

新しい評価関数 $iNorm$ \ アイノルム¹ の紹介

— l^2 ノルムと l^∞ ノルムの間を任意に補間するノルム —

新日鉄住金ソリューションズ株式会社 エンベデッド・ユビキタスシステムセンター
貝塚 洋, 川崎智基

1. $iNorm$ とは？

1.1. 記号の定義

- R^n …… n 次元ベクトル空間

- $\|x\| = \sqrt{\sum_{i=1}^n x_i^2}$ ($x = [x_1 \ x_2 \ \dots \ x_n]^T \in R^n$) …… l^2 ノルム (二乗ノルム)

1.2. ニーズ

世の中には,

『 i 番目の“誤差 or リスク”を

$$\|S_i x + b_i\|^2 \quad (i = 1, 2, \dots, N)$$

where

決定すべき操作変数 $x \in R^m$

所与のデータ $b_i \in R^n, \quad S_i \in R^{n \times m}$

とするとき, 操作変数 x を使って, N 個の“誤差 or リスク”を何らかの意味で最小化したい』というニーズがある. このとき, 下記の2種類の評価関数が採用されることが多い.

$$\text{平均}l^2\text{ノルム(平均二乗ノルム)} \dots \frac{1}{N} \sum_{i=1}^N \|S_i x + b_i\|^2$$

$$l^\infty\text{ノルム(最大ノルム)} \dots \max_{1 \leq i \leq N} (\|S_i x + b_i\|^2)$$

これらの評価関数の最適化における利害得失を簡潔にまとめると, 下表のようになる.

	全体的な最小化	ピーク値の抑制 (平準化)
平均 l^2 ノルム	○	×
l^∞ ノルム	×	○

どちらを選択するかは問題によって変化する. しかし, 問題に最適な評価関数 (ここでは“ノルム”と呼ぶ) は両者のノルムの中間に存在すると思われる. そこで, 下記のようなノルムを設計できれば, たった一つのパラメータ λ を調整することで最適ノルムを選択できるので, 効率的なはずである.

[$iNorm$ のイメージ]

$$iNorm(z, \lambda) \quad (0 \leq \lambda \leq 1)$$

where

$$iNorm(z, 0) = \frac{1}{N} \sum_{i=1}^N z_i^2, \quad \lim_{\lambda \rightarrow 1} iNorm(z, \lambda) \rightarrow \max_{1 \leq i \leq N} (z_i^2)$$

¹ 商標登録出願済み. また, $iNorm$ の実態は非公開としているので, 本論分でもその関数形は記載しない.

(注意) 最適化に $iNorm$ を使用する場合には, $z_i = \|S_i x + b_i\|$ として利用する. しかし $z_i = \|S_i x + b_i\|$ にこだわる必要はなく, $iNorm$ は“誤差 or リスク”の評価関数 (“誤差 or リスク”を定量化するための入れ物) であるので, 一般的に $z_i =$ “ i 番目の誤差 or リスク” として利用すればよい. □

このとき, $iNorm$ はつぎの2つの性質をもつことが望まれる.

(1) λ の調整を容易にするためには, $iNorm(z, \lambda)$ は $\lambda \rightarrow 1$ のとき “ z に関して一様に” $\max_{1 \leq i \leq N} (\|z_i\|^2)$ に近づくことが望ましい. □

(2) 最適化問題が効率的・安定的に解ける必要があるので, $iNorm \left(z = \begin{bmatrix} \vdots \\ z_i = \|S_i x + b_i\| \\ \vdots \end{bmatrix}, \lambda \right)$ としたとき, 当該 $iNorm$

が操作変数である x に関して凸関数であることが望ましい. しかし, 『 $iNorm$ は“誤差 or リスク”の評価関数 (“誤差 or リスク”を定量化するための入れ物)』という観点から, 『 $iNorm(z, \lambda)$ は z に関して凸関数』であることがまずは大切である. □

1.3. l^∞ ノルム類似度 (最大ノルム類似度)

