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. We will denote the time-derivative by either a prime or a dot over the letter, thus the trading rate (in dollars per unit time) is $\dot{q}(t)$, or q' which both are defined to be $dq(t)/dt$. We shall denote the transpose of a matrix A by $A^\top$ in this lecture, so as to avoid using prime for two different things.

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 an APT-inspired model.

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_T.

The trader's problem can be formulated as follows:

$$\begin{gathered}\min_q J[q] \text{ for } q \in C^2([0,T], \mathbb{R}^n) \\ J[q] := \int_0^T L(t, q(t), \dot{q}(t))\, dt \\ \text{subject to } q(0) = q_0, q(T) = q_T\end{gathered} \tag{11.1}$$

where L is called the *Lagrangian*; an example Lagrangian relevant for Almgren–Chriss style execution is given by

$$L(q, v) = c(v) + \frac{1}{2}\kappa(q - q_T)^\top \Sigma (q - q_T)\,. \tag{11.2}$$

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

$J[q]$ is referred to as the cost corresponding to q. A function $q : [0,T] \to \mathbb{R}$ is termed *admissible* if it satisfies the boundary constraints and lies in the appropriate class, in this case $C^2[0,T]$. A *solution* q_* of (11.1) refers to an admissible function q_* such that $J[q_*] \le J[q]$ for all other admissible functions q. We also refer to q_* as a *minimizer* for the problem. The first main question is whether any solutions exist.

Theorem 11.1 (Tonelli 1915). Let the Lagrangian $L(t, q, v)$ be continuous, convex in v, and coercive of degree $r > 1$, i.e. for certain constants $\alpha > 0$ and β, we have

$$L(t, q, v) \ge \alpha \|v\|^r + \beta \quad \forall (t, q, v) \in [0,T] \times \mathbb{R}^n \times \mathbb{R}^n.$$

Then the basic problem (11.1) admits a solution in the class $\mathrm{AC}[0,T]$.

The notation "AC" may be new at this point; we will review it shortly.

11.2. **Review of Absolute Continuity.** Let I be an interval in the real line $\mathbb{R}$. A function $f\colon I \to \mathbb{R}$ is absolutely continuous on I if for every positive number ε, there is a positive number δ such that whenever a finite sequence of pairwise disjoint sub-intervals (x_k, y_k) of I with $x_k, y_k \in I$ satisfies

$$\sum_k (y_k - x_k) < \delta$$

then

$$\sum_k |f(y_k) - f(x_k)| < \varepsilon.$$

The collection of all absolutely continuous functions on I is denoted $\mathrm{AC}(I)$.

The following conditions on a real-valued function f on a compact interval $[a, b]$ are equivalent:

(1) f is absolutely continuous;
(2) f has a derivative f' almost everywhere, the derivative is Lebesgue integrable, and
$$f(x) = f(a) + \int_a^x f'(t)\, dt$$
for all $x \in [a, b]$;
(3) there exists a Lebesgue integrable function g on $[a, b]$ such that
$$f(x) = f(a) + \int_a^x g(t)\, dt$$
for all x in $[a, b]$.

If these equivalent conditions are satisfied then necessarily $g = f'$ almost everywhere. Equivalence between (1) and (3) is known as the fundamental theorem of Lebesgue integral calculus

11.3. **Euler's theorem.** The partial derivatives of the function $L(t, q, v)$ with respect to q and to v are denoted by L_q and L_v.

Theorem 11.2 (Euler 1744). If q^* is a solution of (11.1) then q^* satisfies the Euler equation:

$$\frac{d}{dt}\Big\{L_v(t, q_*(t), \dot{q}_*(t))\Big\} = L_q(t, q_*(t), \dot{q}_*(t)) \quad \forall t \in [0, T].$$

Proof. define a "variation" to be a function $y \in C^2[0, T]$ such that $y(0) = y(T) = 0$. We fix such a y, and proceed to consider the following function g of a single variable:

$$g(\lambda) = J[q_* + \lambda y] = \int_0^T L(t, q_* + \lambda y, q_*' + \lambda y')dt$$

We have simplified notation by leaving out certain arguments in the expression for the integral; thus q_* should be $q_*(t)$ and y is really $y(t)$, and so on.

