + \alpha\mathbf{y}) \text{ for } \alpha \to 0$$

Example (in a Hilbert Space)

For $f(\mathbf{x}) = \|\mathbf{x}\|^4$ we have

$$f(\mathbf{x} + \alpha\mathbf{y}) = \left(\|\mathbf{x}\|^2 + 2\alpha\langle \mathbf{x}, \mathbf{y} \rangle + \alpha^2\|\mathbf{y}\|\right)^2$$

Since $\frac{d}{d\alpha} f(\mathbf{x} + \alpha\mathbf{y}) = 4\|\mathbf{x}\|^2\langle \mathbf{x}, \mathbf{y} \rangle$ for $\alpha \to 0$ we have $f'(\mathbf{x}) = 4\|\mathbf{x}\|^2\mathbf{x}$.

Convexity

THE AUSTRALIAN NATIONAL UNIVERSITY

Convex Set

A set X is called convex if for any $\mathbf{x}, \mathbf{x}' \in X$ and any $\lambda \in [0, 1]$ we have

$$\lambda\mathbf{x} + (1 - \lambda)\mathbf{x}' \in X.$$

Convex Function

A function f defined on a set X (note that X need not be convex itself) is called convex if for any $\mathbf{x}, \mathbf{x}' \in X$ and any $\lambda \in [0, 1]$ such that $\lambda\mathbf{x} + (1 - \lambda)\mathbf{x}' \in X$ we have

$$f(\lambda\mathbf{x} + (1 - \lambda)\mathbf{x}') \leq \lambda f(\mathbf{x}) + (1 - \lambda) f(\mathbf{x}').$$

A function f is called *strictly* convex if the inequality is strict for $\lambda \in (0, 1)$.

Convex Sets as Below Sets

THE AUSTRALIAN NATIONAL UNIVERSITY

Lemma

Denote by $f : \mathcal{X} \to \mathbb{R}$ a convex function on $\mathcal{X}$. Then the set

$$X := \{\mathbf{x} | \mathbf{x} \in \mathcal{X} \text{ and } f(\mathbf{x}) \leq c\} \text{ for some } c \in \mathbb{R}$$

is convex.

Proof

We have to show that for all $\mathbf{x}, \mathbf{x}' \in \mathcal{X}$ we have $\lambda\mathbf{x} + (1 - \lambda)\mathbf{x}' \in \mathcal{X}$. For any $\mathbf{x}, \mathbf{x}' \in \mathcal{X}$ we have $f(\mathbf{x}), f(\mathbf{x}') \leq c$. Moreover, since f is convex, we also have

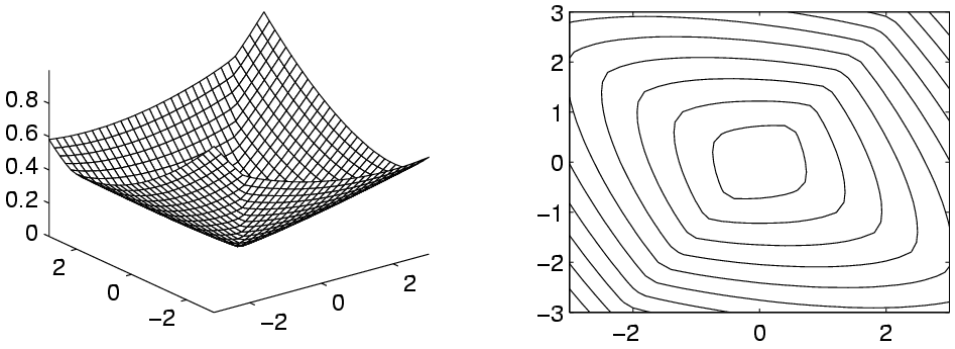
$$f(\lambda\mathbf{x} + (1 - \lambda)\mathbf{x}') \leq \lambda f(\mathbf{x}) + (1 - \lambda) f(\mathbf{x}') \leq c \text{ for all } \lambda \in [0, 1].$$

Hence, for all $\lambda \in [0, 1]$ we have $(\lambda\mathbf{x} + (1 - \lambda)\mathbf{x}') \in X$ which proves the claim.

Example

THE AUSTRALIAN NATIONAL UNIVERSITY

- Level sets of the $\ell_{1.5}$ norm.



Uniqueness of Minimum

THE AUSTRALIAN NATIONAL UNIVERSITY

Theorem

If the convex function $f : \mathcal{X} \to \mathbb{R}$ has a minimum on a convex set $X \subset \mathcal{X}$, then its arguments $\mathbf{x} \in \mathcal{X}$ for which the minimum value is attained, form a convex set. Moreover, if f is strictly convex, then this set will contain only one element.