$iNorm(z, \lambda)$ は z に関して凸関数であれば, $iNorm(z, \lambda)$ の等ノルム曲面は N 次元ユークリッド空間で凸形状となる. そこで, “原点から一番遠い N 次元超立方体 (l^∞ ノルムの等ノルム曲面である) の頂点” P_1 と原点を結ぶ直線が, $iNorm(z, \lambda)$ の等ノルム曲面と交わる交点を P_2 とする. このとき,

$$iNorm(z, \lambda) \text{ の } l^\infty \text{ ノルム類似度 (最大ノルム類似度)} := \frac{\overline{OP_2}}{\overline{OP_1}} \leq 1$$

と定義する. このとき $iNorm(z, \lambda)$ が z に関して凸関数であるため, 『 $iNorm(z, \lambda)$ の l^∞ ノルム類似度が 1 に収束すれば, $iNorm(z, \lambda)$ は “ z に関して一様に” l^∞ ノルムに収束する』と言える.

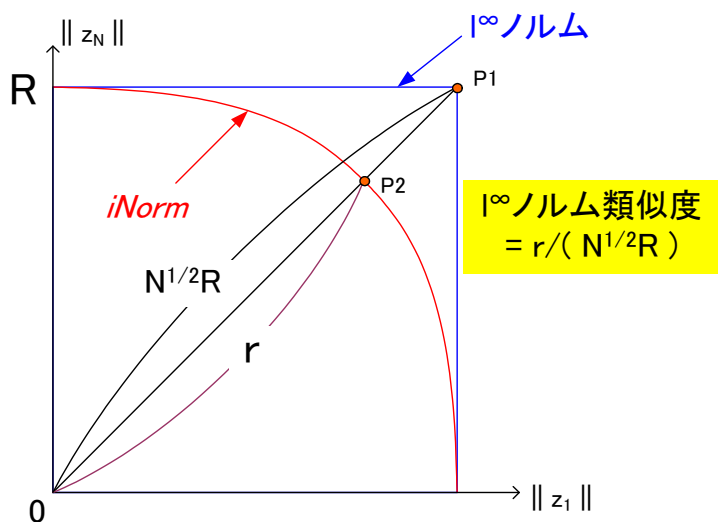


図 1.1 等ノルム曲面

1.4. *iNorm* の構造

iNorm の構造を示す.

[*iNorm* の構造]

$$iNorm_p(z, \lambda) := \sum_{i=1}^N w_i^{(p)}(z, \lambda) z_i^2 \quad (z \in R^N)$$

where

$$\text{重みであるための条件} \dots \quad w_i^{(p)}(z, \lambda) \geq 0, \quad \sum_{i=1}^N w_i^{(p)}(z, \lambda) \equiv 1$$

$$\text{スケール不変性} \dots \quad w_i^{(p)}(\alpha z, \lambda) = w_i^{(p)}(z, \lambda) \quad (\alpha > 0)$$

$$iNorm \text{ の } l^\infty \text{ ノルム類似度の調整パラメータ} \dots \quad 0 \leq \lambda \leq 1$$

$$iNorm_p(z, 0) = \frac{1}{N} \sum_{i=1}^N z_i^2, \quad iNorm_p(z, \lambda = 1) \text{ の } l^\infty \text{ ノルム類似度} \geq \frac{p}{1000} \quad (\forall N \geq 2)$$

$iNorm_p(z, \lambda)$ は z に関して凸関数

□

上記のような *iNorm* を構成することは至難の業である。いくつかのコメントをしておく。

(1) *iNorm* は一つではなく多くの種類が存在する。しかしその多くは、 N が無限大になったときに l^∞ ノルム類似度が 0 に単調収束する（この場合には上記の p は 0 となる）。すなわち N が大きい場合には平準化能力が消滅し、*iNorm* の意味を失ってしまう。 p が正とできる *iNorm* の構成は難しい。□

(2) *iNorm* の性質の中で証明が最も困難なものは、『*iNorm* が z に関して凸関数であること』である。今回われわれが構成に成功した最強の l^∞ ノルム類似度を有する $iNorm_{816}(z, \lambda)$ は、いくつかの数値的な検証と部分的な解析的状况に関して凸性を検証したのみで、解析的証明は今後の課題である。しかしその証明は極めて難しい。 l^∞ ノルム類似度を下げれば証明できる場合もあるが、 l^∞ ノルム類似度を下げるとは実用上意味がない。

z に関して非凸性が大きい *iNorm* もどき（凸関数性以外は *iNorm* の要件を満足するノルム）は、最適解の求解が安定せず、また、最適化計算を強制的に止めて求めた暫定解の精度も $iNorm_{816}(z, \lambda)$ が求める最適解に比べて低い（図 3.1, 図 3.5 に実例を示す）。□

