4. Optimal Control

4.1. **The Problem, and Heuristics.** Let $y(t)$ be a quantity (e.g. production rate at a factory, portfolio value, etc.), that satisfies an ordinary differential equation:

$$\partial_t y(t) = F(y(t), \alpha_q(t), \alpha_2(t), \ldots).$$

Above, $\alpha_1, \alpha_2, \ldots$ are parameters that may depend on time and which influence F.

Example:

$$\partial_t y(t) = ry(t) - \alpha(t), \qquad y(0) = x. \tag{4.1}$$

This can of course be solved, and the solution

$$y(t) = \mathrm{e}^{rt} \left(x - \int_0^t \mathrm{e}^{-rs}\, \alpha(s)\, \mathrm{d}s \right)$$

depends on $\alpha(s)$. Let now h be a utility function, which can depend on y and α; for example, the utility function could be lower when α is large (cost of controlling something), and higher when y is high.

The *problem of optimal control* is this: find *control parameters* $\alpha_1(s), \alpha_2(s), \ldots$ such that h is maximal when averaged over time. In other words, we have to find

$$u(x,t) = \max_{\alpha} \int_t^T h(y(s), \alpha(s))\, \mathrm{d}s, \tag{4.2}$$

where initially (at time t), the value of $y(t) = x$. Importantly, $y(s)$ itself depends both on x and on the control $\alpha(s)$ through the ODE, e.g. (4.1).

There are many variants of the optimal control problem. One can let h depend explicitly on time. The most important instance of this is *discounting*, when we have $\mathrm{e}^{-rs}\, h(y(s), \alpha(s))$ under the integral. Or, we can have a final time utility $g(y(T))$ added to u.

The example (4.1) corresponds to an optimal consumption problem. There $y(t)$ is the wealth at time t, $\alpha(t)$ is the rate of consumption, y_0 is the initial wealth; wealth that has not been consumed earns interest at rate r. The control problem is to maximize

$$u(x,t) = \int_0^T \mathrm{e}^{-\rho s}\, h(\alpha(s))\, \mathrm{d}s. \tag{4.3}$$

h is usually a concave function, such as $h(\alpha) = \alpha^{1/2}$ or, more generally, $h(\alpha) = \alpha^\gamma$ with $0 < \gamma < 1$. The concavity has a nice interpretation: while it may make you happier to spend more money per day, it will not make you twice as happy; or, the less you are used to get by with, the more you will appreciate a bit more money to spend. T is the final time by which you should have spent it all, wealth that is left at that time will not benefit you.

It is a bit confusing that (4.3) seems not to depend on y at all. But in fact it does. Firstly, x is the starting point of $y(s), s \geqslant t$, and secondly, we have the condition that $y(s) \geqslant 0$ for all s, i.e. we adhere to the old-fashioned strategy that you cannot

spend more than you have. This in turn restricts the controls α, since a control that is too large at the beginning will have to be zero after $y(s)$ hits zero, which happens sooner if α is large. Also, of course $\alpha \geqslant 0$, i.e. the consumption should be non-negative.
In the following we will give a general strategy to solve optimal control problems. The problem of optimal consumption will then be done in an exercise.

4.2. **Solving the optimal control problem: dynamic programming.** A general optimal control problem consists of the following parts:
The ODE:

$$\partial_s \boldsymbol{y}(s) = F(\boldsymbol{y}(s), \boldsymbol{\alpha}(s)), \qquad \boldsymbol{y}(t) = \boldsymbol{x},$$

with $\boldsymbol{y} : \mathbb{R} \to \mathbb{R}^n$, $\boldsymbol{\alpha} : \mathbb{R} \to \mathbb{R}^m$ and $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$.
The control constraint: $\boldsymbol{\alpha}(s) \in A \subset \mathbb{R}^m$ for all s. This constraint may not always be present, and A may even depend on time.
The state constraint: $y(t) \in Y \subset \mathbb{R}^n$ for all t. This is usually a difficult constraint. The way we deal with it in the optimal consumption problem is to ignore it and verify afterwards that the solution fulfils it. This is not always possible.
The control problem: Let $h : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be the running utility function, and $g : \mathbb{R}^n \to \mathbb{R}$ be the final utility function. The control problem is to find the optimal value function

$$u(\boldsymbol{x}, t) = \max_{\boldsymbol{\alpha} \in A, \boldsymbol{y} \in Y} \left(\int_t^T h(\boldsymbol{y}_{\boldsymbol{\alpha},\boldsymbol{x}}(s), \boldsymbol{\alpha}(s)) \, \mathrm{d}s + g(\boldsymbol{y}_{\boldsymbol{\alpha},\boldsymbol{x}}(T)) \right). \tag{4.4}$$

