A History of Optimization
Gresham - BSHM 2008

Prof. William Shaw,

King's College London

william.shaw@kcl.ac.uk

www.mth.kcl.ac.uk/staff/w_shaw.html

Introduction: Scope of Talk

Optimization means many things:

Many types of functions we wish to maximize/minimize; profit, risk...

Single or multi-objective;

Different types of constraint;

Data known to varying degrees of certainty;

Integer vs real variables;

Travelling salesman -> advanced vehicle routing problems!

Utility maximization for many applications including option pricing...

....

Focus of this Talk

Note that my title said "A" history, not THE history. I will not claim to be comprehensive, but focus on the key things I have found interesting over the last 25 years, supplemented by some cultural themes! Many books can be written about this topic and I will optimize for the time allowed.

My interests are coloured by (a) having worked with investment banks on practicalities, (b) having an interest in computational aspects.

What assets to buy or short?
How much of each to have?

There are also temporal issues: how long to hold; how to most efficiently get there; minimize transaction costs (market impact).

Riskless Linear Problems

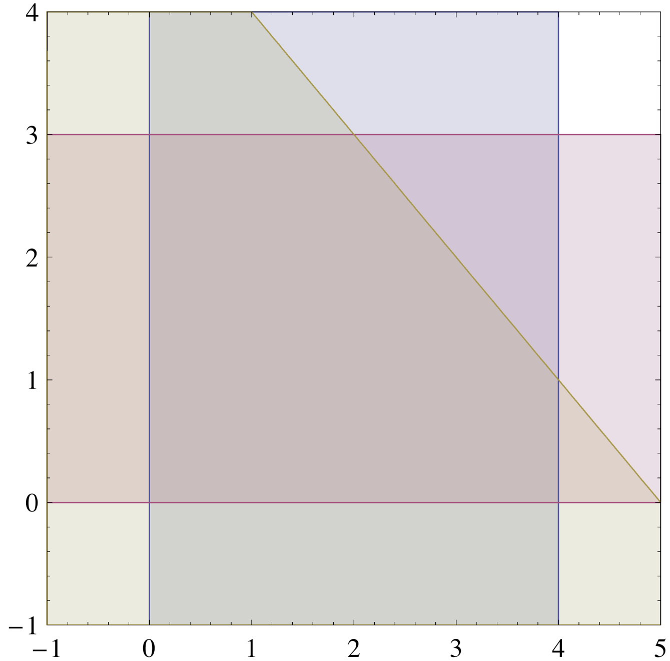
Some school syllabuses have basic ideas about linear programming.

Here is a classic example:

A farmer can keep sheep (X) or goats (Y), makes so much money out of each but there are costs and constraints.

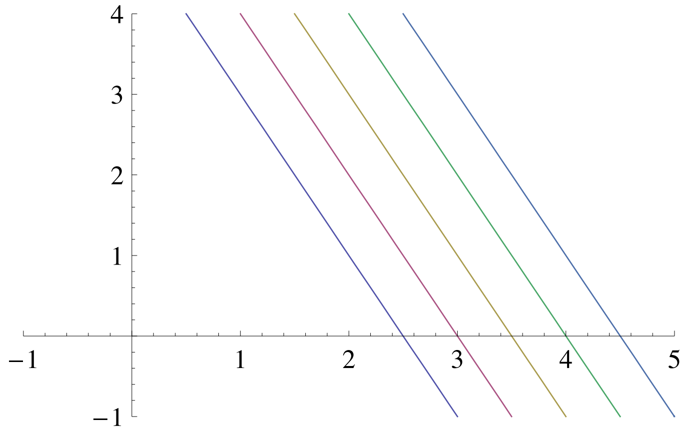
You try to maximize the profit for example, subject to some constraints, shown here:

```
constraints = RegionPlot[{0 ≤ x ≤ 4, 0 ≤ y ≤ 3, x + y ≤ 5}, {x, -1, 5}, {y, -1, 4}, PlotRange → {{-1, 5}, {-1, 4}}
```

