

Componentwise カーネル学習を用いたポートフォリオ選択

Sparse Markowitz Portfolio Selection with Componentwise Kernel Learning

河原吉伸*1 岡田祥吾*2 武田朗子*3 鷲尾隆*1
 Yoshinobu Kawahara Shogo Okada Akiko Takeda Takashi Washio

*1大阪大学 *2マッキンゼー・アンド・カンパニー *3東京大学
 Osaka University McKinsey & Company The University of Tokyo

In this paper, we propose a sparse Markowitz portfolio selection model via multiple kernel learning (MKL), where each kernel is defined on each investable asset. The optimal solution to this model has weights over both assets and samples (*i.e.*, past returns). Hence, we apply the fused regularization on the weights to incorporate time dependencies into the model. The finally-obtained optimization problem is solvable as linear programming (LP). We showed some empirical examples, where the proposed portfolio model was applied to Nikkei 225 data.

1. はじめに

ポートフォリオ選択問題は、経済分野において古くから議論される中心的課題の一つである。特に、ノーベル経済学賞受賞者でもある H. Markowitz によって 1952 年に提案されたリスク最小化に基づくポートフォリオ選択モデルは、その後の統計分野や最適化分野における数理的議論の基礎ともなる重要なものとして挙げられる [4].

近年になって、Markowitz のポートフォリオ選択モデルに基づき、正則化による疎学習を適用したモデルがいくつか提案されている。J. Brodie らは、Markowitz モデルが一種の線形回帰問題となっている事実から $L1$ 正則化を適用し、疎なポートフォリオ選択を可能とする枠組みを提案している [2]. Y-M. Yen & T-J. Yen は、Brodie らのモデルを Elastic Net を用いた枠組みへ拡張している [9]. また著者らにより、 $L0$ 正則化を適用したインデックスの追従ポートフォリオ・モデルも提案されている [7].

基本的に、これらは線形回帰に基づくモデルである。しかし一般に、経済活動がより一般の (つまり非線形な) 現象であると考えるのは自然な発想であると言える。一方で、非線形モデルにおける特徴 (本場合は銘柄) 選択は原理的に困難な問題である事が知られており、一般的に広く受け入れられる手法が存在しているとは言い難い。

本研究では、各銘柄 (変数) 毎に定義される正定値カーネルを用いた複数カーネル学習としてポートフォリオ選択問題を定式化し、線形モデルよりも表現力のある (非線形) モデルを提案する。このモデルは、銘柄方向のみでなく、時間方向 (過去のリターン) への重み付けに基づいたモデルとなっている。本稿ではこの時間方向の重み (係数) に対する Fused 正則化を行う事により、提案モデルへ時間の近接性を組み込む枠組みについても議論する。最後に、提案するポートフォリオ選択モデルを日経 225 へ適用した検証結果も示す。

本稿の構成は、以下のようなものである。まず 2. では、Markowitz により提案されたリスク最小化に基づくポートフォリオ選択と疎学習の適用による方法について述べる。次に 3. では、各銘柄毎に定義された正定値カーネルを用いた複数カーネル学習に基づくポートフォリオ選択モデルを提案する。4. では更に、時間方向の係数に対する Fused 正則化を行う事により、時間の近

接性をモデルへ組み込む方法について述べる。最後に 5. では、提案モデルを日経 225 データへ適用した検証例を示し、6. において結論を述べる。

2. リスク最小化に基づくポートフォリオ選択

対象とする銘柄数を d とし、ある時刻 t における各銘柄のリターンを $\mathbf{r}_t = [r_{1,t}, \dots, r_{d,t}]^T$ とする。このときポートフォリオ選択問題は、一定期間 T にわたるリターンの観測 ($\mathbf{R} := [\mathbf{r}_1, \dots, \mathbf{r}_T]^T$ と表す) を元に、なんらかの基準の下に各銘柄への重み (ポートフォリオ配分) $\mathbf{w} (\in \mathbb{R}^d)$ を決める問題として定式化される (なおこの際、一般に $\mathbf{w}^T \mathbf{1}_d = 1$ という正規化を課す場合が多い)。

投資における最大のリスクの一つは、投資資産価値の下振れに伴う破産である。そのように Markowitz は、構成したポートフォリオの下での、リターン分散の最小化に基づくポートフォリオ選択モデルを提案した [4].