The subscript on the function $\boldsymbol{y}$ is there to remind us that it actually depends on the starting point and the control, but it will be often omitted in the notation.
The idea when solving this is to work backwards from the final time T. We ignore the state constraint for now.
1) For $t = T$, it is clear that there is nothing to control, and $u(\boldsymbol{x}, T) = g(\boldsymbol{x})$.
2) Let δt be a very small time step. Let us consider $t = T - \delta t$. We try to optimize (4.4) by using a control $\boldsymbol{\alpha}$ that is constant on $[T - \delta t, T]$. In the limit when $\delta t \to 0$ there is hope that this is good enough. We approximate the solution of the ODE by Taylor expansion:

$$\boldsymbol{y}(t + s) \approx \boldsymbol{y}(t) + \partial_t \boldsymbol{y}(t) s = \boldsymbol{y}(t) + F(\boldsymbol{y}(t), \boldsymbol{\alpha}(t)) s \tag{4.5}$$

for $s \in [0, \delta t]$. We also approximate $h(\boldsymbol{y}(t+s), \boldsymbol{\alpha}(t+s)) \approx h(\boldsymbol{y}(t)\boldsymbol{\alpha}(t))$ (zero order Taylor), and

$$\int_{t=T-\delta t}^{T} h(\boldsymbol{y}(s), \boldsymbol{\alpha}(s)) \, \mathrm{d}s \approx (T - t) h(\boldsymbol{y}(t), \boldsymbol{\alpha}(t)) = \delta t h(\boldsymbol{y}(t), \boldsymbol{\alpha}(t)). \tag{4.6}$$

Thus

$$
\begin{aligned}
u(\boldsymbol{x}, T-\delta t) &= \max_{\boldsymbol{\alpha}(s)} \left(\int_{T-\delta t}^{T} h(\boldsymbol{y}(s), \boldsymbol{\alpha}(s))\, \mathrm{d}s + g(\boldsymbol{y}(T)) \right) \\
&\approx \max_{\alpha} \left(h(\boldsymbol{x}, \boldsymbol{\alpha})\delta t + g(\boldsymbol{x} + F(\boldsymbol{x}, \boldsymbol{\alpha}))\delta t \right) \\
&= \max_{\alpha} \Big(h(\boldsymbol{x}, \boldsymbol{\alpha})\delta t + u\Big(\boldsymbol{x} + F(\boldsymbol{x}, \boldsymbol{\alpha}), T\Big)\delta t \Big) .
\end{aligned} \tag{4.7}
$$

So, given that we know the optimal value function at time T (which we do, it is g), we also know it at time $T - \delta t$. Indeed, we get it by maximizing the known function on the right hand side of (4.7) over $\boldsymbol{\alpha}$, for all $\boldsymbol{x}$, e.g. by differentiating and looking for zeros. This even gives a numerical scheme for finding the optimal value function and the optimal control, but that scheme is very impractical in higher dimensions. However, the above reasoning will soon lead to a PDE that one can treat easier, at least numerically.
3) Step 2 is now repeated: we know the value function at time $T - \delta t$, so we get it (by step 2) for time $T - 2\delta t$, and so on until we reach the initial time.
The insight that knowing the optimal utility at a time t_1 helps us determine it at an earlier time is important. Indeed, the equation

$$
u(\boldsymbol{x}, t) = \max_{\boldsymbol{\alpha}(s)} \int_{t}^{t_1} h(\boldsymbol{y}(s), \boldsymbol{\alpha}(s))\, \mathrm{d}s + u(\boldsymbol{y}(t_1), t_1),
$$

for $t_1 > t$, can be be seen to be true in a similar way as above. This equation is called *dynamic programming principle* because it is at the basis of the algorithm given above.

