4.4. **Stochastic optimal control.** We will now perturb the equation for the state $\boldsymbol{y}_t$ by noise, leading to the stochastic differential equation

$$\mathrm{d}\boldsymbol{y}_s = f(\boldsymbol{y}_s, \boldsymbol{\alpha}_s)\,\mathrm{d}s + \sigma(\boldsymbol{y}_s, \boldsymbol{\alpha}_s)\,\mathrm{d}\mathcal{W}_s, \tag{4.11}$$

where $\mathcal{W}_s$ is $\mathbb{R}^n$-valued Brownian motion. The control problem is to maximize the *expectation of the various utility functions*, giving the optimal value function

$$u(x,t) = \max_{\boldsymbol{\alpha}} \mathbb{E}_{\boldsymbol{y}(t)=\boldsymbol{x}}\Big(\int_t^T h(\boldsymbol{y}_s, \boldsymbol{\alpha}_s)\,\mathrm{d}s + g(\boldsymbol{y}_T)\Big). \tag{4.12}$$

g and h are utility functions. The space of allowed controls is now such that for some $A \subset \mathbb{R}^m$, we need to have $\boldsymbol{\alpha}_s \in A$ for all s, and an additional condition is that $\boldsymbol{\alpha}_s$ is *adapted* to the Brownian motion; that is, $\boldsymbol{\alpha}_s$ depends only on the values $\{\mathcal{W}_r : r \leqslant s\}$. This condition is very natural, as it means that the controller cannot know the future of the (random) evolution modelled by the Brownian motion. Usually (and also in this lecture) it is enough to let the control $\boldsymbol{\alpha}_s$ depend only on s and $\boldsymbol{y}_s$, i.e. to consider a *feedback control.*
To find the HJB equation in this case, we proceed as in the deterministic case and work backwards from the final time T. Clearly,

$$u(\boldsymbol{x}, T) = \mathbb{E}_{\boldsymbol{y}_T=\boldsymbol{x}}(0 + g(\boldsymbol{y}_T)) = g(\boldsymbol{x}).$$

Assume now that we have found $u(\boldsymbol{x}, t+\delta t)$ for some small δt. Then, by the dynamic programming principle (which applies also to this case, as one can easily see),

$$\begin{aligned} u(\boldsymbol{x},t) &= \max_{\boldsymbol{\alpha}} \mathbb{E}_{\boldsymbol{y}_t=\boldsymbol{x}}\Big(\int_t^{t+\delta t} h(\boldsymbol{y}_s, \boldsymbol{\alpha}_s)\,\mathrm{d}s + u(\boldsymbol{y}_{t+\delta t}, t+\delta t)\Big) \\ &\approx \max_{\boldsymbol{\alpha}} \big(h(\boldsymbol{x},\boldsymbol{\alpha})\delta t + \mathbb{E}_{\boldsymbol{y}_t=\boldsymbol{x}}\big(u(\boldsymbol{y}_{t+\delta t}, t+\delta t)\big)\big)\,. \end{aligned}$$

The approximate identity in the last line is justified by the fact that $s \mapsto \mathbb{E}_{\boldsymbol{y}_t=\boldsymbol{x}}(h(\boldsymbol{y}_s, \boldsymbol{\alpha}_s))$ is continuous, and as δt is very small, the integral is approximately given by the initial value of the integrand times the length of the integration interval. We reformulate this to read

$$0 = \max_{\boldsymbol{\alpha}} \Big(h(\boldsymbol{x},\boldsymbol{\alpha})\,\delta t + \mathbb{E}_{\boldsymbol{y}_t=\boldsymbol{x}}\big(u(\boldsymbol{y}_{t+\delta t}, t+\delta t) - u(\boldsymbol{y}_t, t)\big)\Big). \tag{4.13}$$

