11. Optimal Execution for Portfolios

Our discussion of the Almgren and Chriss (2001) strategy for optimal execution only concerned a single asset at a time, but if part of the goal of optimal execution is to mitigate risk, then this risk-mitigation can only truly be handled at the portfolio level. For example, consider a long and short position of similar sizes in strongly-correlated assets. Suppose the short position were much more liquid. An execution algo which covered the short more quickly due to its higher liquidity would leave the portfolio net long for the remainder of the day.

It is our immediate goal, then, to discuss how to generalize the Almgren-Chriss trading path problem to portfolios. We will also attack the problem directly in continuous time, whereas our previous formulation was in discrete time. The discrete-time formulation is a bit awkward for this problem, because it leads to what is essentially a discretization of a second-order differential equation with boundary conditions. It is simpler to actually treat the differential equations.

Most of the arguments and derivations in this lecture are adapted from Benveniste and Ritter (2017) available at `https://ssrn.com/abstract=3057570`. Check out the link for more details.

11.1. Almgren-Chriss and Hamiltonian Dynamics.

Almgren and Chriss (2001) consider two kinds of impact: permanent and temporary, where temporary is really *instantaneous* as it is assumed to have no impact at all on the price process in any period other than the one in which the trade occurs. Furthermore, Almgren and Chriss (2001) consider only the linear form of permanent impact, and ultimately show that linear permanent impact plays no role in the continuous-time limit. Throughout this lecture, we follow Almgren and Chriss (2001) in treating all temporary impact as instantaneous, and treating permanent impact as linear, hence irrelevant to the continuous-time limit.

These impact assumptions are an approximation which is only valid within a certain regime. For example, if we repeatedly aggress with medium to large order sizes within a short timeframe, it is unrealistic to assume that the impact will revert instantly. We must never push ourselves, by result of our own actions, into a regime where we know that our model no longer holds.

Let $q(t)$ denote a hypothetical continuous-time trading path, so $q(t) \in \mathbb{R}^n$ denotes the portfolio positions (or holdings) at time t, denominated in dollars or any other convenient numeraire.

Much of the literature on market impact (see Gatheral (2010) for example) represents impact as a function of the *trading rate* or the time-derivative of the positions.

It is conventional in physics, and increasingly in finance as well, to denote the time-derivative of a path by a dot over the letter. Hence in our notation the trading rate (in dollars per unit time) is $\dot{q}(t)$, defined to be $dq(t)/dt$.

The *instantaneous trading cost function* is defined as the continuous-time limit of any ordinary trading cost function. In the single-security case, if we trade δq dollars in some small time interval of length δt, and this costs $\lambda \cdot \delta q/\delta t$ times traded notional for some $\lambda > 0$, then the total cost in dollars per unit time is

$$\lambda(\delta q/\delta t)^2 \equiv c(\delta q/\delta t)$$

where $c(v) = \lambda v^2$. However, we stress that $c()$ need not be quadratic, merely convex. In the multiple-security case, $c : \mathbb{R}^n \to \mathbb{R}_+$.

Let $\Sigma \in S^n_{++}$ denote the asset-level covariance matrix, which we assume has strictly positive spectrum. Practically, if n is large the estimation of Σ should be approached via a stable procedure such as (11.16) below.

Consider the problem of starting from a portfolio q_0, and executing a wave of orders over a fixed, pre-determined time interval $[0, T]$, leading to target portfolio q_{opt}. This can be viewed as a problem in the calculus of variations. As in classical mechanics, it has both a Lagrangian and a Hamiltonian formulation, which are convex duals to each other. The Hamiltonian is related to the Lagrangian by the Legendre-Fenchel transform.

The optimal-trading problem is then

$$\min_q \int_0^T L(q(t), \dot{q}(t))\, dt \quad \text{s.t.} \quad q(0) = q_0 \ \text{ and } \ q(T) = q_{\text{opt}} \tag{11.1}$$

where L is the *Lagrangian*

$$L(q, v) = c(v) + \frac{1}{2}\kappa(q - q_{\text{opt}})'\Sigma(q - q_{\text{opt}})\,. \tag{11.2}$$

where $\kappa > 0$ is the risk-aversion constant.

