Convex Optimization for High-Dimensional Portfolio Construction

R/Finance 2017

R. Michael Weylandt

2017-05-20

Rice University

In the beginning...

PORTFOLIO SELECTION*

HARRY MARKOWITZ
The Rand Corporation

THE PROCESS OF SELECTING a portfolio may be divided into two stages. The first stage starts with observation and experience and ends with beliefs about the future performances of available securities. The second stage starts with the relevant beliefs about future performances and ends with the choice of portfolio. This paper is concerned with the second stage. We first consider the rule that the investor does (or should) maximize discounted expected, or anticipated, returns. This rule is rejected both as a hypothesis to explain, and as a maximum to guide investment behavior. We next consider the rule that the investor does (or should) consider expected return a desirable thing *and* variance of return an undesirable thing. This rule has many sound points, both as a maxim for, and hypothesis about, investment behavior. We illustrate geometrically relations between beliefs and choice of portfolio according to the "expected returns—variance of returns" rule.

One type of rule concerning choice of portfolio is that the investor

$$
\begin{aligned}
\text{Given} \quad & \mu, \Sigma \\
\text{minimize} \quad & w^T \Sigma w - \lambda \mu^T w
\end{aligned}
$$

$$\begin{aligned} \text{Given} \quad & \mu, \Sigma \\ \text{minimize} \quad & w^T \Sigma w - \lambda \mu^T w \end{aligned}$$

Sample estimates:

$$\hat{\Sigma} = R^T R / n \qquad \hat{\mu} = R^T 1_n / n$$

$$\begin{aligned} \text{Given} \quad & \mu, \Sigma \\ \text{minimize} \quad & w^T \Sigma w - \lambda \mu^T w \end{aligned}$$

Sample estimates:

$$\hat{\Sigma} = R^T R / n \qquad \hat{\mu} = R^T 1_n / n$$

$$\text{minimize} \quad w^T (R^T R / n) w - \lambda w^T R^T 1_n / n$$

$$\text{Given} \quad \mu, \Sigma$$
$$\text{minimize} \quad w^T \Sigma w - \lambda \mu^T w$$

Sample estimates:

$$\hat{\Sigma} = R^T R / n \qquad \hat{\mu} = R^T 1_n / n$$

$$\text{minimize} \quad w^T (R^T R/n) w - \lambda w^T R^T 1_n / n$$

Simplify and factor:

$$\begin{aligned} w^T (R^T R/n) w - \lambda w^T R^T 1/n &\implies w^T (R^T R/n) w - \lambda w^T R^T 1_n / n \\ &\implies w^T R^T R w - 2 w^T R^T (1_n \lambda/2) + (1_n \lambda/2)^T (1_n \lambda/2) \\ &\implies \|(1_n \lambda/2) - R w\|_2^2 \end{aligned}$$

Linear Regression Connection

$$\underset{w \in \mathbb{R}^p}{\arg\min} \|(1_n \lambda/2) - Rw\|_2^2$$

Linear Regression Connection

$$\underset{w \in \mathbb{R}^p}{\arg\min} \|(1_n \lambda/2) - Rw\|_2^2$$

$\implies$ Ordinary Least Squares (OLS)

$$\underset{\beta \in \mathbb{R}^p}{\arg\min} \|y - X\beta\|_2^2$$

$$\arg\min_{w \in \mathbb{R}^p} \|(1_n \lambda/2) - Rw\|_2^2$$

$\implies$ Ordinary Least Squares (OLS)

$$\arg\min_{\beta \in \mathbb{R}^p} \|y - X\beta\|_2^2$$

We know a lot about OLS!

Portfolio constraints: concentration, diversification, transaction cost management, *etc.*

Portfolio constraints: concentration, diversification, transaction cost management, *etc.*

Example: invest in at most K shares

Constrained Portfolio Optimization

Portfolio constraints: concentration, diversification, transaction cost management, *etc.*

Example: invest in at most K shares

$$\operatorname*{arg\,min}_{w \in \mathbb{R}^p : \|w\|_0 \leq K} \|(1_n \lambda/2) - Rw\|_2^2$$

Constrained Portfolio Optimization

Portfolio constraints: concentration, diversification, transaction cost management, *etc.*

Example: invest in at most K shares

$$\underset{w \in \mathbb{R}^p : \|w\|_0 \leq K}{\arg\min} \|(1_n \lambda/2) - Rw\|_2^2$$

NP-Hard...