Using the same trick that we have applied many times in the first few weeks, we find

$$\begin{aligned} &\mathbb{E}_{\boldsymbol{y}_t=\boldsymbol{x}}\big(u(\boldsymbol{y}_{t+\delta t}, t+\delta t) - u(\boldsymbol{y}_t, t)\big) = \mathbb{E}_{\boldsymbol{y}_t=\boldsymbol{x}}\Big(\int_t^{t+\delta t} \mathrm{d}u(\boldsymbol{y}_s, s)\Big) \\ &= \mathbb{E}_{\boldsymbol{y}_t=\boldsymbol{x}}\Big(\int_t^{t+\delta t} \big(\partial_t u(\boldsymbol{y}_s, s) + \nabla u(\boldsymbol{y}_s, s)\cdot f(\boldsymbol{y}_s, \boldsymbol{\alpha}_s) + \frac{1}{2}\sigma(\boldsymbol{y}_s, \boldsymbol{\alpha}_s)^2 \Delta u(\boldsymbol{y}_s, s)\big)\,\mathrm{d}s\Big) \\ &\approx \Big(\partial_t u(\boldsymbol{x}, s) + \nabla u(\boldsymbol{x}, t)\cdot f(\boldsymbol{x}, \boldsymbol{\alpha}_t) + \frac{1}{2}\sigma(\boldsymbol{x}, \boldsymbol{\alpha}_t)^2 \Delta u(\boldsymbol{x}, t)\Big)\delta t. \end{aligned}$$

The equality between first and second line above follows from the Itô formula and the fact that the expectation of a stochastic integral is zero (it is here that we need $\boldsymbol{\alpha}_s$ to be adapted!). The approximate identity between the second and third line follows in the same way as the one we have just discussed. We now use this in (4.13), and after letting $\delta t \to 0$ we get the following

Theorem: *$u(\boldsymbol{x}, t)$ from (4.12) is the solution of the Hamilton-Jacobi-Bellman equation*

$$\partial_t u(\boldsymbol{x}, t) + \max_{\boldsymbol{\alpha} \in A} \Big(f(\boldsymbol{x}, \boldsymbol{\alpha}) \cdot \nabla u(\boldsymbol{x}, t) + h(\boldsymbol{x}, \boldsymbol{\alpha}) + \frac{1}{2}\sigma^2(\boldsymbol{x}, \boldsymbol{\alpha}) \nabla u(\boldsymbol{x}, t) \Big) = 0, \tag{4.14}$$

with final condition $u(\boldsymbol{x}, T) = g(\boldsymbol{x})$.

Note that if σ does not depend on $\boldsymbol{\alpha}$, then (4.14) becomes

$$\partial_t u + H(\nabla u, \boldsymbol{x}) + \frac{1}{2}\sigma^2 \Delta u = 0,$$

with $H(\boldsymbol{p}, \boldsymbol{x}) = \max_{\boldsymbol{\alpha}}(f(\boldsymbol{x}, \boldsymbol{\alpha}) \cdot \boldsymbol{p} + h(\boldsymbol{x}, \boldsymbol{\alpha}))$ the same as in the deterministic case! As in the deterministic case, the derivation above was not fully rigorous, but once we have the result, we can give a rigorous proof.

Proof of the Theorem. We first show that if v solves the HJB equation (4.14), then for any adapted (not prescient) control $\boldsymbol{\alpha}_s$ we have

$$v(\boldsymbol{x}, t) \geqslant \mathbb{E}_{\boldsymbol{y}_t = \boldsymbol{x}} \Big(\int_t^T h(\boldsymbol{y}_s, \boldsymbol{\alpha}_s) \, \mathrm{d}s + g(\boldsymbol{y}_T) \Big).$$

The proof is similar to the deterministic case: we consider the path $\boldsymbol{y}_s$ resulting from the stochastic differential equation controlled by our chosen control $\boldsymbol{\alpha}_s$, and plug this into the solution v of the HJB equation. The Itô formula then gives

