Maths for Intelligent Systems
Exercise 7

Marc Toussaint
Machine Learning & Robotics lab, U Stuttgart
Universitätsstraße 38, 70569 Stuttgart, Germany

January 18, 2019

1 Lagrangian Method of Multipliers

In a previous exercise we defined the "hole function" $f^c_{\text{hole}}(x)$, assume conditioning $c = 10$ and use the Lagrangian Method of Multipliers to solve on paper the following constrained optimization problem in $2D$.

$$\min_x f^c_{\text{hole}}(x) \quad \text{s.t.} \quad h(x) = 0 \tag{1}$$

$$h(x) = v^\top x - 1 \tag{2}$$

Near the very end, you won't be able to proceed until you have special values for v. Go as far as you can without the need for these values.

2 Lagrangian and dual function

(Taken roughly from 'Convex Optimization', Ex. 5.1)

A simple example. Consider the optimization problem

$$\min_x \; x^2 + 1 \qquad \text{s.t.} \quad (x-2)(x-4) \le 0$$

with variable $x \in \mathbb{R}$.

a) Derive the optimal solution x^* and the optimal value $p^* = f(x^*)$ by hand.

b) Write down the Lagrangian $L(x, \lambda)$. Plot (using gnuplot or so) $L(x, \lambda)$ over x for various values of $\lambda \ge 0$. Verify the lower bound property $\min_x L(x, \lambda) \le p^*$, where p^* is the optimum value of the primal problem.

c) Derive the dual function $l(\lambda) = \min_x L(x, \lambda)$ and plot it (for $\lambda \ge 0$). Derive the dual optimal solution $\lambda^* = \text{argmax}_\lambda\, l(\lambda)$. Is $\max_\lambda l(\lambda) = p^*$ (strong duality)?

3 Equality Constraint Penalties and Augmented Lagrangian

Take a squared penalty approach to solving a constrained optimization problem

$$\min_x \; f(x) + \mu \sum_{i=1}^m h_i(x)^2 \; . \tag{3}$$

The Augmented Lagrangian method adds a Lagrangian term

$$\min_x \; f(x) + \mu \sum_{i=1}^m h_i(x)^2 + \sum_{i=1}^m \lambda_i h_i(x) \; . \tag{4}$$

Assume that if we minimize (3) we end up at a solution $\bar{x}$ for which each $h_i(\bar{x})$ is reasonable small, but not exactly zero. Prove, in the context of the Augmented Lagrangian method, 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 (4) fulfills the constraints $h(x) = 0$.

Tip: Think intuitive. Think about how the gradient that arises from the penalty in (3) is now generated via the λ_i.