"Statistical Perspective":

$$\underset{w \in \mathbb{R}^p : \|w\|_0 \leq K}{\arg\min} \|(1_n \lambda/2) - Rw\|_2^2 \Leftrightarrow \underset{\beta \in \mathbb{R}^p : \|\beta\|_0 \leq K}{\arg\min} \|y - X\beta\|_2^2$$

Best Subsets and the Lasso

"Statistical Perspective":

$$\underset{w \in \mathbb{R}^p : \|w\|_0 \leq K}{\arg\min} \|(1_n \lambda/2) - Rw\|_2^2 \Leftrightarrow \underset{\beta \in \mathbb{R}^p : \|\beta\|_0 \leq K}{\arg\min} \|y - X\beta\|_2^2$$

- Portfolio Optimization: Find the portfolio which has the best risk/return trade-off using at most K assets
- Statistics: Find the statistical model which fits my data best using at most K variables

"Statistical Perspective":

$$\underset{w \in \mathbb{R}^p : \|w\|_0 \leq K}{\arg\min} \|(1_n \lambda/2) - Rw\|_2^2 \Leftrightarrow \underset{\beta \in \mathbb{R}^p : \|\beta\|_0 \leq K}{\arg\min} \|y - X\beta\|_2^2$$

- Portfolio Optimization: Find the portfolio which has the best risk/return trade-off using at most K assets
- Statistics: Find the statistical model which fits my data best using at most K variables

Impossible to solve exactly for reasonable p

Best Subsets and the Lasso

"Statistical Perspective":

$$\underset{w \in \mathbb{R}^p : \|w\|_0 \leq K}{\arg\min} \|(1_n \lambda/2) - Rw\|_2^2 \Leftrightarrow \underset{\beta \in \mathbb{R}^p : \|\beta\|_0 \leq K}{\arg\min} \|y - X\beta\|_2^2$$

- Portfolio Optimization: Find the portfolio which has the best risk/return trade-off using at most K assets
- Statistics: Find the statistical model which fits my data best using at most K variables

Impossible to solve exactly for reasonable p

State of the art MIO [BKM16] solves $p \approx 1000$ in *hours*

Best Subsets and the Lasso

"Statistical Perspective":

$$\operatorname*{arg\,min}_{w \in \mathbb{R}^p : \|w\|_0 \leq K} \|(1_n \lambda/2) - Rw\|_2^2 \Leftrightarrow \operatorname*{arg\,min}_{\beta \in \mathbb{R}^p : \|\beta\|_0 \leq K} \|y - X\beta\|_2^2$$

- Portfolio Optimization: Find the portfolio which has the best risk/return trade-off using at most K assets
- Statistics: Find the statistical model which fits my data best using at most K variables

Impossible to solve exactly for reasonable p

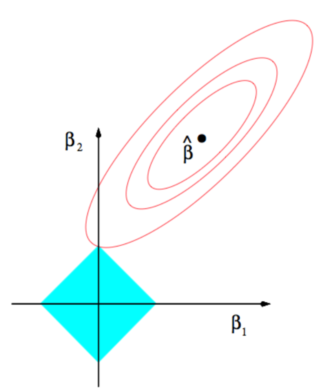
State of the art MIO [BKM16] solves $p \approx 1000$ in *hours*

In statistics, the convex relaxation of this problem is known as the *Lasso* [Tib96] and it has been hugely successful.

Sparsity

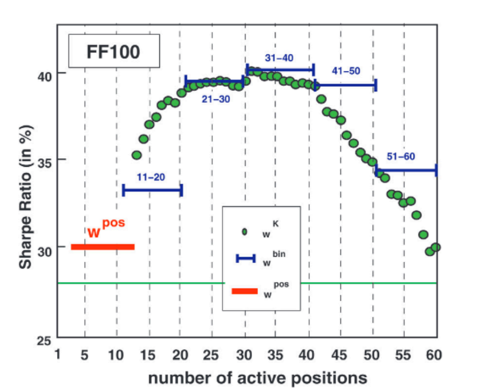
$$\underset{w\in\mathbb{R}^p}{\arg\min} \left\| \underbrace{(1_n\lambda/2)}_{y\in\mathbb{R}^n} - \underbrace{R}_{X\in\mathbb{R}^{n\times p}} \underbrace{w}_{\beta\in\mathbb{R}^p} \right\|_2^2 + \gamma \|w\|_{\ell_1}$$

Tune λ to change risk-return trade-off and γ to change sparsity level



Results

Problem first considered by [BDM$^+$09]



(Figure from [BDM$^+$09])

Why does this work?

Why does this work?

- “Sharp corners” give sparsity [CRPW12]

Why does this work?

- "Sharp corners" give sparsity [CRPW12]
- Lasso can be *prediction consistent* without being *model selection consistent* [GR04, Cha13]

Intuition

Why does this work?

- “Sharp corners” give sparsity [CRPW12]
- Lasso can be *prediction consistent* without being *model selection consistent* [GR04, Cha13]
- Controls over-fitting (degrees of freedom) from search [ZHT07, Tib15]

Further Structure

Statisticians have developed many "Lasso-like" constraints which enforce different structures:

Goal	Relaxed Constraint	Portfolio Analogue
Select at most K variables	Lasso [Tib96]	Invest in at most K assets
Select variables from at most G groups	Group Lasso [YL06]	Invest in at most G asset classes
Select variables from at most B overlapping groups	Group Lasso with Overlap [JOV09]	Trade against at most B brokers
Select variables with at least one in each of D groups	Exclusive Lasso [ZJH10, CA15]	Diversify over at least D asset classes

as well as highly-efficient methods for solving the corresponding optimization problems (*e.g.*, [FHT10]).