1.5. *iNorm* が決定する値

iNorm の典型的な利用シーンは、下記の最適化問題を解くことである。

[解くべき最適化問題]

λ ($0 \leq \lambda \leq 1$) を調整しながら、下記を解く。

$$\min_x iNorm_p(z_1 = \|S_1 x + b_1\|, \dots, z_N = \|S_N x + b_N\|, \lambda)$$

subject to

$$x \in \text{閉凸領域} \subset R^m$$

□

最適化完了後、下記の2種類の値が定まる。

(1) 最適解: x_{opt} □

(2) 最適重み: $w_i^{(p)}(z_1 = \|S_1 x_{opt} + b_1\|, \dots, z_N = \|S_N x_{opt} + b_N\|, \lambda)$ ($i=1, 2, \dots, N$) □

最適重みは、第*i*番目制御量（誤差 or リスク）の『操作量 x から見た“制御困難さ”』あるいは『投資スキーム x から見た“リスク”』と解釈することができる。このような見方が有意味である適応先があれば、*iNorm* が決定する上記2種類の値は共に有益な情報となる。

2. $iNorm_{816}(z, \lambda)$ (関数形は非公開)

今回構成に成功した $iNorm_{816}(z, \lambda=1)$ の l^∞ ノルム類似度は、

$$\sqrt{\frac{1}{3} \left(\frac{1}{\sqrt{N}} + 2 \right)} \geq \sqrt{\frac{2}{3}} = 0.816496\dots$$

となり、 N に無関係に常に 81.6%以上の類似度を保持する。

表 2-1 平均 l^2 ノルム $\frac{1}{N} \sum_{i=1}^N z_i^2$ と

$iNorm_{816}(z, \lambda=1)$ の l^∞ ノルム類似度

N	平均 l^2 ノルム	$iNorm_{816}(z, \lambda=1)$
2	0.7071	0.9499
4	0.5000	0.9129
8	0.3536	0.8857
16	0.2500	0.8660
32	0.1768	0.8518
64	0.1250	0.8416
128	0.0884	0.8343
256	0.0625	0.8292
512	0.0442	0.8255
1024	0.0313	0.8229

↓
 ∞

↓
0

↓
0.8164

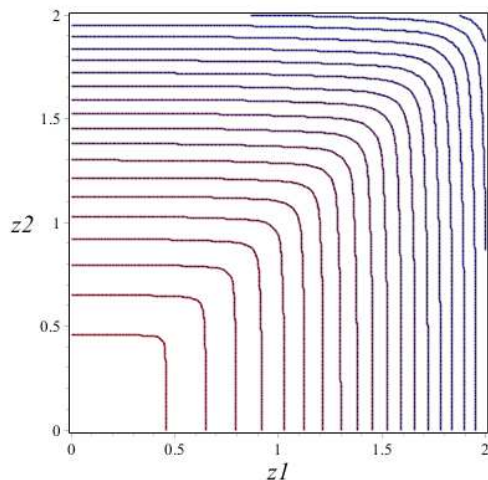


図 2.1 $iNorm_{816}(z, \lambda=1)$ ($N=2$) の等ノルム曲面

(l^∞ ノルム類似度 = 0.950)

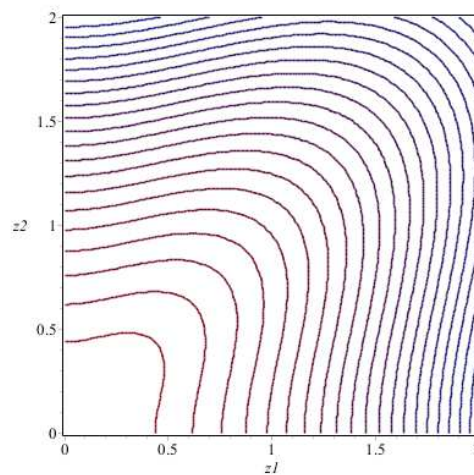


図 3.1 $Q_1(z)$ ($N=2$) の等ノルム曲面

3. 実験結果

3.1. 実験内容と結果概要

下記の最適化問題を実験に使用した.