$$\begin{aligned} \mathrm{d}v(\boldsymbol{y}_s, s) &= \partial_s v(\boldsymbol{y}_s, s) \, \mathrm{d}s + \nabla_{\boldsymbol{y}} v(\boldsymbol{y}_s, s) \cdot \mathrm{d}\boldsymbol{y}_s + \frac{1}{2} \Delta v(\boldsymbol{y}_s, s) (\mathrm{d}\boldsymbol{y}_s)^2 = \\ &= \partial_s v(\boldsymbol{y}_s, s) \mathrm{d}s + \nabla_{\boldsymbol{y}} v(\boldsymbol{y}_s, s) \cdot \Big(f(\boldsymbol{y}_s, \boldsymbol{\alpha}_s) \, \mathrm{d}s + \sigma(\boldsymbol{y}_s, \boldsymbol{\alpha}_s) \mathrm{d}\mathcal{W}_s \Big) + \frac{1}{2}\sigma(\boldsymbol{y}_s, \boldsymbol{\alpha}_s)^2 \Delta v(\boldsymbol{y}_s, s) \, \mathrm{d}s. \end{aligned}$$

Since $\mathbb{E}_{\boldsymbol{y}_t = \boldsymbol{x}}(v(\boldsymbol{y}_T, T)) = \mathbb{E}_{\boldsymbol{y}_t = \boldsymbol{x}}(g(\boldsymbol{y}_T))$ by the final condition of the HJB equation, and since $\mathbb{E}_{\boldsymbol{y}_t = \boldsymbol{x}}(v(\boldsymbol{y}_t, t)) = v(\boldsymbol{x}, t)$, we have

$$\begin{aligned} &\mathbb{E}_{\boldsymbol{y}_t = \boldsymbol{x}}(g(\boldsymbol{y}_T)) - v(\boldsymbol{x}, t) = \mathbb{E}_{\boldsymbol{y}_t = \boldsymbol{x}}(v(\boldsymbol{y}_T, T)) - \mathbb{E}_{\boldsymbol{y}_t = \boldsymbol{x}}(v(\boldsymbol{y}_t, t)) = \mathbb{E}_{\boldsymbol{y}_t = \boldsymbol{x}} \Big(\int_t^T \mathrm{d}v(\boldsymbol{y}_s, s) \Big) = \\ &= \mathbb{E}_{\boldsymbol{y}_t = \boldsymbol{x}} \Big(\int_t^T \big(\partial_s v(\boldsymbol{y}_s, s) + \nabla_{\boldsymbol{y}} v(\boldsymbol{y}_s, s) \cdot f(\boldsymbol{y}_s, \boldsymbol{\alpha}_s) + \frac{1}{2}\sigma(\boldsymbol{y}_s, \boldsymbol{\alpha}_s)^2 \Delta v(\boldsymbol{y}_s, s) + h(\boldsymbol{y}_s, \boldsymbol{\alpha}_s) - h(\boldsymbol{y}_s, \boldsymbol{\alpha}_s) \big) \, \mathrm{d}s \Big) \\ &\leqslant - \mathbb{E}_{\boldsymbol{y}_t = \boldsymbol{x}} \Big(\int_t^T h(\boldsymbol{y}_s, \boldsymbol{\alpha}_s) \, \mathrm{d}s \Big). \end{aligned}$$

The last inequality follows, as in the deterministic case, from the fact that v solves the HJB equation, i.e. the maximum over all controls of all the terms under the integral except the last one is zero. So we find that

$$v(\boldsymbol{x}, t) \geqslant \boldsymbol{E}_{\boldsymbol{y}_t = \boldsymbol{x}} \Big(\int_t^T h(\boldsymbol{y}_s, \boldsymbol{\alpha}_s) \, \mathrm{d}s + g(\boldsymbol{y}_T) \Big),$$

and maximizing over all controls gives that v is at least as large as the optimal value function. Now again, we can see that (for each path of the Brownian motion), the feedback control obtained from the HJB equation leads to the value function $v(\boldsymbol{x}, t)$, so that $v(\boldsymbol{x}, t)$ is not only a solution to the HJB equation, but indeed also a value function. Therefore it must be the optimal value function. □