See [HTF09], [JWHT13], or [HTW15] for details.

Acknowledgments

On ArXiV soon...

On ArXiV soon...

Funding:

- NSF/DMS 1547433: RTG: Cross-Training in Statistics and Computer Science at Rice University;
- Ken Kennedy Institute for Information Technology (K2I) Computer Science & Engineering Enhancement Fellowship;
- K2I 2015/16 BP Graduate Fellowship.

References

References I

[BDM+09] Joshua Brodie, Ingrid Daubechies, Christine De Mol, Domenico Giannone, and Ignace Loris.
Sparse and stable Markowitz portfolios.
Proceedings of the National Academy of Sciences of the United STates of America, 106(30):12267–12272, 2009.
http://www.pnas.org/content/106/30/12267.full.

[BKM16] Dmitris Bertsimas, Angela King, and Rahul Mazumder.
Best subset selection via a modern optimization lens.
Annals of Statistics, 44(2):813–852, 2016.
https://projecteuclid.org/euclid.aos/1458245736.

[CA15] Frederick Campbell and Genevera I. Allen.
Within group variable selection using the exclusive lasso, 2015.
ArXiv 1505.07517.

[Cha13] Sourav Chatterjee.
Assumptionless consistency of the lasso, 2013.
ArXiv 1303.5817.

[CRPW12] Venkat Chandrasekaran, Benjamin Recht, Pablo A. Parrilo, and Alan S. Wilsky.
The convex geometry of linear inverse problems.
Foundations of Computational Mathematics, 12(6):805–849, 2012.

[FHT10] Jerome H. Friedman, Trevor Hastie, and Rob Tibshirani.
Regularization paths for generalized linear models via coordinate descent.
Journal of Statistical Software, 33:1–22, 2010.
`http://www.jstatsoft.org/v33/i01/paper`; CRAN: `http://cran.r-project.org/web/packages/glmnet/index.html`.

[GR04] Eitan Greenshtein and Ya'acov Ritov.
Persistence in high-dimensional linear predictor selection and the virtue of overparametrization.
Bernoulli, 10(6):971–988, 2004.
https://projecteuclid.org/euclid.bj/1106314846.

[HTF09] Trevor Hastie, Robert Tibshirani, and Jerome Friedman.
The Elements of Statistical Learning: Data Mining, Inference, and Prediction.
Springer Series in Statistics. Springer, 2nd edition, February 2009.
http://statweb.stanford.edu/~tibs/ElemStatLearn/.

[HTW15] Trevor Hastie, Robert Tibshirani, and Martin Wainwright.
Statistical Learning with Sparsity: the Lasso and Generalizations.
Monographs on Statistics and Applied Probability. CRC Press, 1st edition, May 2015.
http://web.stanford.edu/~hastie/StatLearnSparsity/.

[JOV09] Laurent Jacob, Guillaume Obozinski, and Jean-Philippe Vert.
Group lasso with overlap and graph lasso.
In Andrea Danyluk, Leon Bottou, and Michael Littman, editors, *ICML '09: Proceedings of the 26th Annual International Conference on Machine Learning*. ACM, 2009.
http://www.machinelearning.org/archive/icml2009/papers/471.pdf.

References V

[JWHT13] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani.
An Introduction to Statistical Learning.
Springer Series in Statistics. Springer, 1st edition, 2013.
`http://www-bcf.usc.edu/~gareth/ISL/`.

[Tib96] Robert Tibshirani.
Regression shrinkage and selection via the lasso.
Journal of the Royal Statistical Society, Series B (Methodological), 58(1):267–288, 1996.
`http://www.jstor.org/stable/2346178`.

[Tib15] Ryan J. Tibshirani.
Degrees of freedom and model search.
Statistica Sinica, 25:1265–1296, 2015.
DOI:10.5705/ss.2014.147.

[YL06] Ming Yuan and Yi Lin.
Model selection and estimation in regression with grouped variables.
Journal of the Royal Statistical Society, Series B (Methodological), 68(1):49–67, 2006.
doi:10.1111/j.1467-9868.2005.00532.x.

[ZHT07] Hui Zou, Trevor Hastie, and Robert Tibshirani.
On the "degrees of freedom" of the lasso.
Annals of Statistics, 35(5):2173–2192, 2007.
https://projecteuclid.org/euclid.aos/1194461726.

[ZJH10] Yang Zhou, Rong Jin, and Steven C.H. Hoi.
Exclusive lasso for multi-task feature selection.
In Yee Whye Teh and Mike Titterington, editors, *AISTATS 2010: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. JMLR, 2010.
http://proceedings.mlr.press/v9/zhou10a.html.