

Optimization Algorithms

Weekly Exercises 7

Marc Toussaint

Learning & Intelligent Systems Lab, TU Berlin

Marchstr. 23, 10587 Berlin, Germany

Winter 2024/25

1 Differentiable Optimization

Let's again consider the problem we had in exercise 5, but this time with parameter $\theta \in \mathbb{R}$,

$$\min_{x \in \mathbb{R}^2} 1^\top x \quad \text{s.t.} \quad g(x) \leq 0, \quad g(x) = \begin{pmatrix} x^\top x - 1 \\ -x_1 + \theta \end{pmatrix}.$$

Here, the second inequality $x_1 \geq \theta$ is parameterized and we want to know $\frac{\partial}{\partial \theta} x^*(\theta)$.

- As in exercise 5, derive the optimum $x^*(\theta)$, but now as a function of θ . (You may assume that both constraints are active. What does this imply for possible values of θ ?)
- Determine the derivative $\frac{\partial}{\partial \theta} x^*(\theta)$ based on the parameteric solution you found in a).
- Derive the KKT Jacobian $\frac{\partial}{\partial x \kappa \lambda} r \in \mathbb{R}^{4 \times 4}$ at the optimum, and the vector $\frac{\partial}{\partial \theta} r \in \mathbb{R}^4$ (see slide 13).
- Check that the first two entries of $\frac{\partial}{\partial \theta} F = -[\frac{\partial}{\partial x \kappa \lambda} r]^{-1} \frac{\partial}{\partial \theta} r$ coincide with the gradient $\frac{\partial}{\partial \theta} x^*(\theta)$ you found in b).

2 Trust Region

Consider a function $f(x) = \frac{1}{2} x^\top A x + b^\top x$ for symmetric positive definite A . For a given x_0 , we would like to solve the problem

$$\min_x f(x) \quad \text{s.t.} \quad \|x - x_0\|^2 \leq \alpha$$

This describes a Newton step under a *Trust Region* approach: It considers a 2nd-order Taylor approximation f , and computes an optimal step under step length limit $\sqrt{\alpha}$, which is called trust region radius.

- Try to solve this problem analytically as far as you can get. Eventually you will not be able to solve it completely analytically and be left with an equation for the dual parameter λ . What numerical algorithm could you use (in an 'inner loop') to efficiently solve for the correct λ ?
- Make explicit, how the λ that arises in the Trust Region step is related to the damping (aka. Levenberg-Marquardt parameter) of a Newton step.
- Derive the dual function $l(\lambda) = \min_x L(x, \lambda)$ for the above problem. Can one solve $\max_{\lambda \geq 0} l(\lambda)$ analytically?