$$\min_{\mathbf{w} \in \mathbb{R}^d} \mathbf{w}^T \Sigma \mathbf{w} \quad \text{s.t.} \quad \mathbf{w}^T \mathbf{1}_d = 1, \mathbf{w}^T \boldsymbol{\mu} = \rho$$

ここで Σ 及び $\boldsymbol{\mu}$ は、リターン \mathbf{r}_t の分散共分散行列と平均ベクトルである。また $\rho \geq 0$ は要求リターンであり事前に設定する必要がある。通常 Σ と $\boldsymbol{\mu}$ は、データを用いて推定された経験値 $\hat{\Sigma} = E[\mathbf{r}_t \mathbf{r}_t^T]$, $\hat{\boldsymbol{\mu}} = E[\mathbf{r}_t]$ で置き換えられる。このときこの最適化問題は、簡単な計算により、次式のような線形回帰問題へと変形される。

$$\min_{\mathbf{w} \in \mathbb{R}^d} \frac{1}{T} \sum_{t=1}^T \left\| \rho - \mathbf{w}^T \mathbf{r}_t \right\|_2^2 \quad \text{s.t.} \quad \mathbf{w}^T \mathbf{1}_d = 1, \mathbf{w}^T \hat{\boldsymbol{\mu}} = \rho \quad (1)$$

上式から分かるように、この投資スタンスは、線形回帰モデル

$$f(\mathbf{r}) = \mathbf{w}^T \mathbf{r}$$

に基づいていると言える。近年この回帰による定式化に基づき、正則化学習を適用したポートフォリオ選択モデルが提案されており、その有用性が報告されている [2, 3, 7, 9].

3. 複数カーネル学習による定式化

一般の (非線形な) 回帰モデルに基づくポートフォリオ選択モデルについて考える。そのためまず、ここでは投資モデルが、

各銘柄毎に定義される関数 f_i を用いた Componentwise な関数であると仮定する.

$$f(\mathbf{r}) = \sum_{i=1}^d f_i(r_i) \quad (2)$$

ただし, 各 f_i ($i = 1, \dots, d$) は正定値カーネル $k_i : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ により得られる再生核ヒルベルト空間 (RKHS) \mathcal{F}_i 上に定義される関数であるとする. つまり, \mathcal{F}_i 内の関数の基底を ϕ_i とすると, 各基底への重みを ξ_i として f_i は $f_i(r) = \xi_i^\top \phi_i(r)$ のように表される. なお Componentwise なカーネル学習モデルは, 統計的に有用な性質を持つ事が知られている [6].

今, 各変数 (銘柄) 上の疎性を誘導するために, (1) 式の線形回帰の評価関数に対応して, 各 f_i の係数をグループとするような, グループ正則化項を持つ次の評価関数を考える.

$$(R(f_1, \dots, f_d) := \frac{1}{T} \sum_{t=1}^T \left\| \rho - \sum_{i=1}^d f_i(r_{i,t}) \right\|_2^2 + \frac{\lambda}{2} \left(\sum_{i=1}^d \|\xi_i\|_{\mathcal{F}_i} \right)^2$$

ただし, $\|\cdot\|_{\mathcal{F}_i}$ は RKHS \mathcal{F}_i 上のノルムである. この正則化項の変分形は, $\sum_{i=1}^d \eta_i = 1$ となる $\boldsymbol{\eta} \in \mathbb{R}_+^d$ を用いて, 次式のようになる.

$$\left(\sum_{i=1}^d \|\xi_i\|_{\mathcal{F}_i} \right)^2 = \sum_{i=1}^d \frac{\|\xi_i\|_{\mathcal{F}_i}^2}{\eta_i}$$

従って, $f_i(r) = \xi_i^\top \phi_i(r)$ である事に注意すると, 次式のようなポートフォリオ選択モデルが得られる.

$$\begin{aligned} \min_{f_1, \dots, f_d, \boldsymbol{\eta} \in \Delta_d} & \frac{1}{T} \sum_{t=1}^T \left\| \rho - \sum_{i=1}^d \eta_i^{1/2} \xi_i^\top \phi_i(r_{i,t}) \right\|_2^2 + \frac{\lambda}{2} \sum_{i=1}^d \|\xi_i\|_{\mathcal{F}_i}^2 \\ \text{s.t.} & \sum_{i=1}^d \eta_i^{1/2} \xi_i^\top \frac{\sum_{t=1}^T \phi_i(r_{i,t})}{T} = \rho \end{aligned} \quad (3)$$