Proof

Denote by c the minimum of f on X. Then clearly the set $X_m := \{\mathbf{x} | \mathbf{x} \in \mathcal{X} \text{ and } f(\mathbf{x}) \leq c\}$ is convex. Moreover $X_m \cap X$ is also convex and $f(\mathbf{x}) = c$ for all $\mathbf{x} \in X_m \cap X$ (otherwise c would not be the minimum).

If f is strictly convex, for any $\mathbf{x}, \mathbf{x}' \in X$, and in particular for any $\mathbf{x}, \mathbf{x}' \in X \cap X_m$ we have (for $\mathbf{x} \neq \mathbf{x}'$ and all $\lambda \in (0, 1)$

$$f(\lambda \mathbf{x} + (1-\lambda)\mathbf{x}') < \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{x}') = \lambda c + (1-\lambda)c = c.$$

This contradicts the assumption that $X_m \cap X$ contains more then one element.

Constrained Convex Minimization

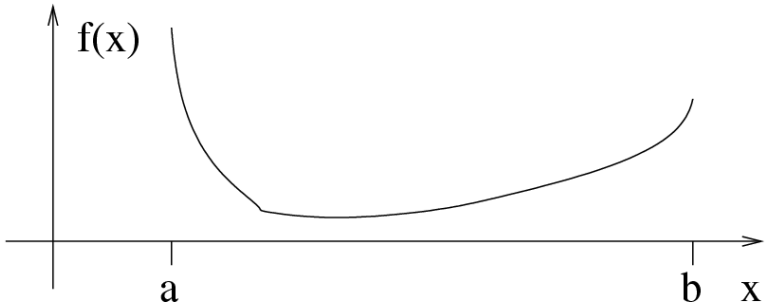
THE AUSTRALIAN NATIONAL UNIVERSITY

Corollary

Denote by $f, c_1, \ldots, c_n$ convex functions on $\mathcal{X}$. Then the problem

$$\begin{aligned} &\text{minimize} && f(x) \\ &\text{subject to} && c_i(x) \leq 0 \text{ for all } i \in [n] \end{aligned}$$

has as solution a convex set if a solution exists and this solution is unique if f is strictly convex.



Interval Cutting Method

THE AUSTRALIAN NATIONAL UNIVERSITY

Basic Idea

If f is a convex function, the slope can only increase. So, all we have to do is look at where the slope changes. This algorithms is the same as the one for finding the root of a monotonic function.

Require: a, b, Precision ϵ
  Set $A = a, B = b$
  repeat
    if $f'\left(\frac{A+B}{2}\right) > 0$ **then**
      $B = \frac{A+B}{2}$
    else
      $A = \frac{A+B}{2}$
    end if
  until $(B - A)\min(|f'(A)|, |f'(B)|) \leq \epsilon$
Output: $x = \frac{A+B}{2}$

Constrained Optimization

THE AUSTRALIAN NATIONAL UNIVERSITY

Optimization Problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & c_i(x) \leq 0 \text{ for all } i \in [n]. \end{array}$$

Here $c_i(x)$ are **convex** functions.

Lagrange Function

The basic idea is to convert the constrained optimization problem into the problem of finding the **saddlepoint** of the Lagrange function

$$L(x, \alpha) := f(x) + \sum_{i=1}^{n} \alpha_i c_i(x) \text{ where } \alpha_i \geq 0.$$

Theorem

Minimize $L(x, \alpha)$ with respect to x while maximizing $L(x, \alpha)$ with respect to α. For an optimal solution we have

$$L(\bar{x}, \alpha) \leq L(\bar{x}, \bar{\alpha}) \leq L(x, \bar{\alpha})$$

Saddlepoint Property of Lagrange Function

THE AUSTRALIAN NATIONAL UNIVERSITY

Proof of $L(\bar{x}, \alpha) \leq L(\bar{x}, \bar{\alpha}) \leq L(x, \bar{\alpha})$

- From the first inequality it follows that $\sum_{i=1}^{n} (\alpha_i - \bar{\alpha}_i) c_i(\bar{x}) \leq 0$.
- Since $\alpha_i \geq 0$ was arbitrary, we can see (by setting all but one of the terms α_i to $\bar{\alpha}_i$ and the remaining one to $\alpha_i = \bar{\alpha}_i + 1$) that $c_i(x) \leq 0$ for all $i \in [n]$. This shows that $\bar{x}$ satisfies the constraints, i.e. it is feasible.
- By setting one of the α_i to 0 we see that $\bar{\alpha}_i c_i(\bar{x}) \geq 0$. The only way to satisfy this is by requiring

  $$\bar{\alpha}_i c_i(\bar{x}) = 0 \text{ for all } i \in [n]. \text{ This yields } L(\bar{x}, \bar{\alpha})$$

  This is often also referred to as the Karush-Kuhn-Tucker (KKT) condition.
- Combining the latter and $c_i(\bar{x}) \leq 0$ with the second inequality in the optimality condition yields $f(\bar{x}) \leq f(x) + \sum_{i=1}^{n} \alpha_i c_i(x) \leq f(x)$ for all feasible x. This proves that $\bar{x}$ is optimal.