4.3. **Hamilton-Jacobi-Bellman (HJB) equation.** Let us replace T by $s + \delta t$ in (4.7). We then have

$$
u(\boldsymbol{x}, s) = \max_{\boldsymbol{\alpha}} \Big(h(\boldsymbol{x}, \boldsymbol{\alpha})\, \delta t + u\Big(\boldsymbol{x} + F(\boldsymbol{x}, \boldsymbol{\alpha})\delta t, s + \delta t\Big)\Big).
$$

We now do a Taylor expansion of the function $\delta t \mapsto u(\boldsymbol{x} + F(\boldsymbol{x}, \boldsymbol{\alpha})\delta t, s + \delta t)$, and find

$$
u(\boldsymbol{x}, s) \approx \max_{\boldsymbol{\alpha}} \Big(h(\boldsymbol{x}, \alpha)\delta t + u(\boldsymbol{x}, s) + \nabla u(\boldsymbol{x}, s) \cdot F(\boldsymbol{x}, \boldsymbol{\alpha})\delta t + \partial_t u(\boldsymbol{x}, s)\, \delta t \Big).
$$

We can now subtract $u(\boldsymbol{x}, s)$ from both sides. The $\approx$ sign really means that there are terms of order $(\delta t)^2$ that we ignored. Dividing by δt and sending $\delta t \to 0$ gives the *Hamilton-Jacobi-Bellman equation* (HJB equation):

$$
\partial_s u(\boldsymbol{x}, s) + \max_{\boldsymbol{\alpha} \in A} \left(\nabla u(\boldsymbol{x}, s) \cdot F(\boldsymbol{x}, \boldsymbol{\alpha}) + h(\boldsymbol{x}, \boldsymbol{\alpha}) \right) = 0. \tag{4.8}
$$

There is a special notation that is often used for the HJB equation: we define the *Hamiltonian* $H(\boldsymbol{p}, \boldsymbol{x})$ as

$$
H(\boldsymbol{p}, \boldsymbol{x}) = \max_{\boldsymbol{\alpha} \in A} (\boldsymbol{p} \cdot F(\boldsymbol{x}, \boldsymbol{\alpha}) + h(\boldsymbol{x}, \boldsymbol{\alpha})), \tag{4.9}
$$

with $\boldsymbol{x}, \boldsymbol{p} \in \mathbb{R}^n$. Then (4.8) reads

$$\partial_s u + H(\nabla u, \boldsymbol{x}) = 0. \tag{4.10}$$

It is clear from (4.10) that for solving a control problem, we have to maximize over $\boldsymbol{\alpha}$ in (4.9) first, and then solve (4.10). In particular, to find the optimal control at space point $\boldsymbol{x}$ and 'momentum' $\boldsymbol{p}$, we need not know the solution u of (4.10). However, of course since we then replace $\boldsymbol{p}$ by ∇u, the solution will be fed back into H.

We have now found the equation for the optimal value function: It is the solution of (4.8) with final data $u(\boldsymbol{x}, T) = g(\boldsymbol{x})$. To know the value function is good enough for e.g. option pricing, when we allow the buyer to change some control parameter: we now know how much an investor can make at most, if they play the game in an optimal way. This should then be the fair price of the option.

But what about actually finding the optimal strategy $\boldsymbol{\alpha}(s)$ for $t \leqslant s \leqslant T$? For this, we have to plug the solution $u(\boldsymbol{x}, s)$ back into the second term of (4.8), and find the argmax for each time. More precisely: from (4.9) we obtain both $H(\boldsymbol{p}, \boldsymbol{x})$ for all $\boldsymbol{p}$ and all $\boldsymbol{x}$, and $\alpha_*(\boldsymbol{p}, \boldsymbol{x})$ as the argmax of the right hand side. Now we solve (4.10) with the H that we just obtained, and the correct final condition. Once this is done, we know $u(\boldsymbol{x}, s)$ for all $\boldsymbol{x}$ and all s, and thus also $\nabla u(\boldsymbol{x}, s)$. The ODE for the optimally controlled $\boldsymbol{y}$ is then

$$\partial_s \boldsymbol{y}(s) = F(\alpha_*(\nabla u(\boldsymbol{y}(s), s), \boldsymbol{y}(s)), s),$$

where now the right hand side only contains known functions of s and $\boldsymbol{y}(s)$. Once we have obtained the solution $\boldsymbol{y}_*(s)$ to this ODE (with initial condition $\boldsymbol{y}(t) = \boldsymbol{x}$), we can finally determine the purely time dependent optimal control for the case when $\boldsymbol{y}(t) = \boldsymbol{x}$. It is given by $\boldsymbol{\alpha}_*(\nabla u(\boldsymbol{y}_*(s), s), \boldsymbol{y}_*(s))$.