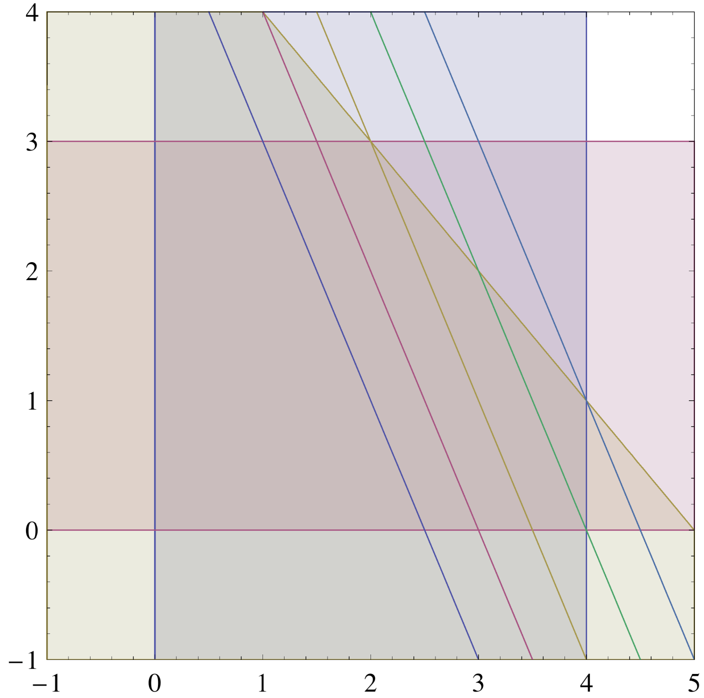


The profit takes the form of a multiple of $2x + y$, so you can look at the lines of constant profit - you would like them to maximize the profit:

```
profitlines = Plot[{-2 x + 5, -2 x + 6, -2 x + 7, -2 x + 8, -2 x + 9}, {x, -1, 5}, PlotRange → {{-1, 5}, {-1, 4}}]
```



```
Show[constraints, profitlines]
```



This type of linear programming has diverse applications.

Linear functions never have stationary values - the derivative never vanishes. So the extremal values are always on a boundary, and typically on a vertex, as in our example above.

So a good idea is to search among vertices. Find a feasible one and work your way around increasing the objective function at each move from vertex to vertex.

This finds expression in the "Simplex Algorithm",

listed by one Journal as one of top 10 algorithms of the 20th Century, as initially developed by George Dantzig just after WW2.

The basic idea is very simple, as per my graphs above.

If you have never looked at it my favourite entry point is in Chapter 10 of Numerical Recipes (3rd Ed), which gives a practical approach and some good references. A concept of 'duality' applies here as introduced by John van Neumann in same time frame.

Several people have worked on either improving the simplex method or doing it differently.

1972 - Klee and Minty showed that Dantzig's method can visit all vertices before getting to the right one, using a distorted cube. Gave a good reason for exploring other options.

1979 - Khachiyan - ellipsoid method O(n^4) but usually less efficient.

1984 - Great excitement in LP commuity as Narendra Karmaker publishes interior point method; 'Karmakar's algorithm' (efficient and polynomial time)

I recall great excitement in Cambridge when this came out! Which was faster? I am not sure this was ever generally resolved - one method may be faster in some situations, the other in others.

