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:

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

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

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

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

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

$x^T x = \sum x_i^2$

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

2 Vector derivatives (5 points)

Use numerator layout notation.

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 $\frac{\partial}{\partial x} x$? $= I_n$.

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

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

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

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

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 ?

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

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

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

can be written as $x^T A^T D^T D A x = (DAx)^T \geq 0$ semi-po. $+ B$ positive-def. (invertible) $\Rightarrow$ invertible

3 Optimization (3 points)

Given $x \in \mathbb{R}^n,\ f : \mathbb{R}^n \to \mathbb{R}$, we want to find $\operatorname{argmin}_x f(x)$. (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 $f(x) - f(x + \alpha \cdot p) \geq \alpha t$, iterate $\alpha = \tau \cdot \alpha$.

③ return α.

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

fixed learning rate

a) gradient descent (steepest) $f(x + \Delta x) = f(x) + J(x) \cdot \Delta x$ $\therefore \Delta x = -J(x)$ $\nabla f(x) = f(x + \Delta x) - f(x) = J(x) \cdot \Delta x$

step size chosen to minimized objective function.

b) second order: $\operatorname*{argmin}_{\Delta x} f(x) + J(x) \cdot \Delta x + \frac{1}{2} \Delta x^T H(x) \cdot \Delta x$ $\nabla^2 f(x) = \nabla f(x + \Delta x) - \nabla f(x) = \frac{1}{2} \Delta x^T \cdot H(x) \cdot \Delta x$

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