[最適化問題 A]

$$\begin{aligned} \min_x f(z_1 = \|S_1 x + b_1\|, \dots, z_N = \|S_N x + b_N\|) \\ \text{subject to} \\ l \leq A(S_i x + b_i) \leq u \quad (i=1, 2, \dots, N=512) \\ x \in R^m = R^{512}, \quad S_i \in R^{100 \times 512}, \quad b_i \in R^{100}, \quad A \in R^{50 \times 100}, \quad l, u \in R^{50} \end{aligned}$$

□

(注意) 制約なしの $m = N = 1024$ の場合も, $f = iNorm_{816}(z, \lambda)$ として最適解が求まることを確認した. □

$f = iNorm_{816}(z, \lambda)$ の場合に, 数理計画法パッケージ ニューオプト²の準ニュートン法 (2 階微係数を準ニュートン法によって求める内点法) を利用して, 最適化問題 A を,
 (ケースその 1) 制約条件がない場合
 (ケースその 2) 制約条件がある場合
 について解いた. すべてのケースで, 安定して最適解を求めることができた. また λ が 1 に近づくほど, 平準化能力が強くなることも確かめられた.

さらに, $iNorm$ が “z に関する凸関数性” にこだわる理由を例示するために, 下記の z に関する非凸関数 (図 3.1 に $Q_1(z)$ ($N=2$) の等ノルム曲面を示しておく)

$$Q_1(z) := \sum_{i=1}^N \left(\left(\frac{z_i^2}{\sum_{j=1}^N z_j^2} \right) \times z_i^2 \right) = \frac{\sum_{i=1}^N z_i^4}{\sum_{j=1}^N z_j^2}$$

に対して, $f = Q_1(z)$ として最適化問題 A (制約条件がない場合) を解いた. $Q_1(z)$ の重みは

$$\frac{z_i^2}{\sum_{j=1}^N z_j^2} = \begin{cases} \text{大きくなる} & (z_i \text{ が相対的に大きい場合}) \\ \text{小さくなる} & (z_i \text{ が相対的に小さい場合}) \end{cases}$$

のように作用するので $iNorm$ が必要とする “平準化機構” を備えている. また, 1.4 節の [$iNorm$ の構造] に記載した $iNorm$ の性質の内, “z に関する凸関数性” 以外はすべて適合する評価関数 (凸関数ではないので l^∞ ノルム類似度は定義できない) である. しかし, $f = Q_1(z)$ では最適化計算が収束せず, 最適化計算を強制的に中断して求めた暫定解の平準化能力も $f = iNorm_{816}(z, 1)$ の場合の最適解より低かった.

² NTT データ数理システム社製

3.2. 実験結果詳細

3.2.1. $f = iNorm_{816}(z, \lambda)$ とした場合

制約条件がない場合もある場合もある場合も安定的に最適解を求解できた。求解状況を下表に示す。

表 3.1 計算時間などの計数値

(制約条件がない場合)

λ	0	0.25	0.5	0.75	1
評価関数値	0.1025	0.7476	1.4043	2.0762	2.5402
内点法の反復回数	19	21	22	31	41
処理時間(sec)	283	297	306	370	457

(制約条件がある場合)

λ	0	0.25	0.5	0.75	1
評価関数値	1.3777	1.8769	2.1507	2.3214	2.6790
内点法の反復回数	61	60	60	63	70
処理時間(sec)	735	724	725	754	822