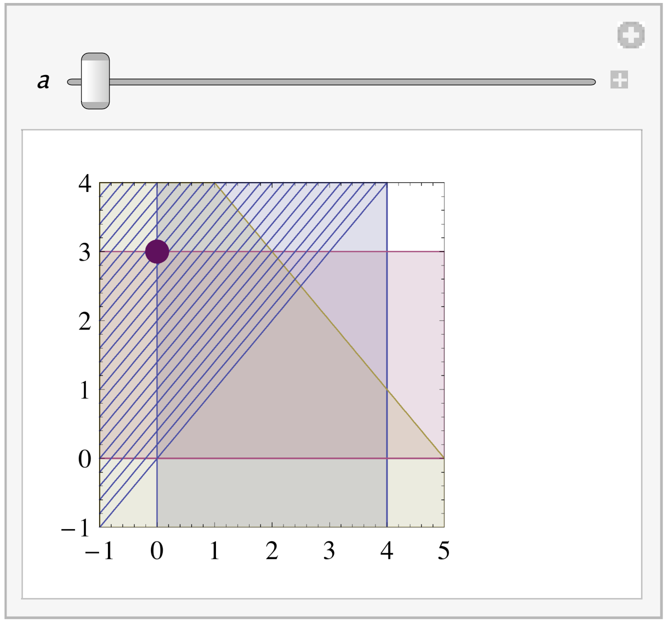
These days functionality is built in to many systems, e.g. here is a function treating all variables as positive (the initial scope) and minimizes m.x subject to $A.x \geq b$.

```
LinearProgramming[{-4, -1}, {{-1, 0}, {0, -1}, {-1, -1}}, {-4, -3, -5}]
```

{4, 1}

But there are open problems, to with efficiency (Smale) and integer unknowns. Let's visualize a problems that illustrates the non-uniqueness and "jumpiness" of optimization

```
Manipulate[profitlines = Plot[Table[-a x + k, {k, 0, 10, 0.2}], {x, -1, 5}, PlotRange → {{-1, 5}, {-1, 4}}]; Optipoint = LinearProgramming[{-a, -1}, {{-1, 0}, {0, -1}, {-1, -1}}, {-4, -3, -5}]; PtPlot = Show[Graphics[{PointSize[0.07], RGBColor[0.3, 0., 0.3], Point[Optipoint]}]];
 Show[constraints, profitlines, PtPlot], {a, -1, 2}]
```



Adding some Risk

In mathematical finance the idea of maximizing the profit is very important, but it is modulated by the need to simultaneously manage the risk (of a loss). It all boils down to a sophisticated version of

"Don't put all your eggs in the same basket".

This is an old idea:

"Tis the part of wise man to keep himself today for tomorrow, and not venture all his eggs in one basket."

(Cervantes, Don Quixote, 17th Century)

Shakespeare similarly illustrated the concept of diversification around the same time in *The Merchant of Venice.*

Antonio told his friends he wasn't spending sleepless nights worrying over his commodity investment:

"Believe me, no. I thank my fortune for it, My ventures are not in one bottom trusted, Nor to one place; nor is my whole estate upon the fortune of this present year. Therefore my merchandise makes me not sad."

Note that he adds the notion of temporal diversification!

The Weighting Problem

But if you are going to spread your assets around, HOW MUCH should you invest in each type of asset?

That is, given weights w_i with

$$\sum_{i=1}^{N} w_i = 1$$

how to you pick the w_i? These are the fractions you put in each asset.

People have worried about this for a long time.

Until relatively recently the weights were assumed to be non-negative - no short selling -in common with the old LP idea of actual positive quantities.

These days we are allowed to go short.

Ancient Views?

Rabbi Issac bar Aha is the oldest authority I am aware of

(if you can read Aramaic it is in the Babylonian Talmud, Tractate Baba Mezi'a, 4th Century):

Equal allocation strategy:

A third in land,
A third in merchandise,
A third in cash ("a third at hand").

OK - so this says $w_i = 1 / N$ in general.

Widely studied, and is a robust approach.

Chinese View?

I recently asked Professor Xunyu Zhou, Nomura Professor of Finance in Oxford, about optimization in ancient China. He came up with the following view:

"It appears to me that the notion of optimisation was non-existent in ancient China, although there had been great ancient Chinese mathematicians working in other areas of mathematics...

Optimisation is a very modern term imported from the west to China.

This may have to do with the central philosophy of Confucian/Chinese culture ``Zhong Yong Zhi Dao'', meaning ``staying in the middle/mediocre/median'', which is in sharp contrast with the very idea of optimisation (aka seeking perfection or zero tolerance). Optimisation means extreme, and extreme is never encouraged in the Chinese culture."

I certainly could not find anything along the lines of our Rabbi in the writings of Confucius, but he did have some good advice for someone who has just had to re-mortgage - it also translates well:

"*He who will not economize will have to agonize*."

