FIN 539 MATHEMATICAL FINANCE
Lecture 2: Dynamic Programming Approach

Philip H. Dybvig
Washington University in Saint Louis

Copyright © Philip H. Dybvig 2021

Dynamic Programming Approach

The idea of dynamic programming is to reduce the optimization into a series of single-period optimization problems (or optimization problems at a point of time in a continuous-time model). Optimal decisions for the future are given, and are encoded in the "continuation value" as a function of what happens this period. The continuation value (or the "value function") is a function of the "state variables" that are important for determining what happens going forward. State variables could include such variables as wealth, time, and current interest rate or mean return on the stock. We use the first-order conditions from the single-period problem to solve for the value function and strategy as a function of the state variables.

The art in dynamic programming is related to the state variables. Solving the problem is at least easier (and perhaps only possible) if we have the correct state variables. Also, in general, the more state variables there are, the harder the problem is to solve, so it usually makes sense to make some strong assumptions to keep the number of state variables low.

A discrete-time dynamic portfolio problem

Given w_0 at time 0,
choose adapted investment θ_s, consumption $c_s \geq 0$, and wealth w_s to maximize $\mathrm{E}[\Sigma_{s=0}^{T-1} \frac{1}{(1+\rho)^s} u(c_s) + b(w_T)]$
subject to
$(\forall s) w_{s+1} = (w_s - c_t)(1 + r_f) + \theta_s(r_{s+1} - r_f)$ (budget constraint)
and $(\forall s) w_s \geq 0$ (nonnegative wealth).

w_s: wealth at time s before consumption
c_s: consumption at time s
w_{T+1}: terminal wealth or bequest
$\Sigma_{s=0}^{T} \frac{1}{(1+\rho)^s} u(c_t) + b(w_{T+1})$ time-separable fon Neumann-Morgenstern utility function
$u(c_s)$: felicity function (also called utility function)
ρ: pure rate of time discount
$b(w_T)$: utility of the bequest
r_f: riskfree rate
r_s: random rate of return on the risky asset from $s-1$ to s

Possible variations

In the previous page, if there is no preference for consumption over time ($u(c) \equiv 0$ and the $c_s \equiv 0$'s are not choice variables), this is a *terminal horizon* problem, which could be a useful model for saving for retirement or a nuclear decommissioning trust. If we include consumption c_s over time, with or without the bequest, this is a *consumption withdrawal* problem. If $T = \infty$, we do not have the term $b(w_T)$ and we call this an *infinite-horizon* model. Infinite-horizon problems can be easier to solve than finite horizon problems because the optimal portfolio does not depend on how much time is left.

If we have cash flows that must be met over time, these liabilities are included in the budget constraint and we would call this an asset-liability management (ALM) problem. ALM problems are common in defined-benefit retirement plans, insurance, and should be important for university endowments and other settings. We could also account for inflows over time, e.g. from salary and wages, within the budget constraint.

Dynamic programming: value function $V(w_t, t)$

In dynamic programming, the *value function* gives the value of continuing as a function of how things stand at a point in time. For our discrete-time problem, the value function $V(w_t, t)$ is the value of the objective function for the optimal solution of the continuation problem:

Given w_t at time t,
choose adapted investment θ_s, consumption $c_s \geq 0$, and wealth w_s to maximize $\mathrm{E}[\Sigma_{s=0}^{T-t-1} \frac{1}{(1+\rho)^s} u(c_s) + b(w_T)]$
subject to
$(\forall s) w_{s+1} = (w_s - c_s)(1 + r_f) + \theta_s(r_{s+1} - r_f)$ (budget constraint)
and $(\forall s) w_s \geq 0$ (nonnegative wealth).

We could take as given the whole history, but instead we condition on *state variables* w_t and t that matter going forward. If we include too many variables, maybe it is possible to prove that the unneeded ones don't matter, but it probably requires more work. If $T = \infty$, $V(w_t, t) = V(w_t)$ and t is not a state variable since the continuation problem is the same at all t and depends only on the starting wealth.

Dynamic programming equation (Bellman equation[1])

Assume returns are i.i.d. in the problem two slides previous, and that $V(w_t,t)$ is the value function before consuming at t. Then the Bellman equation is:

$V(w_t,t) = \max_{c_t,\theta_t}\left\{u(c_t) + \frac{E[V((w_t-c_t)(1+r_f)+\theta_t(r_{t+1}-r_f),t+1)]}{1+\rho}\right\}$

First-order conditions:

$u'(c_t^*) = \frac{E[-(1+r_f)V_w'((w_t-c_t^*)(1+r_f)+\theta_t^*(r_{t+1}-r_f),t+1)]}{1+\rho}$
$E[(r_{t+1}-r_f)V_w'((w_t-c_t^*)(1+r_f)+\theta_t^*(r_{t+1}-r_f),t+1)] = 0$

$$V(w_t,t) = u(c_t^*) + \frac{E[V((w_t-c_t^*)(1+r_f)+\theta_t^*(r_{t+1}-r_f),t+1)]}{1+\rho}$$

Hopefully, this seems natural or at least plausible that we can look at the decision now in isolation given the value of continuing in different contingencies. We will derive this using martingale tools for the continuous case, and almost the same derivation works for this case.