Differentiable Functions

THE AUSTRALIAN NATIONAL UNIVERSITY

Lagrange Function

$$L(x, \alpha) = f(x) + \sum_{i=1}^{n} \alpha_i c_i(x)$$

From the saddle point condition and the fact that $c_i(x)\alpha_i = 0$ for the optimal solution we obtain the following conditions.

Optimality Conditions

$$\begin{aligned} \partial_x L(\bar{x}, \bar{\alpha}) &= \partial_x f(\bar{x}) + \sum_{i=1}^{n} \bar{\alpha}_i \partial_x c_i(\bar{x}) = 0 \text{ (Minimum in } \bar{x} \text{ for } L(x, \bar{\alpha})) \\ \partial_{\alpha_i} L(\bar{x}, \bar{\alpha}) &= c_i(\bar{x}) \leq 0 \text{ (Minimum in } \bar{\alpha} \text{ for } L(\bar{x}, \alpha)) \\ \sum_{i=1}^{n} \bar{\alpha}_i c_i(\bar{x}) &= 0 \text{ (Vanishing KKT-Gap)} \end{aligned}$$

The Primal-Dual HOWTO

THE AUSTRALIAN NATIONAL UNIVERSITY

Primal Objective

Constrained optimization problem with $f(x)$, subject to constraints $c_i(x) \leq 0$.

Dual Objective

From the saddlepoint conditions of the Lagrange functions we can eliminate x as $x(\alpha)$ and write $L(x(\alpha), \alpha)$. This is the **dual objective function**.

We get dual constraints from the optimality conditions on x via $\partial_x L(x(\alpha), \alpha) = 0$.

Trick: Variables and Constraints

Free Variable $\Longrightarrow$ Equality Constraint

Equality Constraint $\Longrightarrow$ Free Variable

Inequality Constraint $\Longrightarrow$ Inequality Constraint

Linear Programs

THE AUSTRALIAN NATIONAL UNIVERSITY

Primal Objective

$$\begin{aligned} \text{minimize} \quad & c^\top x \\ \text{subject to} \quad & Ax + b \leq 0 \end{aligned}$$

Lagrange Function

$$L(x, \alpha) = c^\top x + \alpha^\top (Ax + b) \text{ where } \alpha_i \geq 0$$

Kuhn-Tucker Conditions

$$\alpha_i (Ax + b)_i = 0$$

Dual Objective (also Wolfe's Dual)

Saddlepoint condition in the primal variables x yields

$$\partial_x L(x, \alpha) \; = \; c + A^\top \alpha = 0$$

and therefore

$$\begin{aligned} \text{maximize} \quad & L(x(\alpha), \alpha) = (c + A^\top \alpha)^\top x + \alpha^\top b = \alpha^\top b \\ \text{subject to} \quad & c + A^\top \alpha = 0 \text{ and } \alpha \geq 0 \end{aligned}$$

Quadratic Programs

THE AUSTRALIAN NATIONAL UNIVERSITY

Primal Objective (K is positive definite)

$$\begin{aligned} \text{minimize} \quad & \tfrac{1}{2} x^\top K x + c^\top x \\ \text{subject to} \quad & Ax + b \leq 0 \end{aligned}$$

Lagrange Function

$$L(x, \alpha) = \frac{1}{2} x^\top K x + c^\top x + \alpha^\top (Ax + b)$$

Kuhn-Tucker Conditions

$$\alpha_i (Ax + b)_i = 0$$

Dual Objective (also Wolfe's Dual)

Saddlepoint condition in the primal variables x yields

$$\partial_x L(x, \alpha) \; = \; Kx + A^\top \alpha + c = 0$$

and therefore

$$\begin{aligned} \text{maximize} \quad & L(x(\alpha), \alpha) = -\tfrac{1}{2} \alpha^\top A^\top K^{-1} A \alpha + \left[b - c^\top K^{-1} A^\top \right] \alpha - \tfrac{1}{2} c^\top K^{-1} c \\ \text{subject to} \quad & \alpha \geq 0 \end{aligned}$$