At a stationary point, the Euler-Lagrange equation is satisfied:

$$\frac{d}{dt}\frac{\delta L}{\delta \dot{q}} - \frac{\delta L}{\delta q} = 0.$$

A first-order system can be obtained if we introduce a new variable p, whose components are called *generalized momenta* and defined as

$$p := \frac{\delta L}{\delta v}(q, \dot{q}) \tag{11.3}$$

If the conditions of the implicit function theorem are satisfied, we could in principle algebraically solve (11.3) for $\dot{q}$, obtaining

$$\dot{q} = \phi(q, p)$$

for some function ϕ defined implicitly by (11.3). The Euler equation then takes the form

$$\dot{p} = \frac{\delta L}{\delta q}(q, \dot{q}) = \frac{\delta L}{\delta q}(q, \phi(q,p)) \equiv \psi(q,p)$$

As the functions ϕ, ψ are algebraic (not involving derivatives), we have a system of $2n$ first-order ODE given by

$$\dot{q} = \phi(q,p), \quad \dot{p} = \psi(q,p) \tag{11.4}$$

These equations can be expressed more symmetrically by introducing the *Hamiltonian*

$$H(q,p) = p\phi(q,p) - L(q, \phi(q,p))$$

Eqns. (11.4) are equivalently written in a form known as *Hamilton's equations*:

$$\dot{q} = \frac{\delta H}{\delta p}, \quad \dot{p} = -\frac{\delta H}{\delta q} \tag{11.5}$$

We can now solve the variational problem (11.1), as it is sufficient to solve (11.5) subject to the indicated boundary conditions. If we do this with the same cost function as Almgren and Chriss (2001), we must obtain the same solution.

Suppose $c(v) = \frac{1}{2}v'\Lambda v$ where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ is a diagonal matrix. We assume that no trading is free of cost, so $\lambda_i > 0$ for all i. From (11.2) and (11.3), we see that the generalized momenta are $p = \Lambda\dot{q}$ and hence algebraically solving, one has $\phi(q,p) = \Lambda^{-1}p$. The Hamiltonian is then

$$H(q,p) = \frac{1}{2}p\Lambda^{-1}p - \frac{1}{2}\kappa(q - q_{\text{opt}})'\Sigma(q - q_{\text{opt}})\,.$$

Hamilton's equations then become:

$$\dot{q} = \Lambda^{-1}p, \qquad \dot{p} = \kappa\Sigma(q - q_{\text{opt}}) \tag{11.6}$$

or more simply

$$\frac{dx}{dt} = Ax + b, \tag{11.7}$$

where $x : \mathbb{R} \to \mathbb{R}^{2n}$ and $A \in M(2n; \mathbb{R})$, $b \in \mathbb{R}^{2n}$ are given by

$$x = \begin{pmatrix} q \\ p \end{pmatrix}, \; A = \begin{pmatrix} 0 & \Lambda^{-1} \\ \kappa\Sigma & 0 \end{pmatrix}, \text{ and } \; b = \begin{pmatrix} 0 \\ -\kappa\Sigma q_{\text{opt}} \end{pmatrix}. \tag{11.8}$$

Duhamel's formula yields

$$x(t) = e^{tA}x_0 + \int_0^t e^{(t-s)A}b\,ds \tag{11.9}$$

To make use of (11.9) one must compute the one-parameter group e^{tA}. The usual way to do this involves the Jordan canonical form of A. In fact, the special structure of this problem implies that A is diagonalizable, as we show below.