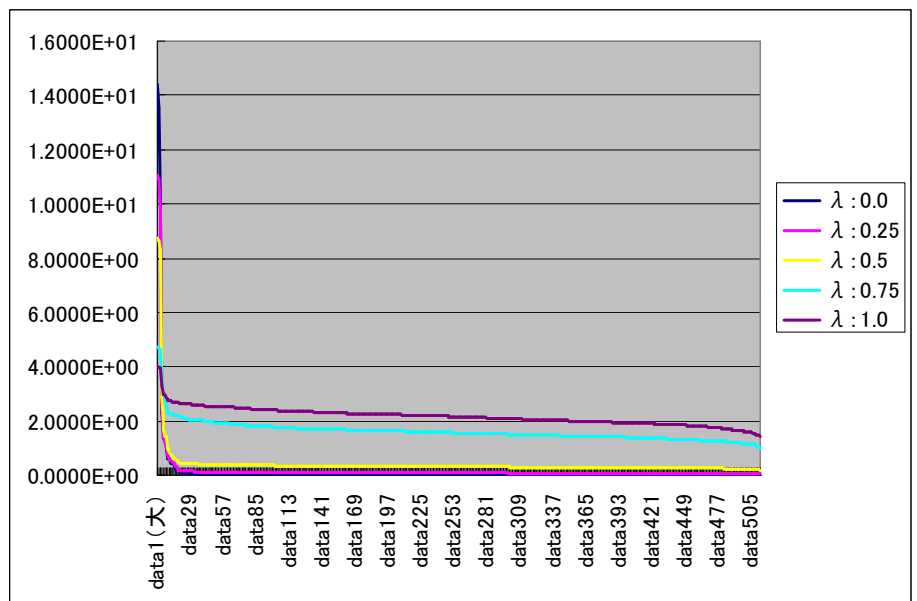
つぎに、 λ を調整することによって平準化能力を連続的に可変にできることを示す。

図 3.2 最適解の平準化能力

(a) 制約条件なしの場合

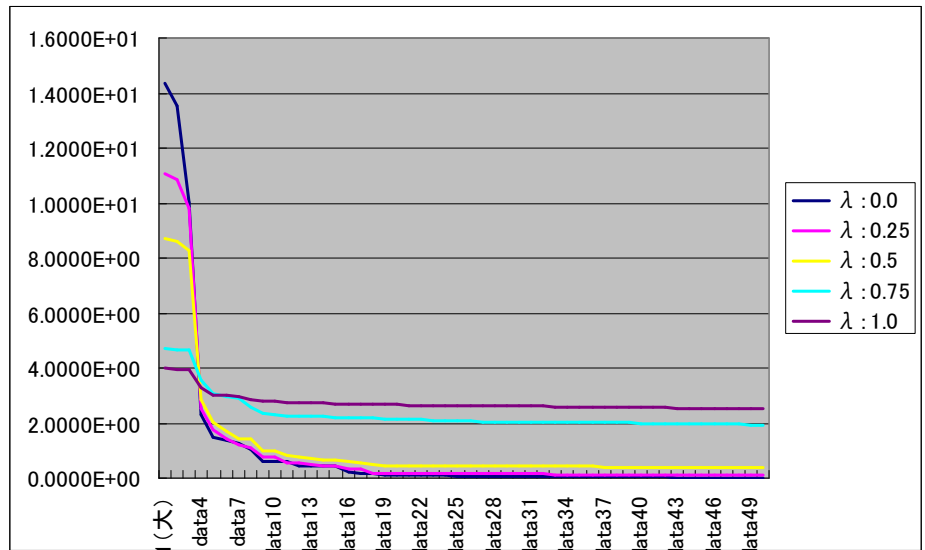
- 縦軸は地点ごとの $\|S_i x_{opt} + b_i\|^2$
- 横軸は 512 地点
($\|S_i x_{opt} + b_i\|^2$ の大きい物順)
- λ が大きいほど平準化されている。

※ 縦軸は(a),(b),(c)共通。



(b) 制約条件なしの場合

- (a)の上位 50 地点のみを表示
- λ が大きいほど平準化されている。



(c) 制約条件ありの場合

- $\|S_i x_{opt} + b_i\|^2$ の大きさの上位 50 地点のみを $\|S_i x_{opt} + b_i\|^2$ の大きい物順に表示
- λ が大きいほど平準化されている。