ただし, $\tilde{\xi}_i := \xi_i / \eta_i^{1/2}$, $\Delta_d := \{\boldsymbol{\eta} \in \mathbb{R}_+^d \mid \sum_{i=1}^d \eta_i = 1\}$ であり, 制約条件は (1) 式における $\mathbf{w}^\top \boldsymbol{\mu} = \rho$ に対応する. なお後に見るように, 各 η_i はポートフォリオ配分そのものであり, η_i 上の制約 $\boldsymbol{\eta} \in \Delta_d$ は, 式 (1) 中の制約 $\mathbf{w}^\top \mathbf{1}_d = 1$ に対応すると言える. ここで, スラック変数

$$\delta_t = \rho - \sum_{i=1}^d \eta_i^{1/2} \tilde{\xi}_i^\top \phi_i(r_{i,t}) \quad (t = 1, \dots, T)$$

を導入すると, 最適化問題 (3) のラグランジュ関数は

$$\begin{aligned} L = & \frac{1}{\lambda T} \sum_{t=1}^T \delta_t^2 + \frac{1}{2} \sum_{i=1}^d \tilde{\xi}_i^\top \tilde{\xi}_i + \sum_{t=1}^T \alpha_t \left(\sum_{i=1}^d \eta_i^{1/2} \tilde{\xi}_i^\top \phi_i(r_{i,t}) + \delta_t - \rho \right) \\ & + \beta \left(\sum_{i=1}^d \eta_i^{1/2} \tilde{\xi}_i^\top \sum_{t=1}^T \phi_i(r_{i,t}) - T\rho \right) \end{aligned}$$

と表される. ここで, α_t ($t = 1, \dots, T$), β はラグランジュ係数である. 従って, Karush-Kuhn-Tucker 条件の一つとして,

$$\tilde{\xi}_i = \eta_i^{1/2} \sum_{t=1}^T (\alpha_t + \beta) \phi_i(r_{i,t}) \quad (i = 1, \dots, d) \quad (4)$$

が得られる. その結果, 式 (2) は次式のように表される.

$$f(\mathbf{r}) = \sum_{i=1}^d \eta_i \sum_{t=1}^T \tilde{\alpha}_t k_i(r_{i,t}, r_i) \quad (5)$$

ただし $\sum_{i=1}^d \eta_i = 1$ かつ $\eta_i \geq 0$ であり, また $\tilde{\alpha}_t := \alpha_t + \beta$ ($t = 1, \dots, T$) である. また, $k_i(\bullet, r_i) := \langle \phi_i(\bullet), \phi_i(r_i) \rangle$ ($i = 1, \dots, d$) はカーネル関数である. なお上述のように, ポートフォリオ選択モデル (5) において η_i が投資配分にあたる.

ポートフォリオ選択モデル (3) は, 各変数 (銘柄) 上に定義された正定値カーネルを用いた複数カーネル学習 (MKL) [1] であるとも捉えられる. そのためこの最適化計算自体は, 一般の MKL 同様, 最急降下法などを用いて行う事ができる. 今, KKT 条件 (4) を元の最適化問題 (3) へ代入する事により, 次式が得られる.

$$\min_{\tilde{\boldsymbol{\alpha}}, \boldsymbol{\eta} \in \Delta_d} \frac{1}{T} \sum_{t=1}^T \left\| \rho - \sum_{i=1}^d \eta_i \sum_{t=1}^T \tilde{\alpha}_t k_i(r_{t,i}, r_i) \right\|_2^2 + \frac{\lambda}{2} \sum_{i=1}^d \eta_i \tilde{\boldsymbol{\alpha}}^\top \mathbf{K}_i \tilde{\boldsymbol{\alpha}} \quad (6)$$

ただし $\tilde{\boldsymbol{\alpha}}$ と $\boldsymbol{\eta}$ は各々 $\tilde{\alpha}_t$ と η_i を並べて得られるベクトルであり, また \mathbf{K}_i ($i = 1, \dots, d$) は k_i に関するカーネル行列である. この問題の最適化変数である $\tilde{\boldsymbol{\alpha}}$ と $\boldsymbol{\eta}$ は互いに独立ではないが, 一般に MKL の最適化計算では, これらを交互に反復的に最適化するという手順をとる. つまり, 正定値行列 $\tilde{\mathbf{K}}$ に対して

$$G(\tilde{\mathbf{K}}) := \min_{\tilde{\boldsymbol{\alpha}}} \frac{1}{T} \left\| \rho \mathbf{1}_d - \tilde{\mathbf{K}} \tilde{\boldsymbol{\alpha}} \right\|_2^2 + \frac{\lambda}{2} \tilde{\boldsymbol{\alpha}}^\top \tilde{\mathbf{K}} \tilde{\boldsymbol{\alpha}} \quad (7)$$