Some time later! H. Markowitz and co-workers

Harry Markowitz went to work for the RAND corporation after WW2 and met Dantzig - got interested in optimization. This lead to some fascinating work which I will now oversimplify.

Minimize the combination

```
risk - λ*profit
```

where λ measures your risk-profit preference.

As it stands this, without constraints, this is an exercise in (partial) differentiation once you have an estimate of risk and profit as a function of the weights. For this and other work with Merton Miller and William Sharpe on financial economics the Nobel prize was awarded in 1990. This is interesting as when Harry's thesis was examined his ideas were thought of as "not economics" by Milton Friedman during his much earlier thesis defence.

There are various related formulations:

Maximize return subject to risk bound;
Minimize risk subject to return goal;

Work relative to an index: fund managers are very fond of this!

Made for some interesting reading (excuses) during the tech fund crash.

The classic choice of function is

$$\sum_{i=1}^{N}\sum_{j=1}^{N} C_{ij}\, w_i\, w_j - \lambda \sum_{i=1}^{N} R_i\, w_i$$

and gives a quadratic. In reality there are constraints, e.g. $w_i \geq 0$, and matrix equalities and inequalities. We get the classical theory of Quadratic Programming, which is a simple conceptual step up from the Linear Programming, combined with differentiation to find stationary points.

In linear programming the optimal solution is always at a boundary point.

In the quadratic case it might also be in the middle.

There is a complex interaction between the maximization problem and the constraints.

Here is a pure risk problem with three assets. We will consider a pure risk function defined by a covariance matrix:

```
cov3[r_] := {{64, 120 r, 25}, {120 r, 225, 50}, {25, 50, 100}};
MatrixForm[cov3[r]]
```

$$\begin{pmatrix} 64 & 120\,r & 25 \\ 120\,r & 225 & 50 \\ 25 & 50 & 100 \end{pmatrix}$$

We use one constraint to eliminate the third weight, and define a risk (total variance) as a function of the remaining two weights:

```
risk[wa_, wb_, r_] :=
Module[{wtvec},
wtvec = {wa, wb, 1-wa-wb};
Simplify[wtvec.cov3[r].wtvec]]
```

Interior minima may be found by ordinary differentiation:

```
eqna = D[risk[wa,wb,r],wa];
eqnb = D[risk[wa,wb,r],wb];
solns = Solve[{eqna==0, eqnb==0}, {wa, wb}];
```

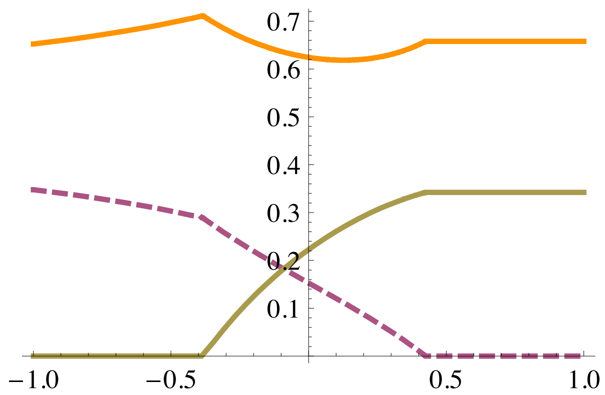
```
wpair[r_] = Simplify[{wa, wb} /. solns[[1]]]
```

$$\left\{\frac{5\,(48\,r-125)}{576\,r^2+240\,r-1001},\ \frac{9\,(40\,r-17)}{576\,r^2+240\,r-1001}\right\}$$