[1]Some people tack on other names, for example, Hamilton-Jacoby-Bellman (HJB) equation.

Continuous-time dynamic portfolio problem (single risky asset)

Given w_0 and initial time 0,
choose adapted risky investment θ_s, consumption $c_s \geq 0$, and wealth w_s to
maximize $\mathrm{E}[\int_{s=0}^{T} e^{-\rho s} u(c_s) ds + e^{-\rho T} b(w_T)]$
subject to:
$(\forall s)(dw_s = rw_s ds + \theta_s((\mu - r)ds + \sigma dZ_s) - c_s ds)$
$(\forall s)(w_s \geq 0)$

In this continuous-time problem, w_0 is initial wealth, θ_s is the portfolio weight in wealth units, c_s is the consumption process, $u(\cdot)$ is the felicity function (also called utility function), $b(w_T)$ is the contribution to utility of the bequest w_T, ρ is the pure rate of time discount, r is the riskfree rate, μ is the mean return on the risky asset, dZ_s is the underlying noise in the risky asset, and σ^2 is the local variance of the risky asset. If $T = \infty$ (an infinite-horizon problem), we exclude the bequest term $b(\cdot)$, and for a terminal horizon problem we exclude consumption choice variables (c_s's) and the integral with the utility $u(c_s)$.

Value function $V(w_t, t)$

For our continuous-time example, the value function $V(w_t, t)$ is the value of the objective function at the optimal solution of the problem

Given w at time t,
choose adapted risky investment quantities θ_s, consumption $c_s \geq 0$, and wealth w_s to
maximize $\mathrm{E}[\int_{s=0}^{T-t} e^{-\rho s} u(c_s) ds + e^{-\rho(T-t)} b(w_{T-t})]$
subject to:
$(\forall s)(dw_s = rw_s ds + \theta_s((\mu - r)ds + \sigma dZ_s) - c_s ds)$
$w_0 = w$ and $(\forall s)(w_s \geq 0)$

The state variables w_t and t summarize everything we need to know about the past to solve the optimization problem going forward. We could also consider a richer choice problem with more state variables (e.g. the current interest rate, current stock return volatility, or estimated future liabilities in an ALM setting). Adding state variables may make the optimization more realistic, but it also can make it harder to estimate the parameters and solve.

Value process M_t: martingale given the optimal strategy

We will follow Fleming and Richel and use the "martingale approach" to deriving the Bellman equation. Recall that a stochastic process M_t is called a *martingale* if it doesn't change on average, so for $s < t$, $E_s[M_t] = M_s$. Conditional expectations are martingales by the law of iterated expectations. Let $M_t \equiv E_t[X]$ for some random variable X. Then $E_s[M_t] = E_s[E_t[X]] = E_s[X] = M_s$.

Now, the value of the objective for our continuous-time is the expectation of an integral. This motivates defining the process

$$M_t \equiv \int_{s=0}^{t} e^{-\rho s} u(c_s) ds + e^{-\rho t} V(w_t, t).$$

If we are following the optimal strategy, this is the conditional expectation of the realized value, given information at time t. The integral in the definition is the part of the integral from what has already happened. The definition of $V(w_t, t)$ tells us that the final term is the value of what is to come if we do what is optimal from now on (which we will since we are assuming the optimal strategy).

Value process M_t: supermartingale given any strategy

Recall that a stochastic process M_t is called a *supermartingale* if it never increases on average, so for $s < t$, $E_s[M_t] \leq M_s$. If we are following an arbitrary strategy, M_t defined in the previous slide may be a supermartingale and is only a martingale if the strategy is optimal. Note that a martingale is a supermartingale; we can call a supermartingale that is not a martingale a *strict supermartingale*.

For an arbitrary strategy, we can define M_t as the expected value given information at time t of following the arbitrary strategy up until time t and then switching to the optimal strategy from t onwards. Given the optimal strategy, changes in M_t only reflect good or bad luck and on average are zero. However, for a sub-optimal strategy, as t increases, changes in M_t reflect both good or bad luck and the impact of following a sub-optimal strategy for a longer time. Therefore, for $s < t$, the decline $E[M_s] - E[M_t]$ is the loss in utility terms of irreversible mistakes made between times s and t.