The thing we now have to check is that $\boldsymbol{y}(t)$ actually fulfils the state constraints. If it does, everything is fine. But if it does not, we have to go back to the dynamic programming approach and incorporate this at each step. There is no nice theory for this, and we will not do it.

We have derived the HJB equation using a lot of heuristic and non-rigorous steps. But once we have it, we can actually prove that it gives the optimal value function. This is the content of the following

Theorem: *Consider the general optimal control problem introduced at at the beginning of this subsection. Assume that $w(\boldsymbol{x}, t)$ solves the HJB equation* (4.8)*, with final condition $w(\boldsymbol{x}, T) = g(T)$. Assume also that $\boldsymbol{y}$ derived from w in the way discussed above fulfils the state constraint. Then $w(\boldsymbol{x}, t) = u(\boldsymbol{x}, t)$, where u is defined is the solution of the control problem.*

Proof. Assume that w is a solution of the HJB equation with final condition $w(\boldsymbol{x}, T) = g(\boldsymbol{x})$. Let $\boldsymbol{\alpha}_0(s)$ be an arbitrary control, and let $\boldsymbol{y}_0(s)$ be the solution of the controlled ODE with start $\boldsymbol{y}_0(t) = \boldsymbol{x}$ and control $\boldsymbol{\alpha}_0(s)$. Let us investigate

the function $s \mapsto w(\boldsymbol{y}_0(s), s)$. We find

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s} w(\boldsymbol{y}(s), s) &= \partial_s w(\boldsymbol{y}(s), s) + \nabla w(\boldsymbol{y}(s), s) \cdot \partial_s \boldsymbol{y}(s) \\ &= \underbrace{\partial_s w(\boldsymbol{y}(s), s) + \nabla w(\boldsymbol{y}(s), s) \cdot F(\boldsymbol{y}(s), \boldsymbol{\alpha}(s)) + h(\boldsymbol{y}(s), \boldsymbol{\alpha}(s))}_{(*)} - h(\boldsymbol{y}(s), \boldsymbol{\alpha}(s)).\end{aligned}$$

Now for an arbitrary control, the expression $(*)$ is smaller or equal to zero. This is because, by the HJB equation that w satisfies, when we take the maximum over all possible values of $\boldsymbol{\alpha}(s)$ we get $(*) = 0$, so any other $\boldsymbol{\alpha}$ will give less. We can then integrate the last equation from t to T and get

$$w(\boldsymbol{y}(T), T) - w(\boldsymbol{y}(t), t) \leqslant -\int_t^T h(\boldsymbol{y}(s), \boldsymbol{\alpha}(s))\, \mathrm{d}s,$$

and using $w(\boldsymbol{y}, T) = g(\boldsymbol{y})$ as well as $\boldsymbol{y}(t) = \boldsymbol{x}$, we find

$$g(\boldsymbol{y}(T)) + \int_t^T h(\boldsymbol{y}(s), \boldsymbol{\alpha}(s))\, \mathrm{d}s \leqslant w(\boldsymbol{x}, t).$$

So any arbitrary control can make the value function at most as large as $w(\boldsymbol{x}, t)$, and maximizing over the possible controls we find $u(\boldsymbol{x}, t) \leqslant w(\boldsymbol{x}, t)$. What remains to show is that $w(\boldsymbol{x}, t)$ is indeed a value function; so far we only know it to be the solution of a HJB equation. To see that w is indeed a value function, notice that when we choose the feedback control that we get from w (see the discussion before the theorem), then $(*)$ above is identically zero (independent of whatever $\boldsymbol{y}(s)$ happens to be). Thus, again by integrating, we see that w is indeed a value function. □