のように表すと, 基本的には次の (1) と (2) を繰り返す.

$$(1) G \left(\sum_{i=1}^d \eta_i \mathbf{K}_i \right) \text{ の計算} \quad (2) \min_{\boldsymbol{\eta} \in \Delta_d} G \left(\sum_{i=1}^d \eta_i \mathbf{K}_i \right)$$

なお後の数値検証においては, このような手順に基づくアルゴリズムの一つである SimpleMKL [5] に基づく実装をベースにして実験を行っている.

4. 時間構造正則化の導入

一般に, 時間的に近くで行われる経済活動は, 離れているそれよりも, 互いに関連性が高いと想像するのは難くない. ここでは, このような時間の近接性を Fused 正則化 [8] に基づきモデルに取り込む方法について述べる. なお本節で扱う構造正則化は, 各時刻における関数への重みでもある $\tilde{\alpha}_t$ のみに関連するため, それは則ち $\tilde{\boldsymbol{\alpha}}$ の最適化部分 (つまり式 (7)) のみに影響を与えるという事に注意する.

まず, 最適化問題が一種の L_2 正則化二乗回帰問題となっているため, サポートベクトル回帰と同様に, ヒンジ損失に基づく再定式化を行う. このためまず, 各 \mathbf{r}_t の式 (5) に基づく予測値を $f_t := f(\mathbf{r}_t)$ と表し, 次のスラック変数を導入する.

$$\begin{aligned} \gamma_t &= \begin{cases} 0 & -(f_t - \rho) < \epsilon \\ |f_t - \rho| - \epsilon & \text{otherwise} \end{cases} \\ \hat{\gamma}_t &= \begin{cases} 0 & f_t - \rho \\ |f_t - \rho| - \epsilon & \text{otherwise} \end{cases} \end{aligned}$$

なお $\epsilon > 0$ であり, γ_t と $\hat{\gamma}_t$ は次式の制約条件を満たす.

$$f_t - \epsilon - \hat{\gamma}_t \leq \rho \leq f_t + \epsilon + \gamma_t \quad (t = 1, \dots, T) \quad (8)$$

このとき, $\tilde{\boldsymbol{\alpha}}$ に関する最適化部分は次式のように書ける.

$$\min_{\tilde{\boldsymbol{\alpha}}, \boldsymbol{\gamma}, \hat{\boldsymbol{\gamma}}} \frac{1}{T} \sum_{t=1}^T (\gamma_t + \hat{\gamma}_t) + \frac{\lambda}{2} \sum_{i=1}^d \eta_i \tilde{\boldsymbol{\alpha}}^\top \mathbf{K}_i \tilde{\boldsymbol{\alpha}} \quad (9)$$

ただし拘束条件は、元の η に関するものに加えて、式 (8) も必要となる。

さらに $\tilde{\alpha}$ に疎性を得るために、最適化問題 (9) の $\tilde{\alpha}$ に関する L_2 正則化項を、 L_1 正則化項へと置き換える。このような操作は、 L_1 ノルム・サポートベクトル器などと同様のものであり、学習性能の向上へとつながる事も知られている [10, 8]. この上で、 $\tilde{\alpha}_t$ に関する Fused 正則化項

$$\sum_{t=1}^{T-1} |\alpha_{t+1} - \alpha_t| \leq \lambda_2$$

を導入すると、最終的に得られる最適化は次式のようにになる。

$$\begin{aligned} \min_{\tilde{\alpha}, \gamma, \hat{\gamma}} \quad & \frac{1}{T} \sum_{t=1}^T (\gamma_t + \hat{\gamma}_t) \\ \text{s.t.} \quad & |\tilde{\alpha}| < \lambda_1, \quad \sum_{t=1}^{T-1} |\alpha_{t+1} - \alpha_t| \leq \lambda_2, \\ & f_t - \epsilon - \hat{\gamma}_t \leq \rho \leq f_t + \epsilon + \gamma_t \quad (t = 1, \dots, T) \end{aligned} \quad (10)$$