Value process M_t: to the Bellman equation

M_t will be an Itô process ($dM_t = a_t dt + b_t dZ_t$ for some random processes a_t and b_t). Then, if M_t is a martingale, the drift $a_t = 0$. If it is a supermartingale, the drift $a_t \leq 0$. Therefore, the optimal strategy maximizes the drift, and the maximized drift is zero. Therefore, we have

$$\max_{\theta_t, c_t} \text{drift}(M_t) = 0,$$

By Itô's lemma and the formula for dw_t from the constraint,

$$\begin{aligned} dM &= e^{-\rho t}(u(c)dt + (V_t - \rho V(w,t))dt + V_w(w,t)(rwdt \\ &\qquad + \theta((\mu - r)dt + \sigma dZ) - cdt) + \frac{\theta^2\sigma^2}{2} V_{ww} dt), \end{aligned}$$

and therefore we have the Bellman equation

$$\max_{\theta, c} \left(u(c) + V_t - \rho V + (rw + \theta(\mu - r) - c)V_w + \frac{\theta^2\sigma^2}{2} V_{ww} \right) = 0.$$

Mini math review: multidimensional Itô's lemma

Let $H : \Re^d \times [0, T] \to \Re$ with continuous partial derivatives $H_x(x, t)$, $H_{xx}(x, t)$, and $H_t(x, t)$. Let $dX_t = g(t)dt + G(t)dZ_t$, where X_t is a d-dimensional process and Z_t is an m-dimentional standard Wiener process. Then $Y_t \equiv H(X_t, t)$ is an Itô process with stochastic differential

$$dY_t = H_t dt + H_x dX + \frac{1}{2}\mathrm{tr}(GG' H_{xx})dt$$

where, for any symmetric matrix A, $\mathrm{tr}(A)$ denotes the trace, which is the sum $\Sigma_i A_{ii}$ of its diagonal elements.

Note: if H takes values in $\Re^k$, we can apply the result elementwise.

Towards a solution: optimal c and θ

Taking the first-order conditions for the maximization in the Bellman equation with respect to c and θ, we have that

$$u'(c) = V_w$$

and

$$(\mu - r)V_w + \frac{2\sigma^2\theta}{2}V_{ww} = 0.$$

Therefore, the optimal choices are

$$c^* = I(V_w)$$

where $I(\cdot)$ is the inverse of the marginal utility function, and

$$\theta^* = -\frac{\mu - r}{\sigma^2}\frac{V_w}{V_{ww}}.$$

Bellman equation with optimized values

Recall that the Bellman equation is

$$\max_{\theta,c}\left(u(c)+V_t-\rho V+(rw+\theta(\mu-r)-c)V_w+\frac{\theta^2\sigma^2}{2}V_{ww}\right)=0.$$

Substituting in the optimal consumption c^* and optimal portfolio θ^*, we have

$$u(I(V_w))+V_t-\rho V+(rw-I(V_w))V_w-\frac{(\mu-r)^2}{2\sigma^2}\frac{(V_w)^2}{V_{ww}}=0.$$

Defining the dual function $\tilde{u}(z)\equiv\max_c u(c)-zc=u(I(z))-zI(z)$,

$$\tilde{u}(V_w)+V_t-\rho V+rwV_w-\frac{(\mu-r)^2}{2\sigma^2}\frac{(V_w)^2}{V_{ww}}=0.$$

We can solve this subject to a boundary conditions at maturity and for large and small wealth. For example, if T is finite, $V(w_T,T)=b(w_T)$.

Exploiting homotheticity

Suppose $u(c) = K_0 \log(c)$ and $b(w) = K_1 \log(w)$ where K_0 and K_1 are nonnegative and not both zero. Then the value function is the value of:

Given w_t and initial time t,
choose adapted investment θ_s, consumption c_s and wealth c_s to maximize $\mathrm{E}[\int_{s=0}^{T-t} e^{-\rho s} K_0 \log(c_s) ds + e^{-\rho(T-t)} K_1 \log(w_T)]$
subject to:
$(\forall s)(dw_s = rw_s ds + \theta_s((\mu - r)ds + \sigma dZ_s) - c_s ds)$
$(\forall s)(w_s \geq 0)$

Now letting $\hat{c}_s \equiv c_s/w_t$, $\hat{w}_s \equiv w_s/w_t$, and $\hat{\theta}_s \equiv \theta_s/w_t$, this becomes