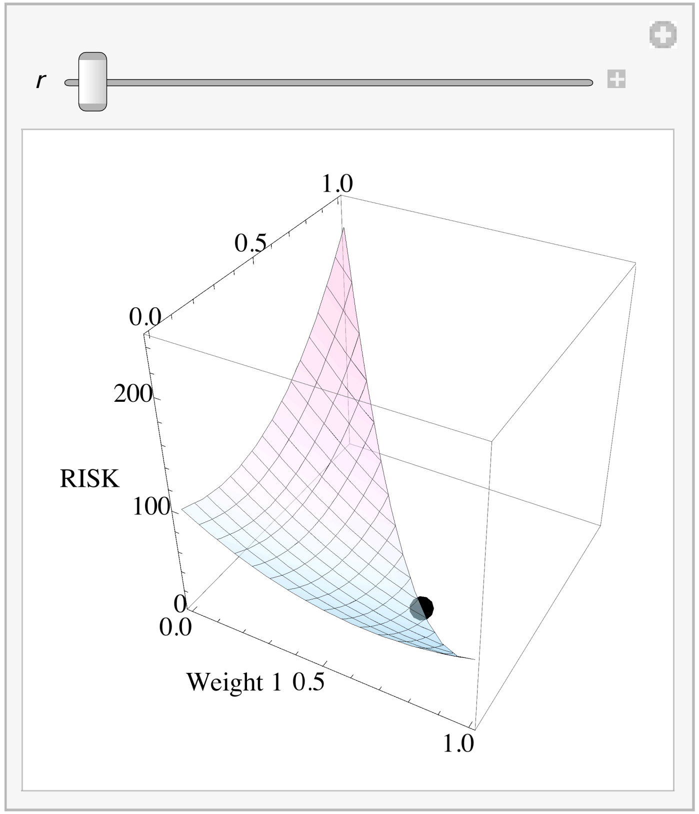
Some further analysis with the constraints applied gives a more complicated function:

```
w3[r_] = Simplify[1-wpair[r][[1]] - wpair[r][[2]]];
weights[r_] :=
Which[r < -0.383782,{(225-120*r)/(289-240*r),1-(225-120*r)/(289-240*r),0},
r < 0.425,{wpair[r][[1]], wpair[r][[2]], w3[r]}, True,
{0.657895, 0, 0.342105}]
```

```
Plot[{weights[r][[1]], weights[r][[2]], weights[r][[3]]}, {r, -1, 1}, PlotStyle→{{Thick, Orange}, {Thi
```



```
Manipulate[riskdata = Table[{wa, y * (1 - wa), risk[wa, y * (1 - wa), r]}, {y, 0, 1, 0.1}, {wa, 0, 1, 0.1}]; wts = weights[r]; surf = ListPlot3D[Flatten[riskdata, 1], BoxRatios → {1, 1, 1}, PlotRange → {0, 250}, Axes → True, PlotStyle -> Opacity[0.5], AxesLabel -
 pointplot = Show[Graphics3D[{PointSize[0.05], Point[{wts[[1]], wts[[2]], risk[wts[[1]], wts[[2]], r]}]}]]; Show[surf, pointplot], {r, -1, 1}]
```



Beyond QP and MV

Much has happened since and I will not pretend to be able to survey it all. I will again focus on matters I have found particularly interesting.

- **Beyond Quadratic Risk**

I will draw attention to some work by Mike Powell in Cambridge which I think is a notable continutation of the thread of analyzing the optimization of increasingly general functions (LP,QP,....) subject to constraints. Powell's1989 work was embodied in a DAMTP research report and F77 code! It has found its way into numerous library functions (e.g. variations of IMSL) and you can get it now for Java!

Mike said that IMSL was not overly happy with the code as it needed you to work out the derivatives! He has made his source availble and this is a valuable resource for anyone who wants to play with portfolio optimization, whether in the *quadratic or more general framework* that it can treat.

I mention this for several reasons.

One is because when reviewing portfolio optimization codes in the early 1990s I found a great many organizations wanting to sell me black box code for a lot of money, but Powell was a bit of an outlier and gave out source for almost nothing.

More important is that this allows for optimization of risk structures that are not necessarily quadratic, but with a complex array of linear equality and inequality conditions. Theoretically this allows for more general risk analysis, though I have always just used it for QP.

But his algorithm allows for

lower bounds on weights;
upper bounds on weights;
budget constraints;
sector exposure inequalities;
sector exposure equalities.

It also allows for "beyond quadratic" risk structures and has been tested against many hard problems.