Before proceeding to the multi-asset case, it is instructive to consider the single-asset case, in which case $\Lambda = (\lambda)$ and $\Sigma = (\sigma^2)$ are 1×1 matrices and

$$e^{tA} = \begin{pmatrix} \cosh(\omega t) & (\lambda\omega)^{-1}\sinh(\omega t) \\ \lambda\omega\sinh(\omega t) & \cosh(\omega t) \end{pmatrix}, \qquad \omega := \sqrt{\frac{\kappa\sigma^2}{\lambda}} \tag{11.10}$$

Applying (11.10) to (11.9), the full solutions to Hamilton's equations (11.6) in the single-asset case are then

$$\begin{aligned} q(t) &= q_{\text{opt}} + (q_0 - q_{\text{opt}})\cosh(t\omega) + (\lambda\omega)^{-1} p_0 \sinh(t\omega) \\ p(t) &= \lambda\omega\,(q_0 - q_{\text{opt}})\sinh(t\omega) + p_0\cosh(t\omega) \end{aligned} \tag{11.11}$$

The path given by (11.11) necessarily satisfies the initial condition $q(0) = q_0$, and imposing the final condition $q(T) = q_{\text{opt}}$ gives

$$p_0 = -\lambda\omega\,(q_0 - q_{\text{opt}})\coth(T\omega). \tag{11.12}$$

Eqns. (11.11) verify the claim made above that, if we use the same (quadratic) cost function as Almgren and Chriss (2001), then the Almgren and Chriss (2001) trading paths can be derived from Hamiltonian dynamics. Amusingly, this provides a rather direct analogy to classical mechanics, in which the variance term plays the role of a "potential energy" and the trading cost term plays the role of a "kinetic energy."

Theorem 11.1. In the general, multi-asset case, A defined by (11.8) is always diagonalizable. The eigenvalues of A are the positive and negative square roots of the eigenvalues of the positive-definite symmetric matrix $\kappa\Lambda^{-1/2}\Sigma\Lambda^{-1/2}$.

Proof. Write

$$A = \begin{pmatrix} 0 & D \\ S & 0 \end{pmatrix}$$

where $D = \Lambda^{-1}$ is diagonal and $S = \kappa\Sigma$ is symmetric. The eigenvalue equation for A then reads $Av = \gamma v$, and if $v = (x \ \ y)^\top$ then this becomes $Dy = \gamma x, \quad Sx = \gamma y$ which gives for x and y:

$$\gamma^2 x = DSx, \quad \gamma^2 y = SDy.$$

The eigenvectors of A must therefore be of the form $(x \ \ y)^\top$ where $x \in \mathbb{R}^n$ is an eigenvector of DS and $y \in \mathbb{R}^n$ is an eigenvector of SD, and the eigenvalues of A are square roots of the associated eigenvalues of DS or SD. Note that DS or SD have the same eigenvalues, and each has the same eigenvalues as the symmetric, positive-definite matrix

$$M := D^{1/2}SD^{1/2}. \tag{11.13}$$

Let $(v_i)_{1 \le i \le n}$ be a basis of eigenvectors for M, so that $Mv_i = m_i v_i$ with $m_i > 0$. Let $x_i = D^{1/2} v_i$ and let $y_i = m_i^{-1/2} S x_i$. Then

$$A \begin{pmatrix} x_i \\ y_i \end{pmatrix} = A \begin{pmatrix} D y_i \\ S x_i \end{pmatrix} = \begin{pmatrix} m_i^{-1/2} D S x_i \\ m_i^{1/2} y_i \end{pmatrix} = \sqrt{m_i} \begin{pmatrix} x_i \\ y_i \end{pmatrix}$$

Similarly, if we take $y_i = -m_i^{-1/2} S x_i$ we construct the eigenvector with eigenvalue $-\sqrt{m_i}$. We have thus constructed a basis for $\mathbb{R}^{2n}$ consisting of eigenvectors of A. Hence A is diagonalizable, completing the proof. $\square$

