Optimization Algorithms
Weekly Exercise 3

Marc Toussaint
Learning & Intelligent Systems Lab, TU Berlin
Marchstr. 23, 10587 Berlin, Germany

Winter 2024/25

1 Gauss-Newton basics

a) Let $f(x) = \|\phi(x)\|^2$ be a sum-of-squares cost for the features $\phi : \mathbb{R}^n \to \mathbb{R}^d$. Derive the Gauss-Newton approximation $\nabla^2 f(x) \approx 2J(x)^\top J(x)$ from a linear approximation (first order Taylor) of features ϕ, with the Jacobian $J(x) = \frac{\partial}{\partial x}\phi(x)$.

b) Show that for any vector $v \in \mathbb{R}^n$ the matrix $vv^\top$ is symmetric and semi-positive-definite.[1] Based on this, argue that the Gauss-Newton approximation $J(x)^\top J(x)$ is also symmetric and semi-positive-definite.

2 Conjugate Gradient

The conjugate gradient methods initialized $\delta_0 = g = -\nabla f(x_0)$ and then iterates the following steps:

```
1: α ← argmin_α f(x + αδ)                          // exact line search
2: x ← x + αδ
3: g' ← g,  g = −∇f(x)                              // store old and compute new gradient
4: β ← max{ g^⊤(g − g') / (g'^⊤ g'), 0 }
5: δ ← g + βδ                                       // conjugate descent direction
```

Consider the quadratic cost function $f(x) = \frac{1}{2}x^\top A x + b^\top x + c$ with $A = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$, $b = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and $c = 0$ whose minimum is achieved at $x^* = (0, 0)$.

a) Compute, by hand, two iterations of the conjugate gradient descent from $x_0 = (1, 1)$ and from $x_0 = (-1, 2)$, respectively.

b) Show that the first and second descent directions are A-orthogonal, i.e., $\delta_0^\top A \delta_1 = 0$.

3 Solve by Sketch and check KKT

In this exercise, for each of the following problems do the following:

– Sketch the problem on paper and, without much maths, figure out where the optimum x^* is.
– State which contraints are active at x^*.
– Compute (by hand) the gradient ∇f and gradients ∇g_i, ∇h_j of active constraints at x^*.
– Identify dual parameters λ_i, κ_j so that the stationarity (1st KKT) condition holds at x^*.

a) A 1D problem:

$$\min_{x \in \mathbb{R}^1} x \quad \text{s.t.} \quad \sin(x) = 0 \;, \quad x^2/4 - 1 \le 0$$

[1] A matrix $A \in \mathbb{R}^{n \times n}$ is semi-positive-definite simply when for any $x \in \mathbb{R}^n$ it holds $x^\top A x \ge 0$. Intuitively: A might be a metric as it "measures" the norm of any x as positive. Or: If A is a Hessian, the function is (locally) convex.

b) 2D problems: (Note that $1^\top x = \sum_i x_i$ is a simple linear cost.)

$$\min_{x\in\mathbb{R}^2} 1^\top x \quad \text{s.t.} \quad |x|^2 - 1 \le 0$$

c)

$$\min_{x\in\mathbb{R}^2} 1^\top x \quad \text{s.t.} \quad |x|^2 - 1 \le 0, \quad -x_1 \le 0$$

d)

$$\min_{x\in\mathbb{R}^2} 1^\top x \quad \text{s.t.} \quad x^2 - 1 \le 0, \; x_2^2 - x_1 \le 0$$