Given w_t and initial time t and $\hat{w}_t \equiv 1$,
choose adapted investment $\hat{\theta}_s$ consumption $\hat{c}_s$, and wealth $\hat{w}_s$ to maximize $\mathrm{E}[\int_{s=0}^{T-t} e^{-\rho s} K_0 \log(w_t \hat{c}_s) ds + e^{-\rho(T-t)} K_1 \log(w_t \hat{w}_T)]$
subject to:
$(\forall s)(d\hat{w}_s = r\hat{w}_s ds + \hat{\theta}_s((\mu - r)ds + \sigma dZ_s) - \hat{c}_s ds)$
$(\forall s)(\hat{w}_s \geq 0)$

Exploiting homotheticity...continued

Now $\log(w_t\hat{c}_s) = \log(w_t) + \hat{c}_s$ and $\log(w_t\hat{w}_T) = \log(w_t) + \log(\hat{w}_T)$. Therefore, the objective function of the second problem can be rewritten as

$$\begin{aligned}
&\textstyle\int_{s=0}^{T-t} e^{-\rho s} K_0 \log(w_t) ds + e^{-\rho(T-t)} K_1 \log(w_t) \\
&\qquad + \mathrm{E}[\textstyle\int_{s=0}^{T-t} e^{-\rho s} K_0 \log(\hat{c}_s) ds + e^{-\rho(T-t)} K_1 \log(\hat{w}_T)] \\
&\ = \ H(T-t)\log(w_t) + \mathrm{E}[\textstyle\int_{s=0}^{T-t} e^{-\rho s} K_0 \log(\hat{c}_s) ds + e^{-\rho(T-t)} K_1 \log(\hat{w}_T)],
\end{aligned}$$

where $H(T-t) \equiv K_0(1 - e^{-\rho(T-t)})/\rho + K_1 e^{\rho(T-t)}$. Since w_t appears only in the leading constant term in the objective (and not in the constraints), the optimal choice of $\hat{c}_s$, $\hat{\theta}_s$, and $\hat{w}_s$ does not depend on w_t. Therefore the value of the problem can be written as $V(w_t, t) = H(T-t)\log(w_t) + v(t)$, where $v(t) = V(1, t)$. Now, the value function has a single state variable (argument), which will make our task easier going forward.

Optimal solution: log utility, infinite horizon

Consider an infinite horizon problem with consumption withdrawal and log utility $u(c) = \log(c)$. For this, $K_0 = 1$, $K_1 = 0$, and $T = \infty$. The value function does not depend on time since there is the same amount of time to the horizon for all t. Similarly, $H = 1/\rho$ and v are constants and do not depend on time. Then $V(w) = \log(w)/\rho + v$, and we have that $V_t = 0$, $V_w = 1/(\rho w)$, and $V_{ww} = -1/(\rho w^2)$. Furthermore, $\tilde{u}(z) \equiv \max_c u(c) - zc = -\log(z) - 1$. Therefore, the Bellman equation (with optimized values) is

$$\begin{aligned} 0 &= -\log(V_w) - 1 + V_t - \rho V + rwV_w - \frac{(\mu - r)^2}{2\sigma^2}\frac{(V_w)^2}{V_{ww}} \\ &= -\log\left(\frac{1}{\rho w}\right) - 1 + \frac{rw}{\rho w} - \rho(\frac{\log(w)}{\rho} + v) - \frac{(\mu - r)^2}{2\sigma^2}\frac{(1/(\rho w))^2}{-1/(\rho w^2)} \\ &= \log(\rho) - 1 + r/\rho - \rho v + (\mu - r)^2/(2\sigma^2\rho) \end{aligned}$$

$$v = \frac{\log(\rho)}{\rho} + \frac{r - \rho}{\rho^2} + \frac{(\mu - r)^2}{2\sigma^2\rho^2}$$

Log utility portfolio choice, consumption and portfolio choice

We can also compute the optimal consumption and portfolio choice in the infinite horizon case.

$$c^* = I(V_w) = \rho w$$

$$\theta^* = -\frac{\mu - r}{\sigma^2}\frac{V_w}{V_{ww}} = \frac{\mu - r}{\sigma^2} w$$

We actually do not need to know the value of v (or more generally the function $v(t)$ if $T < \infty$) to be able to compute c^* and θ^*. For log utility, the portfolio choice is the same fraction $(\mu - r)/\sigma^2$ of wealth independent of K_0, K_1, and T. Optimal consumption does depend on these parameters, but can be inferred from $V(w, t) = H(T - t)log(w_t) + v(t)$ without knowing the function $v(t)$. In particular, we have

$$c^* = I(V_w) = wK_0/H(T - t) = \frac{K_0 \rho w}{K_0(1 - e^{-\rho(T-t)}) + K_1 \rho e^{-\rho(T-t)}}$$