4.5. **Application: Optimal portfolio selection and consumption.** This is a problem considered by Robert Merton (J. Econ. Theory 3, (1971) 373-413). Here is the setup:

b_s is a riskless asset with $\mathrm{d}b_s = rb_s\,\mathrm{d}s$, so $b_s = b_0\,\mathrm{e}^{rs}$.
p_s is a risky asset solving $\mathrm{d}p_s = \mu p_s \mathrm{d}s + \sigma p_s \mathrm{d}\mathcal{W}_s$.
x is our wealth at the starting time t.
The control parameters are $\alpha_1(s)$, the fraction of wealth in the risky asset p_s at time s; clearly, we need $0 \leqslant \alpha_1 \leqslant 1$.
$\alpha_2(s)$ is our rate of consumption at time s. We want $\alpha_2 \geqslant 0$.
The equation for the total wealth controlled by α_1 and α_2 is then

$$\mathrm{d}y_s = (1-\alpha_1(s))y_s r\mathrm{d}s + \alpha_1(s)y_s(\mu\,\mathrm{d}s + \sigma\,\mathrm{d}\mathcal{W}_s) - \alpha_2(s)\,\mathrm{d}s. \tag{4.15}$$

We impose the state constraint $y_s \geqslant 0$. The most elegant way to do this is to define $\tau(x) = \inf\{s \geqslant t : y_s = 0\}$; we then have the optimal value function (with discounting) given by

$$u(x.t) = \max_{\alpha_1,\alpha_2} \mathbb{E}_{y_t = x}\Big(\int_t^{\min(T,\tau(x))} \mathrm{e}^{-\rho s}\, h(\alpha_2(s))\,\mathrm{d}s \Big).$$

We want the utility function h to be monotone increasing and concave, as in the deterministic case. Our eventual choice will be $h(\alpha) = \alpha^\gamma$ with $0 < \gamma < 1$.
The derivation of the HJB equation with discounting is entirely parallel to the general case that we just treated. The result is

$$\partial_t u + \max_{\alpha_1,\alpha_2} \Big(\mathrm{e}^{-\rho t}\, h(\alpha_s) + (xr + \alpha_1(\mu - r)x - \alpha_2)\partial_x u + \frac{1}{2}x^2\sigma^2\alpha_1^2\partial_x^2 u \Big) = 0. \tag{4.16}$$

We can find the optimal α_1 by simple differentiation: the determining equation is

$$x(\mu - r)\partial_x u + \sigma^2 x^2 \partial_x^2 u \alpha_1 = 0,$$

giving for the optimal α_1:

$$\alpha_1^* = -\frac{(\mu - r)\partial_x u}{\sigma^2 x \partial x^2 u}. \tag{4.17}$$

Have we actually found a maximum? Only if $\partial_x^2 u > 0$, otherwise it would be a minimum! We need to keep the in mind and check at the end that it holds. Also, we have so far ignored the constraint $0 \leqslant \alpha_1 \leqslant 1$. We will have to come back to this later, too.
The optimal α_2 is now determined by the equation

$$\mathrm{e}^{-\rho t}\, h'(\alpha_2) = \partial_x u(x,t), \tag{4.18}$$

where h' is the derivative of h. It is intuitively clear that $\partial_x u > 0$, as greater initial wealth will give greater optimal value (try to find a mathematical argument for this!). Also, $h' > 0$ by the assumption that h is monotone increasing, $h'' < 0$ by concavity. So the optimal $\alpha_2^* > 0$ is nonnegative.
We now specialize to the case $h(\alpha) = \alpha^\gamma$ to make further progress. In the same way as on last week's problem sheet, we see that u must be of the form $u(x,t) = g(t)x^\gamma$. Since the optimal value is certainly nonnegative, we will have $g(t) \geqslant 0$. Thus $\partial_x u = \gamma g(t) x^{\gamma-1}$, and $\partial_x^2 u = \gamma(\gamma-1)g(t)x^{\gamma-2}$. Note that this means $\partial_x^2 u < 0$, which was one of the conditions that we had to remember checking.
Now (4.17) becomes

$$\alpha_1^* = \frac{\mu - r}{\sigma^2(1-\gamma)},$$

and (4.18) reads

$$\alpha_2^* = (\,\mathrm{e}^{\rho t}\, g(t))^{1/(\gamma-1)} x.$$

We can see that $0 \leqslant \alpha_1^* \leqslant 1$ if

$$0 \leqslant \mu - r \leqslant \sigma^2(1-\gamma). \tag{4.19}$$

We will assume for the moment that this extra condition holds. Putting $u(x,t) = g(t)x^\gamma$ back into the equation, we find that g needs to satisfy

