Mathematics for Intelligent Systems
Lecture 9 Homework
(Optimization)

Marc Toussaint, Andrea Baisero

1 Quadratics Review

Take the quadratic function $f_{\text{sq}}(x) = x^\top C x$ with diagonal matrix C and entries $C(i,i) = \lambda_i$.

a) Which 3 fundamental shapes does the 2-dimensional quadratic function f_{sq} assume? Plot the surface of f_{sq} for various values of λ_1, λ_2 (big/small, positive/negative/zero). Could you predict these shapes before plotting them?

b) For which values of λ_1, λ_2 does $\min_x f_{\text{sq}}(x)$ *not* have a solution? For which does it have *infinite* solutions? For which does it have *exactly* 1 solution? Find out empirically first, if you have to, then analytically.

c) Use the eigen-decomposition of a generic (non-diagonal) matrix C to prove that the same 3 basic shapes appear and that the values of λ_1 and λ_2 have the same implications on the existence of one or more solutions. (In this scenario, λ_1 and λ_2 don't indicate the diagonal entries of C, but its *eigenvalues*).

2 Lagrangian Method of Multipliers

Take the "hole function" $f^c_{\text{hole}}(x) = 1 - \exp(-x^\top C x)$, and assume C is a diagonal matrix with $C(i,i) = c^{\frac{i-1}{n-1}}$, with n the number of dimensions. Assume a conditioning[1] $c = 10$.

a) Use the Lagrangian Method of Multipliers to solve on paper the following constrained optimization problem in $2D$.

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

$$v^\top x = 1 \tag{2}$$

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

[1] The word "conditioning" generally denotes the ratio of the largest and smallest Eigenvalue of the Hessian.

3 Backtracking

Consider again the functions:

$$f_{\text{sq}}(x) = x^\intercal C x \tag{3}$$
$$f_{\text{hole}}(x) = 1 - \exp(-x^\intercal C x) \tag{4}$$

with diagonal matrix C and entries $C(i,i) = c^{\frac{i-1}{n-1}}$, with conditioning $c = 10$.

a) Implement gradient descent with backtracking, as described on slide 44 (Algorithm2 Plain gradient descent). Test the algorithm on $f_{\text{sq}}(x)$ and $f_{\text{hole}}(x)$ with start point $x_0 = (1,1)$. To judge the performance, create the following plots:

- function value over the number of function evaluations.
- number of inner (line search) loops over the number of outer (gradient descent) loops.
- function surface, this time including algorithm's search trajectory.

b) Test also the *alternative* in step 3. Further, how does the performance change with ρ_{ls} (the backtracking stop criterion)?