Optimization Algorithms
Exercise 4

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

Winter 2020/21

1 Equality Constraint Penalties and Augmented Lagrangian

The squared penalty approach to solving an equality constrained optimization problem minimizes in each inner loop:

$$\min_x \; f(x) + \mu \sum_{i=1}^m h_i(x)^2 \; . \tag{1}$$

The Augmented Lagrangian method adds a Lagrangian term and minimizes in each inner loop:

$$\min_x \; f(x) + \mu \sum_{i=1}^m h_i(x)^2 + \sum_{i=1}^m \lambda_i h_i(x) \; . \tag{2}$$

Assume that we first minimize (1) we end up at a minimum $\bar{x}$.
Now prove that setting $\lambda_i = 2\mu h_i(\bar{x})$ will, if we assume that the gradients $\nabla f(x)$ and $\nabla h(x)$ are (locally) constant, ensure that the minimum of (2) fulfills the constraints $h(x) = 0$.

2 Alternative Barriers & Penalties

Propose 3 alternative barrier functions, and 3 alternative penalty functions. To display functions, gnuplot is useful, e.g., `plot -log(-x)`.

3 Log Barrier method

Consider the following constrained problem

$$\min_x \sum_{i=1}^n x_i \quad \text{s.t.} \quad g(x) \le 0 \tag{3}$$

$$g(x) = \begin{pmatrix} x^\top x - 1 \\ -x_1 \end{pmatrix} \tag{4}$$

a) First, assume $x \in \mathbb{R}^2$ is 2-dimensional, and draw on paper what the problem looks like and where you expect the optimum.

b) Implement the Log Barrier method. Tips:
   - Initialize $x = (\frac{1}{2}, \frac{1}{2})$ and $\mu = 1$
   - First code an inner loop:
     - In each iteration, first compute the gradient of the log-barrier function. Recall that

$$F(x;\mu) = f(x) - \mu \sum_i \log(-g_i(x)) \tag{5}$$

$$\nabla F(x;\mu) = \nabla f - \mu \sum_i (1/g_i(x)) \nabla g_i(x) \tag{6}$$

  - Then perform a backtracking line search along $-\nabla F(x;\mu)$. In particular, backtrack if a step goes beyond the barrier (where $g(x) \not\leq 0$ and $F(x;\mu) = \infty$).
  - Iterate until convergence; let's call the result $x^*(\mu)$. Further, compute $\lambda^*(m) = -(\mu/g_1(x), \mu/g_2(x))$ at convergence.
- Decrease $\mu \leftarrow \mu/2$, recompute $x^*(\mu)$ (with the previous x^* as initialization) and iterate this.

Plot the optimization path and report on the total number of function/gradient evaluations needed. What does λ converge to? Is this intuitive?

Note: The path $x^*(\mu) = \text{argmin}_x F(x;\mu)$ (the optimum in dependence of μ) is called *central path.*

4 Squared Penalty method

Consider the same problem as in the previous exercise

Implement the Squared Penalty method. Choose as a start point $x = (\frac{1}{2}, \frac{1}{2})$. Plot the optimization path and report on the total number of function/gradient evaluations needed.