I have written an interface to his model for the QP case:

```
Import["optidata.txt"]
```

```
   3
   8.000    15.000    20.000
   0.000
  64.000  -120.000    25.000
-120.000   225.000    50.000
  25.000    50.000   100.000
   0.000     0.000     0.000
   1.000     1.000     1.000
   1
   1
   1.000     1.000     1.000
   1.000
```

Can load complex constraint set, e.g. for sector exposures.

```
Run["rm weights.txt"];
```

```
Run["./optimizer"];
```

```
Import["weights.txt"]
```

```
weights[-1] // N
```

{0.652174, 0.347826, 0.}

Contact me if you want this model - for free!

CAPM, APT,... variations

Use of the full covariance structure is not the only way to go. Use of reduced structures has been proposed in many forms for various reasons. Focus on "essential part" of covariance:

- **CAPM - correlation with index is main driver - one "real" factor**

Ideas by Treynor, Sharpe, Lintner, Mosson

$R_{i\,t} = \alpha_i + \beta_i I_t + \epsilon_{it}$

$C_{i\,j} = \text{Diag}(\sigma_i^2) + \beta_i \beta_j \text{Var}(I)$

- **APT - many real or abstract factors - projection of Covariance onto relevant subset. Think of as vector of β**

Ideas by Ross, Roll and others, commercial variations by Barra and others...

$C_{i\,j} = \text{Diag}(\sigma_i^2) + \beta^T \Lambda \beta$

These all give different answers for optima!

Time-dependence

There are many types of generalization:

Analysis of periodic rebalancing: transaction costs and turnover constraints;

Getting from state A to state B - market impact and detailed timing (Almgren, Chriss, 1999+)

Continuous-time Markowtiz analysis - Zhou and collaborators (2000s)

Optimal control - stopping and starting processes (many)

"Post-modern" Portfolio Optimization Theory

Recognition that, e.g. in index-relative optimization, large positive deviations are just fine! It's the downside you need to worry about. Markowitz was clear on this, but practical computations took him back to variance.

Recognition that Variance is not all of Risk in realistic non-normal world.

More realistic objective functions under consideration.

Semi-Variance

Adding Skewness/Kurtosis

Avoiding moments - functions of distribution: Keating's Omega, Sortini ratio, relatives....

The Robustness Issue

We face the constant difficulty of unstable weights. Near-degeneracy in covariance can amplify problem.
Covariance never that well known anyway. What to do - work mostly in 2000s:

Imperial group : Manage with Scenarios

Goldfarb-Iyengar: Ellipsoidal uncertainty sets

Developments by Tütüncü, Hauser, Zuev and others

Common approach is to take the best worst case

- nested manageable problem
- is this right approach?

Most work keeps classic MV - need to link to realistic Objective Function with PMPT

"130/30" and 1X0/X0 family

I will end by discussing a current favourite of the investment community - the "130/30 investment strategy". What does this mean? Who knows? When I was asked to look into this a few months ago I found it very confusing as it is not actually well defined.

Common theme - start with index, go short 30% worst stocks, use cash raised to have a 130% \long side, and optimize it all (or not). LOTS of choices:....

What is α selection criteria?

Is the shorting absolute, or relative to the background index?

How sophisticated is the subsequent asset weighting choice, from equal to fully optimized?

Are other constraints applied? E.g. $\beta = 1$, and is this exact, $\beta = 1$, or approximate: e.g. $0.95 < \beta < 1.05$.

Once established, choices for turnover control on rebalancing, frequency of doing so

....

Are there benchmarks and what do they do?

A 130/30 "benchmark" is awkward because of all the choices, especially over alpha model

Let's look at some recently proposed:

S&P: D. M. Blitzer, P. Murphy, T. Eisenhauer; S&P 500 130/30 Strategy Index
(Index Methodology Note, Nov 2007)

AlphaSimplex-Credit-Suisse, A. Lo and P. Patel, 130/30: The New Long-Only, Dec. 2007

- **S&P model**

  Asset sorting from STARS rating from S&P Global Equity Research

  Shorting is RELATIVE to background

  Equally weighted 1% on 30 assets

  No forced beta constraint

  Apparently no optimization...