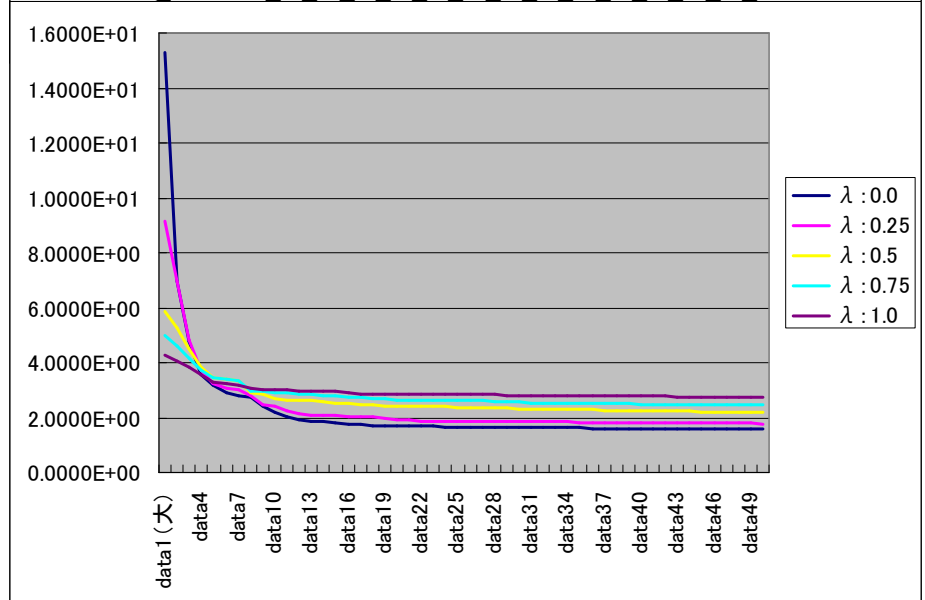


図 3.3 最適化計算中の

$\left\{ \|S_i x_{opt} + b_i\|^2 \right\}$ の分散

の変化

- 制約条件なし
- 縦軸は $\left\{ \|S_i x_{opt} + b_i\|^2 \right\}$ の分散
- 横軸は解 x の変化回数 (内点法の反復回数とほぼ同義)
- λ が大きいほど分散は小さい。

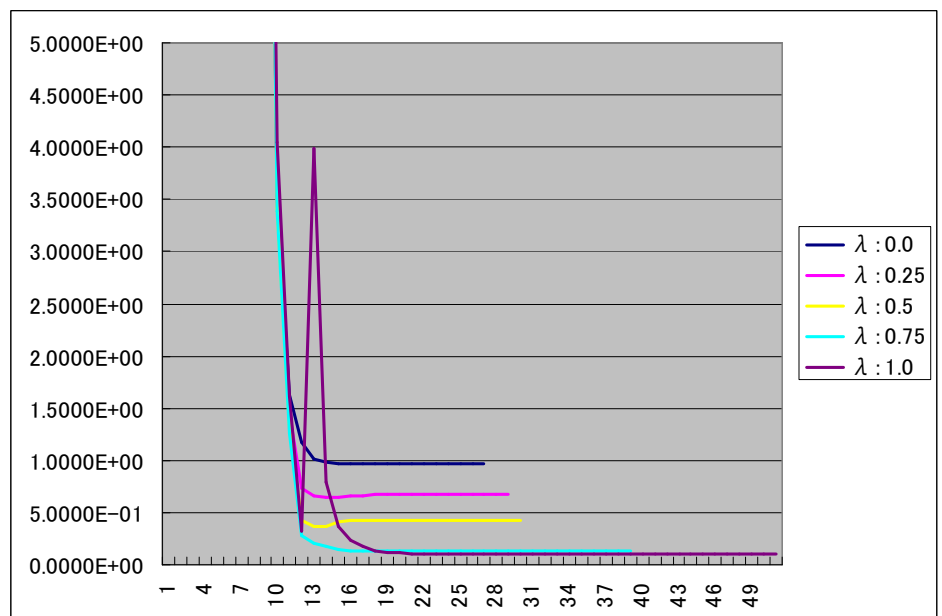
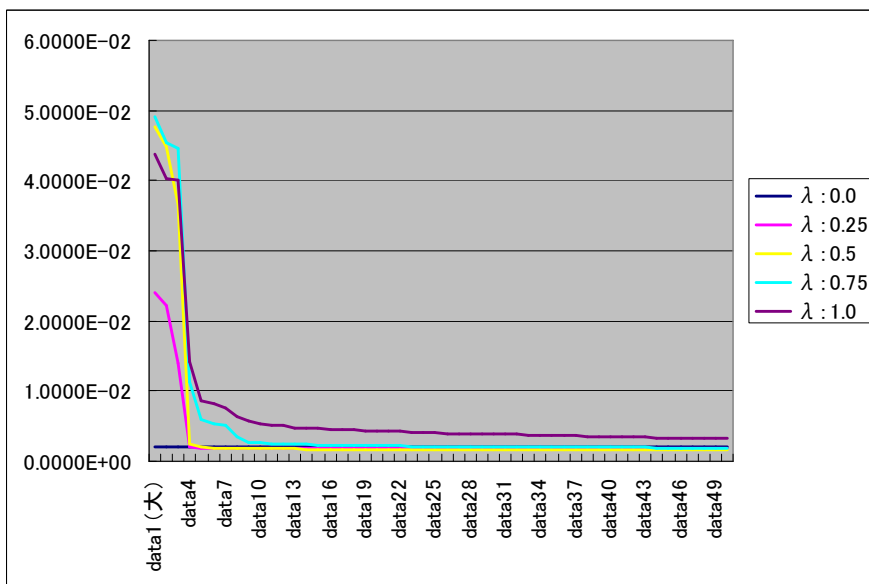


