Homework #3: Non-linear Optimization Solutions

2.a Theoretical Question. Consider the linear system that we currently have and we want to minimize the quadratic cost

$$\frac{1}{2}\sum_t x_t^\top Q x_t \tag{1}$$

$$x_{t+1} = Ax_t + Bu_t \tag{2}$$

Hence, we have a linear quadratic regulator problem. Derive the gradient update for the action variables for both optimization methods: shooting and collocation. In the case of collocation, do not include the update due to the constraints.

Explain in a few lines why the shooting method might become unstable while the collocation method does not.

Solution:

- Shooting: For the shooting case, we need to incorporate the constraints in the objective. Then, the state x_t is re-written as

  $$x_t = A^t x_0 + \sum_{k=1}^{t} A^{t-k} B u_{k-1} \tag{3}$$

  Then, using the chain rule:

  $$\nabla_{u_j} \frac{1}{2}\sum_{t=0}^{H} x_t^\top Q x_t = \sum_{t=j+1}^{H} x_t^\top Q \nabla_{u_j} x_t \tag{4}$$

  Given $j \geq t$, we have that $\nabla_{u_j} x_t = 0$. For $j < t$, the gradient $\nabla_{u_j} x_t$ is the gradient w.r.t u_j of Equation 3.

  $$\nabla_{u_j} x_t = \nabla_{u_j} A^t x_0 + \sum_{k=1}^{t} A^{t-k} B u_{k-1} = \sum_{k=1}^{t} \nabla_{u_j} A^{t-k} B u_{k-1} = A^{t-j-1} B \tag{5}$$

  Combining the previous three equations we have the desired result:

  $$\nabla_{u_j} \frac{1}{2}\sum_{t=0}^{H} x_t^\top Q x_t = \sum_{t=j+1}^{H} \left(A^t x_0 + \sum_{k=1}^{t} A^{t-k} B u_{k-1} \right)^\top Q A^{t-j-1} B$$

- Collocation: The collocation case, if we do not take into account the constraints, the solution is much simpler. Given that the cost is not dependent on the actions we have that:

  $$\nabla_{u_j} \frac{1}{2}\sum_{t=0}^{H} x_t^\top Q x_t = 0 \tag{6}$$

The gradient with respect to the state variables is

$$\nabla_{x_j} \frac{1}{2} \sum_{t=0}^{H} x_t^\top Q x_t = Q x_j \tag{7}$$

The shooting method is much more unstable than the collocation method, specially when the horizon H is large and the eigenvalues of the matrix A are larger than 1. The reason is that the gradient is the result of a summation of exponentially large elements. Note that if all the elements in the gradient were exponentially large but with the same magnitude, one could simply choose an exponentially small learning rate. However, in the shooting method, the gradient of each action has a different (exponentially large) magnitude.

In contrast, the collocation method presents an update that is equal in magnitude to all the variables regardless the horizon used.

MPC vs. Policy with noise. Why does the MPC method perform better than having a policy? Is there a way we could make the performance of the policy better?

Solution: Some of the reasons why MPC is better than a policy in this case are: 1) The policy has not been learned in the presence of noise, and hence there is a distribution shift between the training of the policy and the testing of it. Note that both approaches are closed-loop. Another reason could be that MPC is a non-paramtetric approach that is more expressive than the policy. As a result, the MPC is can obtain the sequence of actions that are optimal (or almost optimal), while the policy might not be able to represent it.

In order to improve the performance of the policy, we should learn the policy with stochastic dynamics (i.e., with noise in the dynamics) and make it more expressive (for instance, using a deeper and/or wider neural network).