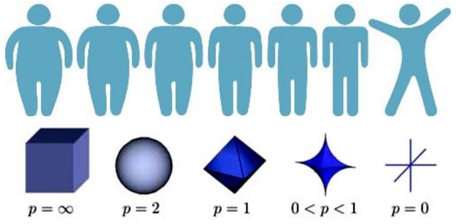
Day 4

スパース解をもつ凸緩和問題



- BP: Basis pursuit

$$\underset{\boldsymbol{x}}{\text{Minimize}}\ \|\boldsymbol{x}\|_1 \quad \text{subject to} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$$

- BP_δ: Constrained BP denoising

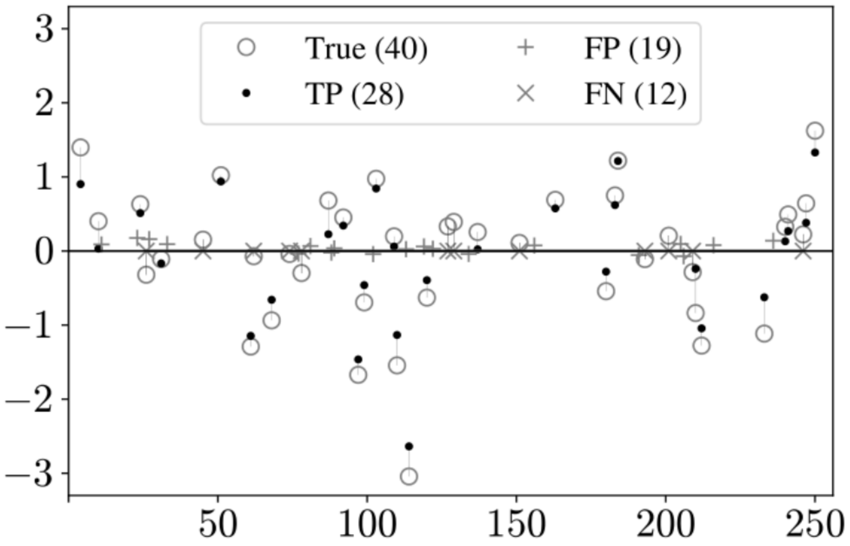
$$\underset{\boldsymbol{x}}{\text{Minimize}}\ \|\boldsymbol{x}\|_1 \quad \text{subject to} \quad \|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|_2 \leq \delta$$

- LASSO: least absolute shrinkage and selection operator

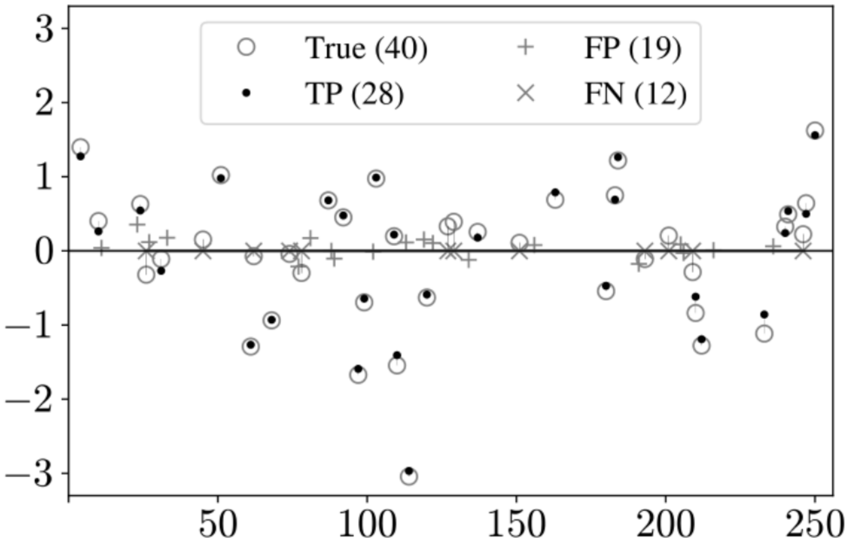
$$\underset{\boldsymbol{x}}{\text{Minimize}}\ \|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|_2 \quad \text{subject to} \quad \|\boldsymbol{x}\|_1 \leq \tau$$