図 3.4 最適重み

- 制約条件なし
- 縦軸は最適重み

$$w_i^{(p)}(\dots, z_k = \|S_k x_{opt} + b_k\|, \dots, \lambda)$$

- 横軸は図 3.2(a)の上位 50 地点



3.2.2. $f = “z$ に関して非凸関数である $Q_1(z)”$ とした場合

$f = Q_1(z)$ では最適化計算が収束せず、最適化計算を強制的に中断して求めた暫定解の平準化能力も

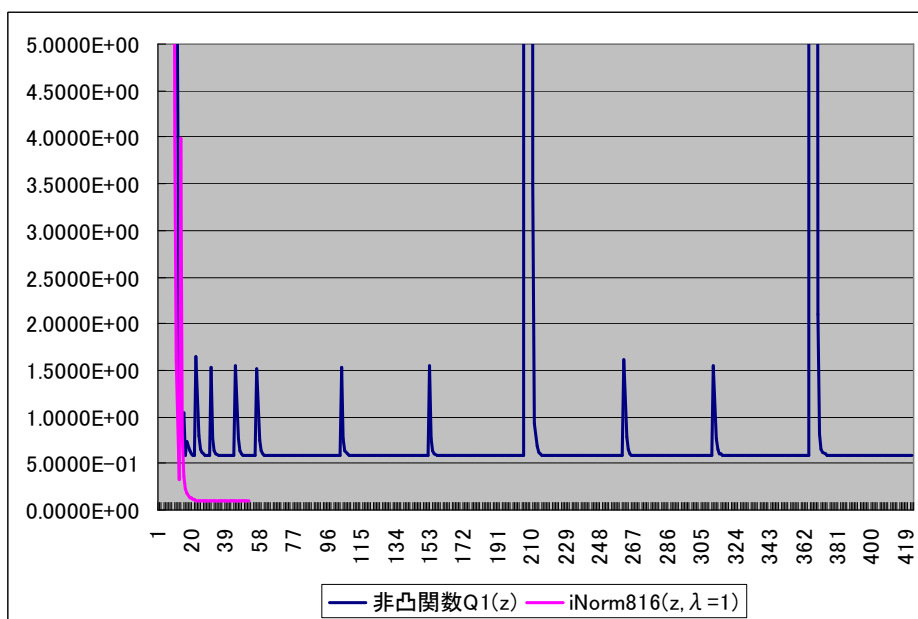
$f = iNorm_{816}(z, 1)$ の場合の最適解より低かったことを下図に示す。

図 3.5 最適化計算中の

$\left\{ \|S_i x_{opt} + b_i\|^2 \right\}$ の分散変化

- 制約条件なし
- 縦軸は $\left\{ \|S_i x_{opt} + b_i\|^2 \right\}$ の分散
- 横軸は解 x の変化回数 (内点法の反復回数とほぼ同義)
- $f = Q_1(z)$ …ピークが何回も立っているグラフ

$f = iNorm_{816}(z, 1)$ …最適化計算が横軸 = 51 の所で収束し終了しているグラフ



4. 結論

$iNorm_{816}(z, \lambda)$ を評価関数に利用して 1.5 節に記載した非線形最適化問題を解く。このとき、“ $\left\{ \|S_i x_{opt} + b_i\|^2 \right\}$ に関する全体的な最小化と最悪値の抑制のバランス”を、一つのパラメータ λ だけを調整することによって効果的に実現することができる。