ただし、 $\lambda_1 \geq 0$ と $\lambda_2 \geq 0$ は正則化パラメータである。この問題は、 $\tilde{\alpha}_t = \tilde{\alpha}_t^+ - \tilde{\alpha}_t^-$ ($\tilde{\alpha}_t^+, \tilde{\alpha}_t^- \geq 0$), $\theta_t = \tilde{\alpha}_t - \tilde{\alpha}_{t-1}$, 及び $\theta_t = \theta_t^+ - \theta_t^-$ ($\theta_t^+, \theta_t^- \geq 0$) と変数を導入する事により、線形計画問題 (LP) として高速に解く事ができる [8].

5. 数値例

上述のポートフォリオ選択モデルの性能を検証するため、日経 225 データを用いた実験を行った結果を示す。ここでは、2000 年 1 月から 2010 年 12 月までの日経 225 に継続して含まれる 187 銘柄を対象とした。

以下では、上記のデータを用いて次のように行った実験の結果を示す。まず上記のデータから、各時刻・各銘柄のリターン $r_{t,i}$ は各々の対前週最終値の割合 (当週終値/前週終値) として得られる。データを連続する 2 年分ずつを組とし、1 年分を訓練データ、残り 1 年分を評価データとして、10 つのデータ組を作成し実験を行った。また要求リターン ρ としては、ここではそれぞれの訓練期間内での平均収支 ($\sum_{i=1}^d \sum_{t=1}^T r_{t,i} / T$) とした。またここでは、予備実験より適切と判断された $\lambda = 10$, $\lambda_1 = 15$, 及び $\lambda_2 = 20$ を用いて実験を行った。

まず、得られたポートフォリオの性能を評価するため、金融商品の評価に一般に広く用いられるシャープ・レシオ (SR) を検証した。シャープ・レシオは、どの程度のリスクをおって、どの程度のリターンを得ることができるかを表す指標であり、一般的には次のように定義される。

$$SR = \frac{\text{年率の収益率 (Return)} - \text{リスクフリーレート}}{\text{標準偏差 (Risk)}}$$

リスクフリーレートは、理論的にリスクがゼロか極小のリスクフリー商品から得る事のできる利回りのことを言い、具体的には元利金の支払いが保障された預貯金や国債などがある。現在の日本ではリスクフリーレートが限りなくゼロに近いので、今回は無視する。また Return と Risk は、本稿の記述を用いると各々次のように定義される。

$$\begin{aligned} \text{Return} &= \prod_{t=1}^T (1 + f(\mathbf{r}_t)) - 1 \\ \text{Risk} &= \sqrt{\frac{1}{T} \sum_{t=1}^T \left(f(\mathbf{r}_t) - \frac{1}{T} \sum_{t=1}^T f(\mathbf{r}_t) \right)^2} \end{aligned}$$

