Multi-period Optimization

General Multiperiod Problems

- Gârleanu and Pedersen (2013) studied the multiperiod quantitative-trading problem under the somewhat restrictive assumptions that the alpha models follow mean-reverting dynamics and that the only source of trading frictions are purely linear market impacts (leading to purely quadratic impact-related trading costs).
- We are going to do something similar, but not so restrictive and general enough to apply to real trading scenarios.

- We now place ourselves into the position of a rational agent planning a sequence of trades beginning presently and extending into the future.
- Specifically, a *trading plan* for the agent is modeled as a specific portfolio sequence

$$\boldsymbol{h} = (h_1, h_2, \ldots, h_T),$$

where h_t is the portfolio the agent plans to hold at time t in the future.
- If r_{t+1} is the vector of asset returns over $[t, t+1]$, then the trading profit (ie. difference between initial and final wealth) associated to the trading plan $\boldsymbol{h}$ is given by

$$\pi(\boldsymbol{h}) = \sum_t [h_t \cdot r_{t+1} - c_t(h_{t-1}, h_t)] \tag{11.1}$$

where $c_t(h_{t-1}, h_t)$ is the total cost (including but not limited to market impact, spread pay, borrow costs, ticket charges, financing, etc.) associated with holding portfolio h_{t-1} at time $t-1$ and ending up with h_t at time t.

- Trading profit $\pi(\boldsymbol{h})$ is a random variable, since many of its components are future quantities unknowable at time $t = 0$.
- Thus the problem we treat initially is that of maximizing $u(\boldsymbol{h})$, where

$$u(\boldsymbol{h}) := \mathbb{E}[\pi(\boldsymbol{h})] - (\kappa/2)\mathbb{V}[\pi(\boldsymbol{h})] \tag{11.2}$$

- Our task is to find the maximum-utility path $\boldsymbol{h}^* = \operatorname{argmax}_{\boldsymbol{h}} u(\boldsymbol{h})$.

- Since they will come up over and over again, let us define the shorthand notations

$$\alpha_t := \mathbb{E}[r_{t+1}] \quad \text{and} \quad \Sigma_t := \boldsymbol{V}[r_{t+1}], \quad y_t := (\kappa\Sigma_t)^{-1}\alpha_t. \tag{11.3}$$

- Then combining (11.1) with (11.2), one has

$$u(\boldsymbol{h}) = \sum_t \left[h_t' \alpha_t - \frac{\kappa}{2} h_t' \Sigma_t h_t - c_t(h_{t-1}, h_t) \right] \qquad (11.4)$$

Note that any symmetric, positive-definite matrix Q defines a bilinear form

$$b_Q(x, y) = N_Q(x - y).$$

Lemma 11.1

One has

$$-b_{\kappa\Sigma_t}(h_t, y_t) = O(y^2) + h_t'\alpha_t - \frac{1}{2}h_t'(\kappa\Sigma_t)h_t$$

where $O(y^2)$ denotes a collection of terms which is a quadratic function of y_t and which doesn't contain h_t.

- The proof is left as an exercise.

- Therefore the first two terms in the utility calculation (11.4) (ie. all the terms not dealing with costs) are given by

$$-b_{\kappa\Sigma_t}(h_t, y_t)$$

- Then up to $\boldsymbol{h}$-independent terms,

$$u(\boldsymbol{h}) = -\sum_t \left[b_{\kappa\Sigma_t}(h_t, y_t) + c_t(h_{t-1}, h_t) \right] \tag{11.5}$$

- Note that maximizing (11.5) is naturally a tracking problem.
- We are finding the sequence h_t that minimizes tracking error $b_{\kappa\Sigma_t}(h_t, y_t)$ but also minimizes cost $c_t(h_{t-1}, h_t)$.

- If you had written a computer program that maximizes (11.5), you could also apply it to other cases where y_t was something other than (11.3).
- For example, applying it to the sequence $y_t = 0$, and adding the appropriate constraint, we recover the Almgren-Chriss problem.
- For hedging exposure to derivatives, y_t should be our expectation of the offsetting replicating portfolio at all future times until expiration.

- Tracking the portfolios of Black and Litterman (1992) is also a special case of our framework in which y_t is the solution to a mean-variance problem with a Bayesian posterior distribution for the expected returns.
- Since the posterior is Gaussian in the original Black-Litterman model, the two-moment approximation to utility is exact, and one simply replaces α_t and Σ_t with the appropriate quantities.

- We are now starting to think that a computer program that maximizes (11.5) would be pretty useful.
- Next we describe how to write such a program, in practical terms.
- It's surprisingly easy, and you'll do it on your homework.

Non-differentiable Optimization

- Given a convex, differentiable map $f : \mathbb{R}^n \to \mathbb{R}$, if we are at a point x such that $f(x)$ is minimized along each coordinate axis, have we found a global minimizer?
- In other words, does

$$f(x + d \cdot e_i) \geq f(x) \quad \text{for all} \quad d, i$$

  imply that $f(x) = \min_z f(z)$?
- Here $e_i = (0, ..., 1, ...0) \in \mathbb{R}^n$, the i-th standard basis vector.

- The answer is: Yes!

- Now consider the same question, but without the differentiability assumption.

The answer changes to no, and Fig. 11.1 below gives a counterexample.

