1 Insights

Idea 1. Convexity and non-uniqueness are different things. If convex problem has large eps-null space, i.e. large flat area around the optimum, then

- it is bad for consistency of solver: using different methods will give you different argmaxes;

- it is good in practice, since this allows for a secondary objective

Idea 2. $\max \mu^T w - \gamma w^T \Sigma w$ does not make sense because you are comparing terms that have $units$ and $units^2$, but $\max \mu^T w - \gamma' \sqrt{w^T \Sigma w}$ makes sense. By sweeping γ and γ' we get exactly same efficient frontier. It is also better computationally (more modern)!

Proof. When you have a bi-objective optimization problem and apply monotonic transformation to either or both and then take sum you get same efficient frontier. $\square$

Idea 3. If you solve LeastSquare problem within x ms, the cost of solving a convex problem with the same structure is about 10-100 times more, maybe 20. Because interior point method does about 20 iterations.

Idea 4. Risk neutrality. Instead of

$$m^T w = 0 \tag{1}$$

we should do

$$m^T \Sigma w = 0 \tag{2}$$

because this corresponds to orthogonality of market portfolio P&L($M = m^T r$) and portfolio P&L($R = w^T r$).

Idea 5. V@R is not convex, but cV@R is. And it turns out you don't want to do but V@R as it is not a:

- V@R – you want prob of your drawdown exceeding 10% is no more than 5%.

- cV@R – I want conditional expectation of the drawdown to be less than this given that is greater than 10%.

Idea 6. If you can add a lot of constraints and they don't change result (objective value) too much, then add as many and tighten up. It is like insurance for which you hardly pay anything, but if bad event happens you are in a much better shape.

Idea 7. $w \in \mathbb{R}^n_+ \Leftrightarrow \|w\|_1 \leq 1$, i.e. non negativity is constrained on leverage.

Idea 8. Non-negativity constraint, l1-norm and alpha uncertainty generate sparse solutions.

If something is small, in l1 you pay small, but in l2 you pay very small. If you have something already small, for l2 there is no particular motivation to make it zero. But for l2 the marginal advantage of making it smaller keeps up.

And for non-negativity, that's like having infinite value to the left of zero. As long as I have non-zero negative slope from right at zero, I'll have sparsity also. E.g. $max(0, (x+1)^2 - 1)$ or $max(0, x)$. See Figure 1
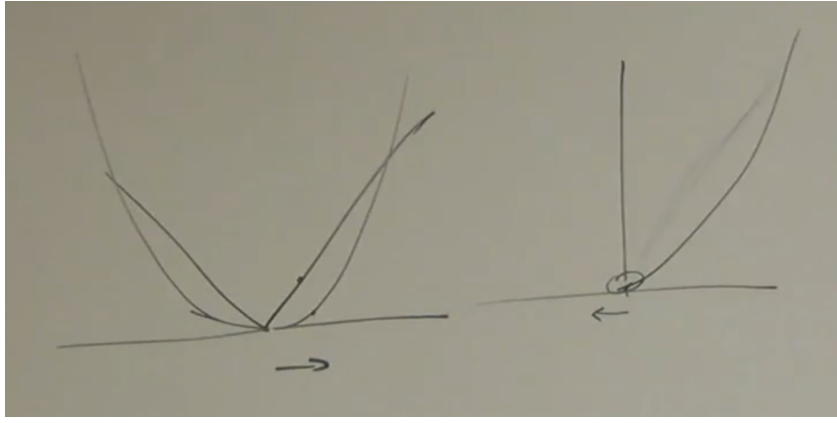
Idea 9. Every reasonable transaction cost is convex

Figure 1

Idea 10. k-day liquidation cost: $k\phi^{tc}(w/k)$. Because of convexity, you need to trade evenly. If part of ϕ^{tc} is linear (spread) cost, then this will involve $l1$ penalty on positions and we'll start seeing zero positions, because these stocks are too illiquid to hold and get out when necessary...

Idea 11. Doing optimization vs doing simulation are different things: people who do simulation certainly care whether impact has power $3/2$ or 2. People who do optimization just think of it as "something that discourages large trades", while spread cost term as "something that discourages small trades".

Idea 12. Introducing variables for exposures $f = F^t x$ should <u>always</u> be done:

- it reduces computational complexity from $\mathcal{O}(n^3)$ to $\mathcal{O}(nk^2)$;

- now we can put constraints on factor exposures.

Idea 13. Fitting low-rank + diagonal model is not a convex problem but it is perfectly solvable.

Idea 14. It is a very good idea to add constraints on exposures, because it is a better risk control. The quadratics is a BIG LIE. If you look at covariance matrix, there are only a few significant eigenvalues and it says that there are 3600 directions with zero risk! Constrain those directions!!!

Idea 15. Forecast uncertainty. Let's say instead of point estimate vector mu we are given intervals $\mu^{low} \le \mu \le \mu^{high}$. Introduce <u>vectors</u>

$$\bar{\mu} = \frac{\mu^{low} + \mu^{high}}{2} \tag{3}$$

$$\bar{\mu} = \frac{\mu^{high} - \mu^{low}}{2} \text{(vector of uncertainty)} \tag{4}$$

Then worst-case μ is easy to get analytically

$$\min_{\mu^{low} \le \mu \le \mu^{high}} \mu^T h = \bar{\mu} - \underbrace{\sigma^T |h|}_{\text{weighted l1-norm}} \tag{5}$$

The second term is alpha-uncertainty risk. L1-norm weighted by alpha uncertainties.

Idea 16. How to get covariance scenarios?

- manually (labeled)

- perturb nominal cov matrices in a reasonable way (remember to adjust corr matrix as it may stop being corr matrix: svd and replace negative eigenvalues, or add identity or project onto S_+ or closest matrix in S_+)