$$\partial_t g(t) + \nu\gamma g(t) + (1-\gamma) g(t) \Big(\,\mathrm{e}^{\rho t}\, g(t)\Big)^{1/(\gamma-1)} = 0,$$

with final condition $g(T) = 0$, and $\nu = r + \frac{(\mu-r)^2}{2\sigma^2(1-\gamma)}$. This is of the same form as the equation that we have seen in the deterministic optimal consumption problem, and by following the steps given in Problem 1 on Sheet 5, we find that the solution is

$$g(t) = \,\mathrm{e}^{-\rho t}\, \Big(\frac{1-\gamma}{\rho - \nu\gamma}\Big(1 - \,\mathrm{e}^{-\frac{(\rho-\nu\gamma)(T-t)}{1-\gamma}}\,\Big)\Big)^{1-\gamma}.$$

So, the optimal value function is $g(t)x^\gamma$ with the above $g(t)$. The optimal control α_1^* is constant, i.e. it depends neither on time nor on the current wealth. This means that our investment decision is not influenced by our current wealth, and also not by the time we still have to consume. Instead, it is fully determined by the difference $\mu - r$ of the expected return of the asset and the bond rate, divided by a factor $\sigma^2(1-\gamma)$. This factor σ^2 is easy to interpret: large uncertainty σ makes it more unattractive to invest in the risky asset. The factor $(1-\gamma)$ is less obvious. It means that if we can consume larger amounts of wealth with relatively little penalty (i.e. γ close to 1), then we should invest less into the risky asset. Now, all of this is for $\mu > r$. In the case $\mu < r$, our extra condition (4.19) does not hold; it is not difficult to see that in this case, $\alpha_1^* = 0$ is the optimal allowed control. So, if the expected rate of return for the risky asset is less than the bond rate, it is not worth investing into it at all. On the other hand, if $\mu - r > \sigma^2(1-\gamma)$, then the return of the risky asset is so much better that the bond rate, that we will put all our wealth into it, and $\alpha_1^* = 1$ in that case.

Unlike α_1^*, the optimal α_2^* in (4.18) does depend on time. Now that we know $g(t)$, we put it into (4.18) (this is part of the feedback!) and find

$$\alpha_2^*(x,t) = \frac{1-\gamma}{\rho-\nu\gamma}\Big(1 - \mathrm{e}^{-\frac{(\rho-\nu\gamma)(T-t)}{1-\gamma}}\Big)x.$$

At given time s, our wealth will also be known to be y_s. So at that time, we replace x with y_s in the above equation. This is the second part of the feedback. This means that the optimally controlled asset solves the SDE

$$\mathrm{d}y_s = (r + (\mu - r)\alpha_1^*)y_s r\mathrm{d}s - \frac{1-\gamma}{\rho-\nu\gamma}\Big(1 - \mathrm{e}^{-\frac{(\rho-\nu\gamma)(T-s)}{1-\gamma}}\Big)y_s + \alpha_1^*\sigma y_s\,\mathrm{d}\mathcal{W}_s.$$

The fact that $\alpha_2^*(y_s, s)$ is proportional to y_s guarantees that the whole right hand side of the SDE is proportional to y_s. This means that $y_s \geqslant 0$ (since, should it ever hit zero (from above, obviously), its time derivative will be zero, and it will be stuck there). Thus luckily the state constraint is automatically fulfilled, and indeed we do not need $\tau(x)$ in the end. The optimal consumption rate is easy as a function of wealth (proportional to it), but rather difficult as a function of time. It seems strange that it goes to zero as $t \to T$. One would have thought it should go to infinity then, as there is nothing to lose by consuming it all in the last instant. I don't fully understand this intuitively. One explanation is that we are optimizing the *expected* utility, and so the optimal consumption is made so that in the last instant, on average there won't be much left to consume. But this is not fully clear to me.