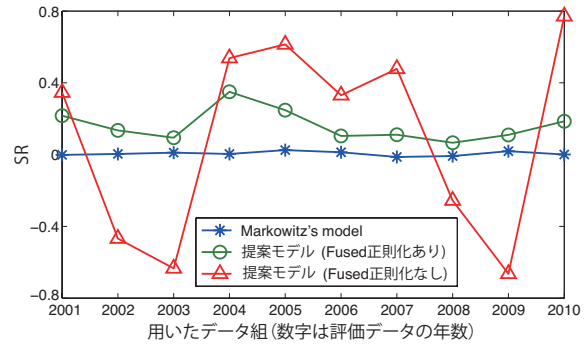


図 1: 各ポートフォリオ選択モデルにより得られた各データ組のシャープレシオ。

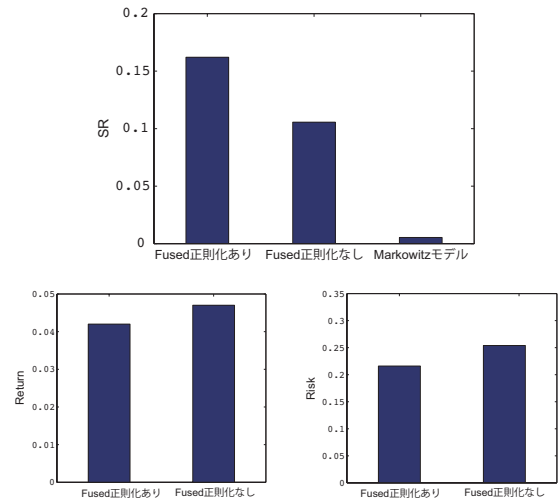


図 2: 各ポートフォリオ選択モデルにより得られたシャープレシオ平均値 (上) とその内訳 (下)。

まず図 1 は、提案モデル (10) と、このモデルから Fused 正則化項を除いたもの、及び Markowitz モデルによる、各データ組の SR を示したものである。Fused 正則化を導入する事により、ポートフォリオの各年における変動が著しく軽減されていることが分かる。また図 2 は、これらの平均値と、その Return と Risk の内訳を表したものである。図からも分かるように、Fused 正則化項を加える事で、Risk (投資に伴う資産変動) を軽減したポートフォリオが得られている事が分かる。

なお図 3 は、推定された α_t の例として、2002 年のデータを訓練データとした場合を図示したものである。12~15 週目辺りは多くの企業の決算期にあたるため、このモデルはその付近への重みが大きくなっており、解釈性の高いモデルが得られていることが分かる。

6. まとめ

本稿では、銘柄毎に定義された正定値カーネルを用いた複数カーネル学習に基づくポートフォリオ選択モデルを提案した。更に、このモデルが銘柄方向に加えてサンプル方向 (過去のリターン) への重み係数を持つため、Fused 正則化を用いて時間の近接性をモデルに取り込む方法について述べた。日経 225 データを用いた実験により、提案するモデルのパフォーマンスについても検証を行なった。

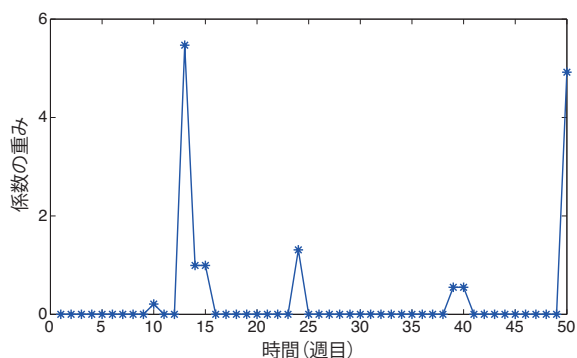


図 3: α_t ($t = 1, \dots, 50$) の分布. 各 α_t は, 訓練期間の各週のリターンの関数に対する重みとなっている.

参考文献

- [1] F. Bach, G. Lanchriet, and M. Jordan. Multiple kernel learning, conic duality, and the SMO algorithm. In *Proc. of the 21st Int'l Conf. on Machine Learning*, pages 41–48, 2004.
- [2] J. Brodie, I. Daubechies, C. De Mol, D. Giannone, and I. Loris. Sparse and stable markowitz portfolios. *Proceedings of the National Academy of Sciences of the USA (PNAS)*, 106(30):12267–12272, 2009.
- [3] J. Fan, J. Zhang, and K. Yu. Vast portfolio selection with gross-exposure constraints. *Journal of the American Statistical Association*, 107(498):592–606, 2012.
- [4] H. Markowitz. Portfolio selection. *The Journal of Finance*, 7:77–91, 1952.
- [5] A. Rakotomamonjy, F. Bach, S. Canu, and Y. Grandvalet. Simplemkl. *The Journal of Machine Learning Research*, 9:2491–2521, 2008.
- [6] J.A.K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle. *Least Squares Support Vector Machines*. World Scientific Pub. Co. Inc., 2003.
- [7] A. Takeda, M. Niranjana, J. Goto, and Y. Kawahara. Simultaneous pursuit of out-of-sample performance and sparsity in tracking portfolio. *Computational Management Science*, 10(1):21–49, 2013.
- [8] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight. Sparsity and smoothness via the fused Lasso. *Journal of the Royal Statistical Society: Series B*, 67(1):91–108, 2005.
- [9] Y. Yen and T. Yen. Solving norm constrained portfolio optimization via coordinate-wise descent algorithms. *Computational Statistics & Data Analysis*, in press.
- [10] J. Zhu, S. Rosset, T. Hastie, and R. Tibshirani. 1-norm support vector machines. In *Advances in Neural Information Processing Systems 16*, pages 49–56. MIT Press, 2004.