- **Lo-Patel**

  Alpha Model based on Credit-Suisse factor model

  Shorting is absolute to background, with cap on relative weights

  Optimization of weights

  Imposition of beta constraint as firm β=1

Why 30?

- On-line forum indicates there are some skeptics:

"Given the deeply mathematical and highly quantitative nature of equity fund investing, I would think that a new and highly proprietary multi-factor N-dimensional, hybrid-hyperbolic stochastic volatility and stochastic correlation matrix Ansatz is the basis for establishing 130/30 as the ultimate optimum for a long-short mix.

More likely though it is becase the MBA in charge of client presentation could not count beyond 130....."

- Patel (of Lo-Patel): "there is no mathematical justification"
- Regulatory

"The Federal Reserve Regulation T regulates the collateral requirements for margin loans and shorted stocks.These collateral requirements theoretically limit U.S.-domiciled mutual funds regulated by the Investment Company Act of 1940 (‘40 Act funds’) to a maximum 133 % long and 33 % short position."

Source: J.D. Martielli, SEI Investments, Jan 2005

"Long/Short Extensions: How Much is Enough?"
Clarke et al, Analytic Investors, 2007

Recent analysis makes several simplifying assumptions and steps for N-dimensional portfolio

assume covariance 2-parameter - constant sigma and rho, e.g. in 3D, N=3, the covariance is

$$\Omega = \sigma^2 \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}$$

Grinold's Alpha Generation Process

$$\alpha = IC * \Omega^{1/2} S$$

where S is N-vector of standard normal scores summing to zero, IC is information coefficient

Assume unconstrained mean-variance optimization

Without constraints the optimization of risk given a risk bound ${\sigma_A}^2$ gives

$$w_A = \sigma_a \, \Omega^{-1/2} \, S$$

and the cash constraint $1.\,w_A = 0$ is in fact satisfied, given the special nature of Ω, provided S sums to zero. The special form of Ω allows further simplification to

$$w_A = \sigma_a \, S \Big/ \left(\sigma \sqrt{N[1-\rho]}\right)$$

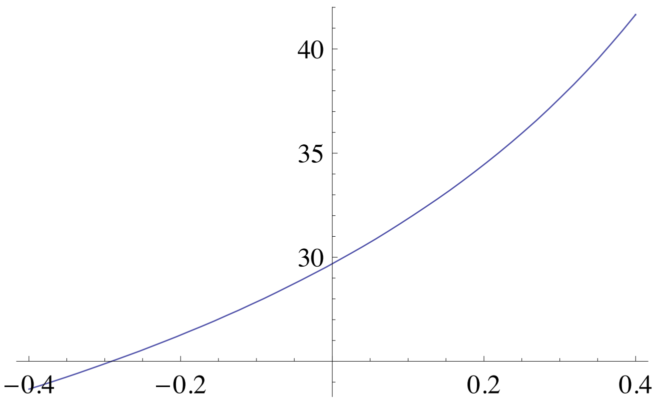
Treatment of S as normal random variable and introduction of background index weight gives expected short weight in terms of probability distributions.

Assumptions about index weight distributions, e.g geometric

Simplified form of model has five parameters:

n = no of assets;
ρ = correlation parameter;
σ = vol paramater;
TE = tracking error;
indexconc = index concentration parameter (SMALLER is more concentrated)

```
Plot[ShortWeight[100, ρ, 0.3, 0.04, 0.95], {ρ, -0.4, 0.4}]
```



Closing Remarks

Optimization is a topic that is under active investigation by many researchers, even within the narrow thread I have presented today.

We should all value it whether we are:

mathematicians interested in the mathematical ideas;

numerical analysts or computer scientists interested in efficient algorithms;

fund managers interested in a bonus.

Almost all of us have pensions we want managed well and on a rational basis.

Beyond that we want to see food distributed with minimal environmental impact, tasks allocated efficiently (Dantzig!), chips designed efficiently, video pathways in our AV amplifiers shortened, and so on and so on.....

william.shaw@kcl.ac.uk