Robotics
Exercise 1

Marc Toussaint
Lecturer: Duy Nguyen-Tuong
TAs: Philipp Kratzer, Janik Hager, Yoojin Oh
Machine Learning & Robotics lab, U Stuttgart
Universitätsstraße 38, 70569 Stuttgart, Germany

April 22, 2021

1 Matrix equations (4 points)

a) Let X, A be arbitrary matrices, A invertible. Solve for X:

$$XA + A^\top = \mathbf{I}$$

$X = (I - A^T) A^{-1}$

b) Let X, A, B be arbitrary matrices, $(C - 2A^\top)$ invertible. Solve for X:

$$X^\top C = [2A(X + B)]^\top$$

$X^T C = (B^T + X^T) A^T \cdot 2$

$X^T (C - 2A^T) = 2B^T A^T$

$X^T = 2B^T A^T (C - 2A^T)^{-1}$

$X = 2(C^T - 2A)^{-1} AB$

c) Let $x \in \mathbb{R}^n, y \in \mathbb{R}^d, A \in \mathbb{R}^{d\times n}$. A obviously *not* invertible, but let $A^\top A$ be invertible. Solve for x:

$$(Ax - y)^\top A = \mathbf{0}_n^\top$$

$A^T (Ax - y) = 0_n$

$\therefore x = (A^T A)^{-1} \cdot A^T \cdot y$

d) As above, additionally $B \in \mathbb{R}^{n\times n}$, B positive-definite. Solve for x:

$$(Ax - y)^\top A + x^\top B = \mathbf{0}_n^\top$$

$B^T x + A^T (Ax - y) = 0_n$

$(B^T + A^T A) x = A^T y$

$\because B^T$ positive-definite, $A^T A$ semi-positive-definite $\therefore B^T + A^T A$ still positive-definite

$\therefore x = (B^T + A^T A)^{-1} A^T y$

$\underline{x}^T \underline{x} = \sum x_i^2$

$\frac{\partial \underline{x}^T \underline{x}}{\partial \underline{x}} = \begin{pmatrix} \frac{\partial \sum x_i^2}{\partial x_1} \\ \vdots \\ \frac{\partial \sum x_i^2}{\partial x_n} \end{pmatrix} = \begin{pmatrix} 2x_1 \\ \vdots \\ 2x_n \end{pmatrix} = 2\underline{x}$

2 Vector derivatives (5 points)

Let $x \in \mathbb{R}^n,\ y \in \mathbb{R}^d,\ f, g : \mathbb{R}^n \to \mathbb{R}^d,\ A \in \mathbb{R}^{d\times n},\ C \in \mathbb{R}^{d\times d}$. (Also ~~provide the dimensionality~~ of the results.)

a) What is $\frac{\partial}{\partial x} x$? $= I$

b) What is $\frac{\partial}{\partial x} [x^\top x]$? $= 2X$

c) What is $\frac{\partial}{\partial x} [f(x)^\top f(x)]$? $= 2 f(x) \cdot J_f(x)$, $J_f(x) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & & \vdots \\ \frac{\partial f_d}{\partial x_1} & & \cdots & \frac{\partial f_d}{\partial x_n} \end{pmatrix}$ $d \times n$

d) What is $\frac{\partial}{\partial x} [f(x)^\top C g(x)]$? $= J_f^T(x) \cdot C \cdot g(x) + J_g^T(x) \cdot C^T \cdot f(x)$

$x^T C x$, $\frac{\partial Cx}{\partial x} = C$, $\frac{\partial \underset{1\times n}{d} \cdot x}{\partial x} = d^T$

$\underset{n\times n}{C} \underset{n\times 1}{X} + \underset{1\times n \cdot n \times 1}{(x^T C)^T}$

$A^T C \cdot (Ax - y) + A^T C^T (Ax - y)$

e) Let B and C be symmetric (and pos.def.). What is the minimum of $(Ax - y)^\top C(Ax - y) + x^\top Bx$ w.r.t. x ?

$(n\times d) \times (d \times d) \times (d \times 1)$ $\quad (n \times d) \times (d \times d) \times (d \times 1)$

$J_f^T \cdot C \cdot g + (\underset{1\times d}{f^T} \cdot \underset{d\times d}{C} \cdot \underset{d\times n}{J_g})^T$

$(C_1\, C_2 \cdots C_d) \begin{pmatrix} g_1 \\ \vdots \\ g_d \end{pmatrix} = C_1 g_1 + C_2 g_2 + \cdots + C_d g_d$

$\frac{\partial (\,)}{\partial g} = \frac{\partial\, C_1 g_1 + C_2 g_2 \cdots}{\partial g_1} \cdots \frac{\partial (\,)}{\partial g} = C$

$\frac{\partial F(x)}{\partial x} = 2 A^T C (Ax - y) + 2Bx \overset{!}{=} 0$

$x = (A^T C A + B)^{-1} A C y$

$x^T A^T D^T D A x = (DAx)^T \geq 0$ Semi-po. $+ B$ positive-def. (invertible)

3 Optimization (3 points)

Given $x \in \mathbb{R}^n$, $f :\ \mathbb{R}^n \to \mathbb{R}$, we want to find $\text{argmin}_x f(x)$. (We assume f is uni-modal.)

a) What 1st-order optimization methods (querying $f(x), \nabla f(x)$ in each iteration) do you know?

b) What 2nd-order optimization methods (querying $f(x), \nabla f(x), \nabla^2 f(x)$ in each iteration) do you know?

c) What is backtracking line search? find a step size α, $\Delta x = \alpha \cdot p$ (p is search direction) to minimize $f(x + \Delta x)$ in an iterative way (not analytic). (using 2 control parameter $\tau, c \in (0,1)$)

Algorithm: ① $t = -c(p^T \cdot \nabla f)$

② until $f(x) - f(x + \alpha p) \geq \alpha t$, iterate $\alpha = \tau \cdot \alpha$.

③ return α.

$f(x + \Delta x) = f(x) + J(x) \cdot \Delta x + \frac{1}{2} \Delta x^T \cdot H(x) \Delta x + \ldots$

a) gradient descent (Steepest) — fixed learning rate; step size chosen to minimized objective function.

$f(x + \Delta x) = f(x) + J(x) \cdot \Delta x \quad \therefore \Delta x = -J(x)$

$\nabla f(x) = f(x + \Delta x) - f(x) = J(x) \cdot \Delta x$

b) second order: $\underset{\Delta x}{\text{argmin}}\ f(x) + J(x) \cdot \Delta x + \frac{1}{2} \Delta x^T H(x) \cdot \Delta x$

$\nabla^2 f(x) = \nabla f(x + \Delta x) - \nabla f(x) = \frac{1}{2} \Delta x^T \cdot H(x) \cdot \Delta x$

$\frac{\partial H(x)}{\partial \Delta x} = J(x) + H(x) \cdot \Delta x \overset{!}{=} 0 \quad \therefore H(x) \cdot \Delta x = -J(x)$