By Theorem 11.1 we can write $A = P^{-1} \Gamma P$ for some invertible matrix P whose columns are the eigenvectors of A, and some diagonal matrix $\Gamma = \operatorname{diag}(\gamma_1, \ldots, \gamma_{2n})$, where $\gamma_\alpha, \alpha = 1 \ldots 2n$ are the eigenvalues of A. This is also the Jordan canonical form of A, and gives the usual way of computing the matrix exponential:

$$\exp(tA) = P^{-1} \exp(t\Gamma) P \tag{11.14}$$

Moreover, as the proof of Theorem 11.1 actually constructs the eigenvectors, we can see that the columns of P are of the form $(x_i \quad y_i)^\top$ where $x_i = \Lambda^{-1/2} v_i$ and $y_i = \pm m_i^{-1/2} \kappa \Sigma \Lambda^{-1/2} v_i$ for some eigenvector v_i of M with $M v_i = m_i v_i$. As an aside, we have also shown that A is *traceless*, $\operatorname{tr}(A) = 0$, and hence that $\det(e^{tA}) = \exp(\operatorname{tr}(tA)) = 1$ which checks out for (11.10).

To complete the solution, we need to impose the final condition $q(T) = q_{\text{opt}}$ as before. In so doing we find a multi-asset analogue of (11.12). For simplicity, take $q_{\text{opt}} = 0$ (ie. liquidation as opposed to trading towards a fixed target), and introduce notation for the $n \times n$ blocks within $\exp(tA)$ as follows:

$$U(t) = \exp(tA) = P^{-1} \exp(t\Gamma) P = \left(\begin{array}{c|c} U_{11}(t) & U_{12}(t) \\ \hline U_{21}(t) & U_{22}(t) \end{array} \right)$$

Then $p_0 \in \mathbb{R}^n$ is determined by solving the n equations $[U(T)(q_0 \quad p_0)^\top]_i = 0$ for all $i = 1, \ldots, n$. These equations determine p_0, implicitly, as a function of q_0 and T. Explicitly,

$$p_0 = \Xi(q_0, T) = -U_{12}(T)^{-1} U_{11}(T) q_0 \tag{11.15}$$

where this equation serves to define $\Xi(q_0, T)$. Note (11.15) is the n-dimensional analogue of (11.12).

This formulation allows a computationally efficient implementation. Computing $U(t) = P^{-1} \exp(t\Gamma) P$ for various values of t is fast once the diagonalization of A (yielding P and Γ) has been computed. Furthermore, the matrix $-U_{12}(t)^{-1} U_{11}(t)$ could be computed once at the beginning of the day for all t in a sufficiently fine grid, ie. it does not need to be recomputed each time one has a new portfolio q_0 and wishes to find the associated generalized momenta from (11.15). Hence an execution desk could effect the liquidation of hundreds of different portfolios using the same set of $U(t)$, computed once before the trading day begins.

Eq. (11.14) together with the observations just made thus provide a complete prescription for computing the paths $q(t)$ and generalized momenta $p(t)$ from the Duhamel formula (11.9). The only expensive step is obtaining the eigenvalue decomposition of M, defined above in (11.13), but this need only be done once at the beginning of the execution path (not at every step of the algorithm to be presented below). To build intuition in a simple special case, note that when all assets have the same cost function, then Λ is proportional to the identity matrix, and v_i and x_i are proportional to the principal component directions.

Finally, to implement the above procedure in practice, the asset-level covariance $\Sigma \in S^n_{++}$ must be estimated. We adopt the standard factor model approach, in which we assume that a stock's returns over a given period is the sum of components arising from (unobserved) returns to certain common factors and an idiosyncratic component. This implies a reduction of Σ to the form

$$\Sigma := D + XFX' \tag{11.16}$$

where X is an $n \times k$ vector matrix of factor loadings, F is the covariance matrix of the factors, and

$$D := \mathrm{diag}(\sigma_1^2, \dots, \sigma_n^2) \quad \text{with all} \quad \sigma_i^2 > 0. \tag{11.17}$$

is the diagonal matrix of idiosyncratic variances.

