Reinforcement Learning and Optimal Control

IFT6760C, Fall 2021

Pierre-Luc Bacon

November 15, 2021

Bolza Problems

$$
\begin{aligned}
&\text{minimize} \;\; c_T(x_T) + \sum_{t=1}^{T-1} c_t(x_t, u_t) \\
&\text{subject to} \;\; x_{t+1} = f_t(x_t, u_t), \;\; t = 1, \ldots T-1 \\
&\text{given} \;\; x_1
\end{aligned}
$$

We then know that if there exists feasible $x_1, \ldots, x_T, u_1, \ldots, u_{T-1}$, then it must be that there exists a unique set $\{\lambda_t^\star\}_1^{T-1}$ such that $DL(x^\star, u^\star, \lambda^\star) = 0$ in:

$$
L(x, u, \lambda) \triangleq c_T(x_T) + \sum_{t=1}^{T-1} c_t(x_t, u_t) + \sum_{t=1}^{T-1} \lambda_t(f_t(x_t, u_t) - x_{t+1}) \;\; .
$$

Equality-Constrained Problem (ECP)

Let $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^m$ in:

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & h(x) = 0 \quad . \end{aligned}$$

First-order optimality condition tells us that if $x^\star$ is a regular feasible local minimum then there must be a $\lambda^\star$ such that $DL(x^\star, \lambda^\star) = 0$ (for both partial derivatives).

Idea: Let's view this problem as a root-finding problem, ie solve the nonlinear equations: $DL(x, \lambda) = 0$ in the variables x and λ.

Newton's Method for solving ECP

Let $y \triangleq (x, \lambda)$, so that $\varphi(y) \triangleq DL(x, \lambda)$. The iterates of Newton's method would look like:

$$y_{t+1} = y_t - [D\varphi(y_t)]^{-1}\varphi(y_t) \quad .$$

Note that $L : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, therefore $\varphi(y) : \mathbb{R}^{n+m} \to \mathbb{R}^{n+m}$ and $[D\varphi(y)] \in \mathbb{R}^{(n+m)+(n+m)}$:

$$\begin{aligned}
\begin{pmatrix} x_{t+1} \\ \lambda_{t+1} \end{pmatrix} &= \begin{pmatrix} x_t \\ \lambda_t \end{pmatrix} - \begin{pmatrix} D_1^2 L(x_t, \lambda_t) & D_2 D_1 L(x_t, \lambda_t) \\ D_1 D_2 L(x_t, \lambda_t) & D_2^2 L(x_t, \lambda_t) \end{pmatrix}^{-1} \begin{pmatrix} D_1 L(x_t, \lambda_t) \\ D_2 L(x_t, \lambda_t) \end{pmatrix} \\
&= \begin{pmatrix} x_t \\ \lambda_t \end{pmatrix} - \begin{pmatrix} D^2 f(x_t) + \lambda D^2 h(x_t) & [Dh(x_t)]^\top \\ Dh(x_t) & 0 \end{pmatrix}^{-1} \begin{pmatrix} Df(x_t) + \lambda_t Dh(x_t) \\ h(x_t) \end{pmatrix}
\end{aligned}$$

Solving for Delta

As usual, we don't want to take an explicit inverse and we write:

$$[D\varphi(y_t)]\Delta_t = \varphi(y_t) \;\text{ and }\; y_{t+1} = y_t - \Delta_t \quad .$$

In our setting, we must solve for Δ_t in:

$$\begin{pmatrix} D^2 f(x_t) + \lambda D^2 h(x_t) & [Dh(x_t)]^\top \\ Dh(x_t) & 0 \end{pmatrix} \Delta_t = \begin{pmatrix} Df(x_t) + \lambda_t Dh(x_t) \\ h(x_t) \end{pmatrix}$$

We refer to that matrix (that we want to avoid inverting explicitely) the **KKT matrix**.

Assumptions

The KKT matrix is nonsingular under the following assumptions (see Nocedal 18.1):

1. The Jacobian $Dh(x)$ has full row rank
2. The Hessian $D_1^2 L(x, \lambda)$ is positive definite on the tangent space of the constraints. That is given any $z \neq 0$ such that $Dh(x)z = 0$, then $z^\top D_1^2 L(x, \lambda) z > 0$.

(these were the assumptions of the Lagrange multiplier theorem in the Bertsekas book)

Under those assumptions, this method converges quadratically if the primal-dual pair is chosen close enough to the optimum.

Approximation by a QP

The idea behind this method is to approximate the ECP by a simplier one. If f is twice continuously differentiable, then function can be approximated locally by a quadratic model:

$$\tilde{f}_k(x) \triangleq f(x_t) + Df(x_t)(x - x_t) + \frac{1}{2}(x - x_t)^\top D^2 f(x_t)(x - x_t) \quad .$$

Similarly, we can approximate the constraint by a linear model:

$$\tilde{h}_k(x) \triangleq h(x_t) + Dh(x_t)(x - x_t) \quad .$$

Approximation by a QP

We now have a Quadratic Program (QP):

$$\begin{aligned} \text{minimize} \quad & f(x_t) + Df(x_t)p + \frac{1}{2}p^\top D^2 f(x_t)p \\ \text{subject to} \quad & h(x_t) + Dh(x_t)p = 0 \end{aligned} \quad ,$$

where the optimization variable is the vector p. Note that the quadratic term acts as a *regularizer*, penalizing for values that are too far from where our approximation is taken. Instead of using the hessian of f, we choose instead to take $D_1^2 L(x_t, \lambda_t)$.

Sequential Quadratic Program (SQP)

The idea behind SQP is to keep on solving a QP instead of the original ECP. The final form is:

$$
\begin{aligned}
\text{minimize} \quad & f(x_t) + Df(x_t)p + \frac{1}{2}p^\top D_1^2 L(x_t, \lambda_t)p \\
\text{subject to} \quad & h(x_t) + Dh(x_t)p = 0
\end{aligned} \quad ,
$$

Quadratic Programs in General

An equality-constrained QP is a problem of the form:

$$\begin{aligned} \text{minimize} \quad & c^\top x + \frac{1}{2}x^\top Qx \\ \text{subject to} \quad & Ax = b \quad . \end{aligned}$$

The Lagrangian is:

$$L(x, \lambda) = c^\top x + \frac{1}{2}x^\top Qx + \lambda(Ax - b) \quad .$$

The first-order optimality condition tell us that $DL(x^\star, \lambda^\star) = 0$, hence:

$$\begin{aligned} D_1L(x, \lambda) = c^\top + x^\top Q + \lambda Ax \\ D_2L(x, \lambda) = Ax - b \quad . \end{aligned}$$

Matrix Form

The first-order condition then coincide with the linear system:

$$\begin{pmatrix} Q & A^\top \\ A & 0 \end{pmatrix} \begin{pmatrix} x^\star \\ \lambda^\star \end{pmatrix} = \begin{pmatrix} -c \\ b \end{pmatrix}$$