From standard results in calculus, g is differentiable, and that we can "differentiate through the integral" to obtain

$$g'(\lambda) = \int_0^T [L_q(t, q_* + \lambda y, q'_* + \lambda y')y + L_v(q_* + \lambda y, q'_* + \lambda y')y']dt$$

Observe now that for each λ, the function $q_* + \lambda y$ is admissible, whence

$$g(\lambda) = J[q_* + \lambda y] \geq J[q_*] = g(0)$$

It follows that g attains a minimum at $\lambda = 0$, and hence that $g'(0) = 0$; thus:

$$\int_0^T [\alpha(t)y(t) + \beta(t)y'(t)]\, dt = 0$$

where we have set

$$\alpha(t) = L_q(t, q_*(t), q'_*(t)), \quad \beta(t) = L_v(t, q_*(t), q'_*(t))$$

Using integration by parts, we deduce

$$\int_0^T [\alpha(t) - \beta'(t)]y(t)\, dt = 0$$

Since this is true for any variation y, it follows that the continuous function which is the coefficient of y under the integral sign must vanish identically on $[0, T]$ (exercise: show this). But this conclusion is precisely Euler's equation $\square$.

A function $q \in C^2([0, T], \mathbb{R}^n)$ satisfying Euler's equation is referred to as an *extremal*. Not every extremal is a solution; one sometimes encounters a path-space analogue of saddle points, but every solution is an extremal. The Lagrangian L is said to be *autonomous* if it has no explicit dependence on the t variable. We will mostly work with autonomous Lagrangians for simplicity.

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

$$p := L_v(q, \dot{q}) \tag{11.3}$$

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

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

where this defines ψ.

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*:

$$\dot{q} = H_p(q,p), \quad \dot{p} = -H_q(q,p) \tag{11.5}$$

In many cases, Hamilton's equations can be solved explicitly; for example, take $q_T = 0$, and $\alpha(t) = 0$. Suppose

$$c(v) = \frac{1}{2}v'\Lambda v$$

where

$$\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$$

is a diagonal matrix with $\lambda_i > 0$ for all i. From (11.3), the generalized momenta are $p = \Lambda\dot{q}$ and hence algebraically solving, one has $\phi(q,p) = \Lambda^{-1}p$. The Hamiltonian is then

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

11.5. **Solving the First-Order System.** 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

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

Hamilton's equations then become:

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

or more simply

$$\frac{d\xi}{dt} = A\xi + b, \tag{11.7}$$

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

$$\xi = \begin{pmatrix} q \\ p \end{pmatrix}, \ A = \begin{pmatrix} 0 & D \\ S & 0 \end{pmatrix} \text{ and } b = \begin{pmatrix} 0 \\ -\kappa\Sigma q_{\mathrm{opt}} \end{pmatrix}. \tag{11.8}$$

where $D = \Lambda^{-1}$ is diagonal and $S = \kappa\Sigma$ is symmetric.

Duhamel's formula yields

$$\xi(t) = e^{tA}\xi_0 + \int_0^t e^{(t-s)A} b\, ds \tag{11.9}$$

To make use of (11.9) one must compute the one-parameter group e^{tA}. Computing matrix exponentials such as $\exp(tA)$ is especially easy when A is diagonalizable, in which case it reduces to computing functions $\exp(t\gamma_i)$ where $\{\gamma_i\}$ are the eigenvalues of A. 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

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

Note that p_0 is still unspecified. The path given by (11.11) satisfies the initial condition $q(0) = q_0$ for any p_0, but 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." What is kinetic energy, if not a number that's higher if you're moving faster? Trading cost has this character: higher trading rates cost more in terms of market impact.

Theorem 11.3. 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. Recall that

$$A = \begin{pmatrix} 0 & D \\ S & 0 \end{pmatrix}, \quad D = \Lambda^{-1}, S = \kappa\Sigma \tag{11.13}$$

where D is diagonal and S is symmetric. Let v denote any generic eigenvector for A, with corresponding eigenvalue γ. Denote the first n components of v by $x \in \mathbb{R}^n$ and the second n components by $y \in \mathbb{R}^n$, ie

$$v = \begin{pmatrix} x \\ y \end{pmatrix}$$

We shall write the eigenvalue equation for A as separate equations for x and y.