If n is large, we choose $k \ll n$ and use a model of the form (11.16) to estimate Σ. Practically, (and in our numerical experiments) we take the matrix X to consist of several "style factors" including beta, size, momentum, value, and volatility inspired by Fama and French (1993) and Carhart (1997), augmented with an industry classification based on GICS. We estimate D and F empirically. Bayesian estimation of σ_i^2 ensures that σ_i^2 are bounded away from zero; the inverse-gamma posterior has vanishingly small probability mass around the origin. Practically, this is important: with the diagonal elements of D bounded away from zero, $\Sigma = D + XFX'$ is stably and efficiently invertible because it is diagonal plus low-rank.

11.2. **Generalized Momenta and the Value Function.** In this section, we wish to connect the generalized momenta discussed above with optimization theory and dynamic programming. In the notation from the previous section, define the *value function*

$$\begin{aligned} V(t,x) \quad &= \quad -\min_q \int_t^T L(q(s), \dot{q}(s))\, ds \\ \text{s.t.} \quad & \quad q(t) = x \quad \text{and} \quad q(T) = q_{\mathrm{opt}}. \end{aligned} \tag{11.18}$$

Thus $V(t,x)$ is the remaining utility gain from time t obtained from following the best policy, when the current state at time t is x. (We use the negative of the integral so that a higher value is better.)

Theorem 11.2. The function V above satisfies the Hamilton-Jacobi differential equation:

$$\frac{\partial V}{\partial t} + H(x, \nabla V) = 0 \tag{11.19}$$

with the singular final condition:

$$V(T,x) = \begin{cases} 0, & \text{if } x = q_{\text{opt}} \\ \infty, & \text{if } x \neq q_{\text{opt}}. \end{cases}$$

Proof: Let q^* be the path solving the problem (11.18), and let $h > 0$ be a small number. By Bellman's principle, we have

$$V(t,x) = -\int_t^{t+h} L(q^*(s), \dot{q}^*(s))ds + V(t+h, q^*(t+h)) \tag{11.20}$$

$$\approx -L(x, \dot{q}^*(t))h + V(t,x) + \frac{\partial V}{\partial t} h + \nabla V \cdot \dot{q}^*(t)h. \tag{11.21}$$

where we have dropped terms of order greater than 1 in h.

Then the optimality of q^* implies that the path $q^*|_{[t,t+h]}$ must maximize the right hand side of (11.20) among all paths on the interval $[t, t+h]$ such that $q^*(t) = x$; since h is small, such paths can be well approximated by linear paths; thus $\dot{q}^*$ must solve

$$\sup_v \left\{ -L(x,v) + \nabla V \cdot v \right\}. \tag{11.22}$$

Inserting this expression into (11.21), collecting terms gives

$$\begin{aligned} 0 &\approx \frac{\partial V}{\partial t} h + \sup_v \left\{ -L(x,v) + \nabla V \cdot v \right\} h \\ &= \frac{\partial V}{\partial t} h + H(x, \nabla V) h; \end{aligned}$$

dividing through by h and letting $h \to 0$ gives the Hamilton-Jacobi equation, completing the proof. □

The above proof also gives an interpretation of the generalized momenta p in terms of the value function. Indeed, the first-order conditions for (11.22) imply that along an optimal trajectory q, the equation

$$\nabla V = \frac{\partial L}{\partial v}(q, \dot{q})$$

holds. The right-hand side is precisely the definition of p from (11.3). Thus p is the direction in which the "value of the position" increases most rapidly.

References

Almgren, Robert and Neil Chriss (2001). "Optimal execution of portfolio transactions". In: *Journal of Risk* 3, pp. 5–40.

Benveniste, Elie Jerome and Gordon Ritter (2017). "Optimal Microstructure Trading with a Long-Term Utility Function". In: *Available at SSRN*. URL: `https://ssrn.com/abstract=3057570`.

Carhart, Mark M (1997). "On persistence in mutual fund performance". In: *The Journal of Finance* 52.1, pp. 57–82.

Fama, Eugene F and Kenneth R French (1993). "Common risk factors in the returns on stocks and bonds". In: *Journal of financial economics* 33.1, pp. 3–56.

Gatheral, Jim (2010). "No-dynamic-arbitrage and market impact". In: *Quantitative finance* 10.7, pp. 749–759.