Optimization Methods.
Seminar 5. Introduction to Matrix calculus.

Alexandr Katrutsa

Moscow Institute of Physics and Technology
Department of Control and Applied Mathematics

October 3, 2016

Reminder

- Conjugate sets
- Properties of conjugate sets
- Farkas' lemma

Basic definitions

More details see here. Let $f : D \to E$ be some function, then its derivative $\frac{\partial f}{\partial x} \in G$:

D	E	G	Name
$\mathbb{R}$	$\mathbb{R}$	$\mathbb{R}$	Derivative, $f'(x)$
$\mathbb{R}^n$	$\mathbb{R}$	$\mathbb{R}^n$	Gradient, $\frac{\partial f}{\partial x_i}$
$\mathbb{R}^n$	$\mathbb{R}^m$	$\mathbb{R}^{n \times m}$	Jacobian, $\frac{\partial f_i}{\partial x_j}$
$\mathbb{R}^{m \times n}$	$\mathbb{R}$	$\mathbb{R}^{m \times n}$	$\frac{\partial f}{\partial x_{ij}}$

Also square matrix $n \times n$ of the second derivatives $\mathbf{H} = [h_{ij}]$ in case of $f : \mathbb{R}^n \to \mathbb{R}$ is called hessian and its elements equal $h_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}$.

The main technique

Matrix representation of function: $f = \mathbf{x}^\top \mathbf{A}\mathbf{x}$

$\downarrow$

Scalar representation of function: $f = \sum\limits_{ij} a_{ij} x_i x_j$

$\xrightarrow{\nabla f_i = \frac{\partial f}{\partial x_i}}$

Scalar representation of gradient: $\nabla f_i = \sum\limits_{j} a_{ij} x_j$

$\uparrow$

Matrix representation of gradient: $\nabla f = \mathbf{A}\mathbf{x}$

Examples

1. Linear function: $f(\mathbf{x}) = \mathbf{c}^\intercal \mathbf{x}$
2. Quadratic form: $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\intercal \mathbf{A}\mathbf{x} + \mathbf{b}^\intercal \mathbf{x}$
3. ℓ_2 norm of difference squared: $f(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2$
4. Determinant: $f(\mathbf{X}) = \det \mathbf{X}$
5. Trace: $f(\mathbf{X}) = \mathrm{Tr}(\mathbf{A}\mathbf{X}\mathbf{B})$
6. $f(\mathbf{x}) = (\mathbf{x} - \mathbf{A}\mathbf{s})^\intercal \mathbf{W}(\mathbf{x} - \mathbf{A}\mathbf{s})$
7. $f(\mathbf{A}) = (\mathbf{x} - \mathbf{A}\mathbf{s})^\intercal \mathbf{W}(\mathbf{x} - \mathbf{A}\mathbf{s})$
8. $f(\mathbf{s}) = (\mathbf{x} - \mathbf{A}\mathbf{s})^\intercal \mathbf{W}(\mathbf{x} - \mathbf{A}\mathbf{s})$

Function composition

Let $f(\mathbf{x}) = g(u(\mathbf{x}))$, then $\nabla f(\mathbf{x}) = \frac{\partial g}{\partial u}\frac{\partial u}{\partial \mathbf{x}}$
Check dimensions and understand how to write $\frac{\partial g}{\partial u}$.
Examples:

1. ℓ_2 norm of vector: $f(\mathbf{x}) = \|\mathbf{x}\|_2$
2. Bilinear form: $f(\mathbf{x}) = u^\intercal(\mathbf{x})\mathbf{R}v(\mathbf{x})$, $\mathbf{R} \in \mathbb{R}^{m\times n}$
3. Exponent: $f(\mathbf{x}) = -e^{-\mathbf{x}^\intercal \mathbf{x}}$

Recap

- Derivative by scalar
- Derivative by vector
- Derivative by matrix
- Derivative of function composition