The eigenvalue equation $Av = \gamma v$ then becomes

$$Dy = \gamma x, \quad Sx = \gamma y.$$

Apply S to the equation $Dy = \gamma x$ and apply D to the equation $Sx = \gamma y$, 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

$$M := D^{1/2} S D^{1/2}. \tag{11.14}$$

Indeed, if $M\nu = m_i \nu$ for any vector $\nu \in \mathbb{R}^n$, then $DSD^{1/2}\nu = m_i D^{1/2}\nu$ and hence $D^{1/2}\nu$ is an eigenvector for DS with eigenvalue m_i.

We now construct eigenvectors for A from those of M. Let $(v_i)_{1 \leq i \leq n}$ 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 $y_i = m_i^{-1/2} S x_i$. It follows that

$$DSx_i = DSD^{1/2} v_i = D^{1/2} M v_i = m_i D^{1/2} v_i = m_i x_i$$

Then

$$A \begin{pmatrix} x_i \\ y_i \end{pmatrix} = \begin{pmatrix} Dy_i \\ Sx_i \end{pmatrix} = \begin{pmatrix} m_i^{-1/2} DSx_i \\ m_i^{1/2} y_i \end{pmatrix} = \sqrt{m_i} \begin{pmatrix} x_i \\ y_i \end{pmatrix}$$

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 11.3 we can write $A = P\Gamma P^{-1}$ for some invertible matrix P whose columns are the eigenvectors of A, and some diagonal matrix $\Gamma = \mathrm{diag}(\gamma_1, \ldots, \gamma_{2n})$, 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:

$$\exp(tA) = P \exp(t\Gamma) P^{-1} \tag{11.15}$$

Moreover, as the proof of Theorem 11.3 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*, $\mathrm{tr}(A) = 0$, and hence that $\det(e^{tA}) = \exp(\mathrm{tr}(tA)) = 1$ which checks out for (11.10).

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

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

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

$$p_0 = \Xi(q_0, T) = -U_{12}(T)^{-1} U_{11}(T) q_0 \tag{11.16}$$

where this equation serves to define $\Xi(q_0, T)$. Note (11.16) is the n-dimensional analogue of (11.12).

This formulation allows a computationally efficient implementation. Computing $U(t) = P \exp(t\Gamma) P^{-1}$ 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 (11.16). 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. (11.15) 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 (11.14), 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.

11.6. **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*

$$
\begin{aligned}
V(t,x) \quad &= \quad -\min_q \int_t^T L(q(s),\dot q(s))\,ds \qquad (11.17)\\
\text{s.t.} \qquad & q(t)=x \ \text{ and } \ q(T)=q_{\mathrm{opt}}.
\end{aligned}
$$

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, which we find more intuitive.

Theorem 11.4. The function V above satisfies the Hamilton-Jacobi differential equation:

$$\frac{\partial V}{\partial t} + H(x,\nabla V) = 0 \qquad (11.18)$$

with the singular final condition:

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

Proof: Let q^* be the path solving the problem (11.17), and let $h>0$ be a small number. By Bellman's principle, we have

$$
\begin{aligned}
V(t,x) &= -\int_t^{t+h} L(q^*(s),\dot q^*(s))ds + V(t+h,q^*(t+h)) \qquad (11.19)\\
&\approx -L(x,\dot q^*(t))h + V(t,x) + \frac{\partial V}{\partial t}h + \nabla V\cdot \dot q^*(t)h. \qquad (11.20)
\end{aligned}
$$

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 (11.19) 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

$$\sup_v \left\{-L(x,v)+\nabla V\cdot v\right\}. \qquad (11.21)$$

Inserting this expression into (11.20), 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 (11.21) imply that along an optimal trajectory q, the equation

$$\nabla V = L_2(q,\dot q)$$

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`.

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