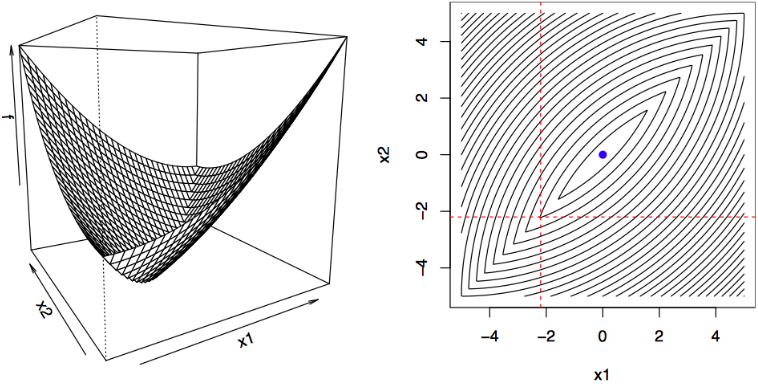

Figure: A convex function for which coordinate descent will get "stuck" before finding the global minimum.

Consider the same question again: "if we are at a point x such that $f(x)$ is minimized along each coordinate axis, have we found a global minimizer?" only now

$$f(x) = g(x) + \sum_{i=1}^{n} h_i(x_i)$$

with g convex, differentiable and each h_i convex? (In this case, we say the non-differentiable part is *separable*.)

- If the non-differentiable term is separable, the answer is yes once again.
- This is a special case of a deep general result proved by Tseng (2001), which we will call "Tseng's theorem."
- The main take-away is: we can easily optimize

$$f(x) = g(x) + \sum_{i=1}^{n} h_i(x_i)$$

with g convex, differentiable and each h_i convex, by coordinate-wise optimization.

Tseng's results also suggest an algorithm, called *blockwise coordinate descent* (BCD).

Algorithm 11.1

Chose an initial guess for x. Repeatedly iterate cyclically through $i = 1, \ldots, N$, and perform the following optimization and update:

$$x_i = \underset{\omega}{\mathit{argmin}}\, f(x_1, \ldots, x_{i-1},\ \omega,\ x_{i+1}, \ldots, x_N)$$

- Tseng (2001) shows that for functions of the form above, any limit point of the BCD iteration is a minimizer of f.
- The order of cycling through coordinates is arbitrary; and we can use any scheme that visits each of $\{1, 2, \ldots, n\}$ every M steps for fixed constant M.
- We can also everywhere replace individual coordinates with blocks of coordinates.

Single-asset trading paths

- Now let us consider the multiperiod problem for a single asset, in which case the ideal sequence $\boldsymbol{y} = (y_t)$ and the holdings (or equivalently, hidden states) $\boldsymbol{h} = (h_t)$ are both univariate time series.
- Since the multiperiod many-asset problem can be reduced to iteratively solving a sequence of single-asset problems, the methods we develop in this section are important even if our main interest is in multi-asset portfolios.

- A very important class of examples arises when there are no constraints, but the cost function is a convex and non-differentiable function of the difference

$$\delta_t := h_t - h_{t-1}.$$

- This allows for non-quadratic terms as in Almgren et al. (2005) and non-differentiable terms such as linear proportional costs.

- In this case, we can use Tseng's theorem, applied to *trades* rather than *positions*.
- Writing

$$h_t = h_0 + \sum_{s=1}^{t} \delta_s,$$

the objective function becomes

$$u(\boldsymbol{h}) = -\sum_t \left[b\Big(h_0 + \sum_{s=1}^{t} \delta_s, y_t\Big) + c_t(\delta_t) \right] \tag{11.6}$$

- Eq. (11.6) satisfies the convergence criteria of Tseng (2001) that the non-differentiable term is separable *across time*, while the non-separable term is differentiable.
- One then performs coordinate descent over the trades $\delta_1, \delta_2, \ldots, \delta_T$.
- Almost any reasonable starting point will do to initialize the iteration, but if a warm start from a previous optimization is available, that may speed things along.
- This approach was introduced in Kolm and Ritter (2015).

Problem 11.2

Prove Lemma 11.1

Problem 11.3

Consider optimally trading a single stock over $T = 30$ days. Each period is one day, and you can trade once per day. The stock's daily return volatility is σ. Suppose your forecast is 50 basis points for the first period, and decays exponentially with half-life 5 days. This means that

$$\alpha_t := \mathbb{E}[r_{t,t+1}] = 50 \times 10^{-4} \times 2^{-t/5}. \tag{11.7}$$

Let $c(\delta)$ be the cost, in dollars, of trading δ dollars of this stock. For selling, $\delta < 0$. Following Almgren, we assume that

$$c(\delta) = PX \left(\frac{\gamma\sigma}{2} \frac{X}{V} \left(\frac{\Theta}{V} \right)^{1/4} + \operatorname{sign}(X)\, \eta\sigma \left| \frac{X}{V} \right|^{\beta} \right), \quad X = \delta / P$$

where P is the current price in dollars, X is the signed trade size in shares, V is the daily volume in shares, Θ is the total number of shares outstanding, and finally $\gamma = 0.314$ and $\eta = 0.142$ and $\beta = 0.6$ are constants fit to market data. For concreteness, suppose the asset we are trading has

Almgren, Robert et al. (2005). "Direct estimation of equity market impact". In: *Risk* 18.7, pp. 58–62.

Black, Fischer and Robert Litterman (1992). "Global portfolio optimization". In: *Financial Analysts Journal*, pp. 28–43.

Gârleanu, Nicolae and Lasse Heje Pedersen (2013). "Dynamic trading with predictable returns and transaction costs". In: *The Journal of Finance* 68.6, pp. 2309–2340.

Kolm, Petter N and Gordon Ritter (2015). "Multiperiod Portfolio Selection and Bayesian Dynamic Models". In: *Risk* March 2015.

Tseng, Paul (2001). "Convergence of a block coordinate descent method for nondifferentiable minimization". In: *Journal of optimization theory and applications* 109.3, pp. 475–494.