- LASSO regression / unconstrained BP denoising (BPDN) / ℓ_1-regularized least squares (ℓ_1-LS)

$$\underset{\boldsymbol{x}}{\text{Minimize}}\ \frac{1}{2}\|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{x}\|_1$$



(a) FISTA による復元 $\hat{\boldsymbol{x}} = \boldsymbol{x}^{(30)}$



(b) デバイアス後

図 3.22 正規乱数の非ゼロ成分をもつスパース解をFISTA で復元した例．True は正解 $\boldsymbol{x}^\circ \in \mathbb{R}^{256}$ の非ゼロ成分である．正答 TP，誤検出 FP，検出漏れ FN の意味は図 3.8 を参照のこと．

fig:OMP solution Gaussian

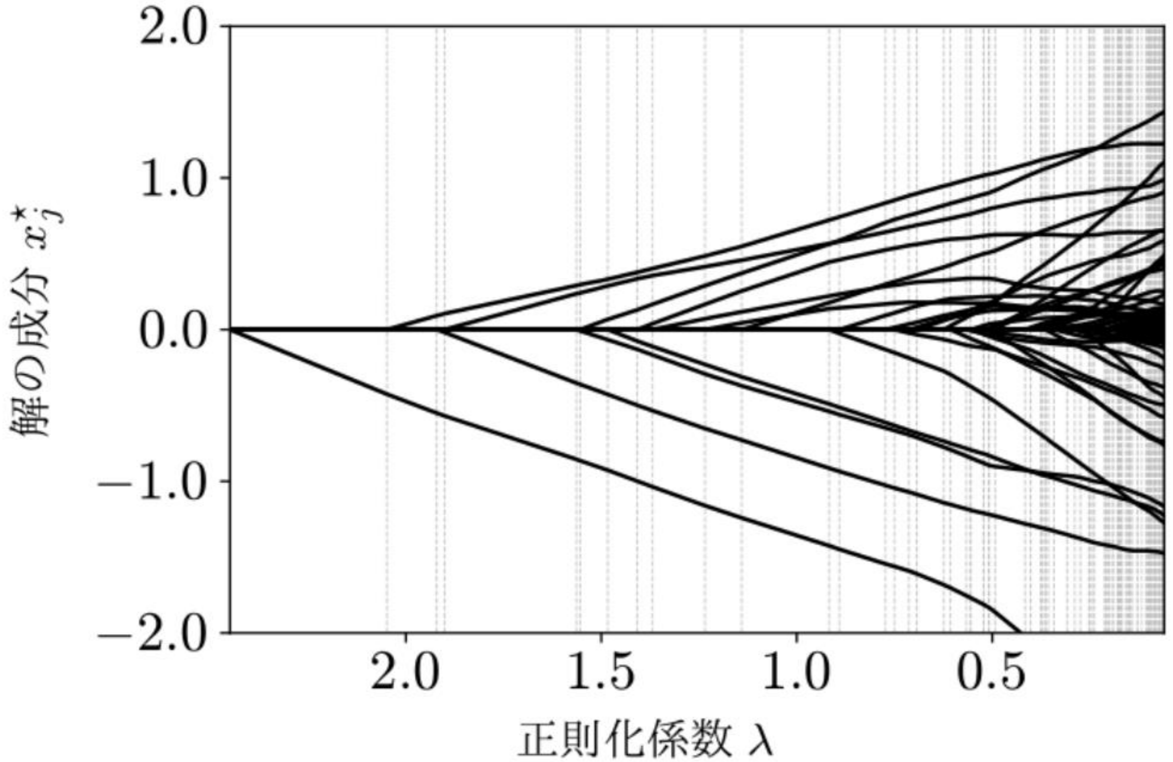


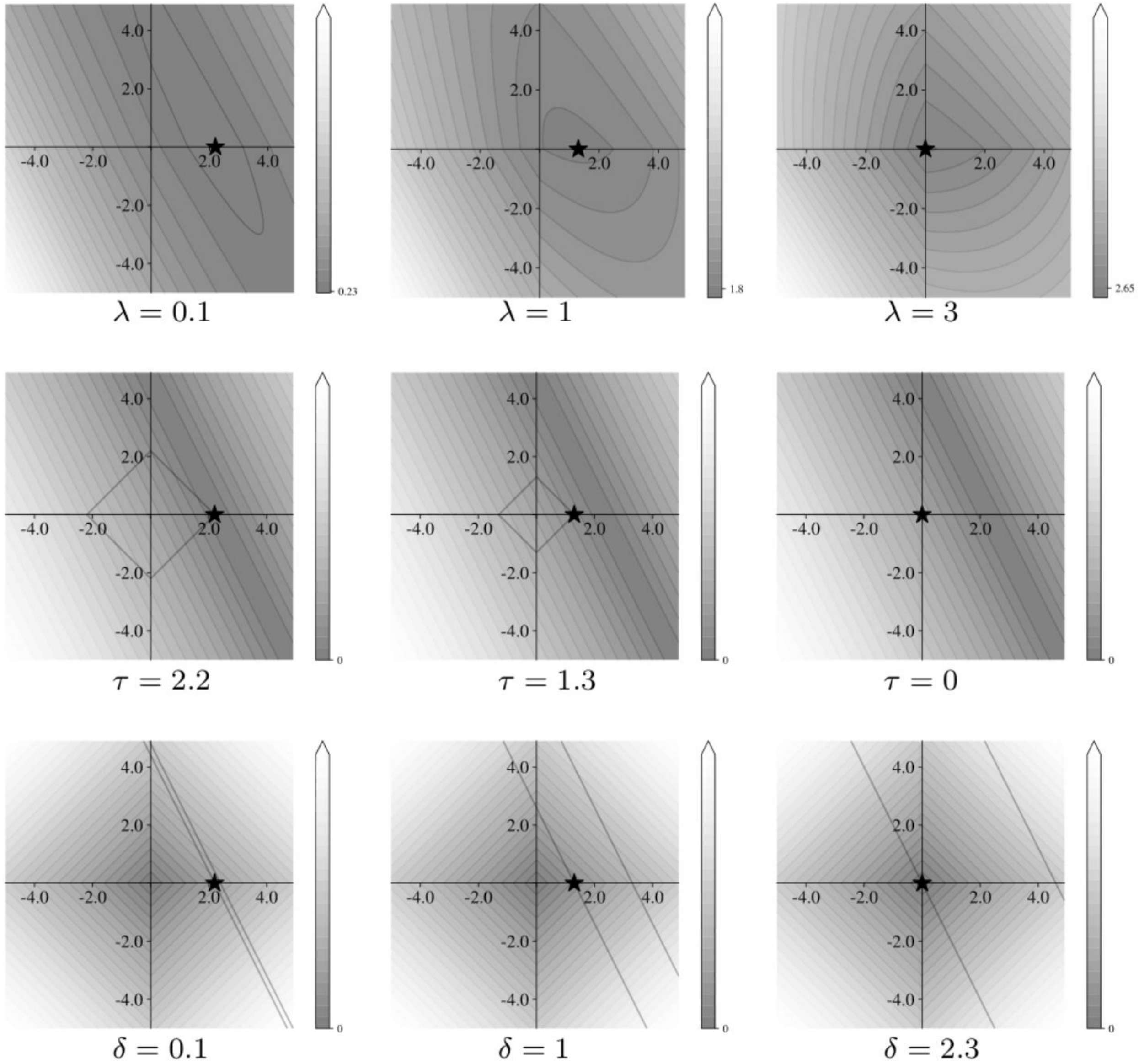
図 3.15 解パスの例．非ゼロ成分の数と値は λ に依存する．

$$\lambda = \sigma_\nu \sqrt{2 \log n} \approx 0.16$$

例 3.9 $m=1$，$n=2$，$\boldsymbol{A}=\begin{bmatrix} 1 & 0.5 \end{bmatrix}$，$b=2.3$ とする．

- LASSO 回帰の目的関数

$$F_\lambda(\boldsymbol{x})=\frac{1}{2}\left(2.3-(x_1+0.5x_2)\right)^2+\lambda(|x_1|+|x_2|)$$



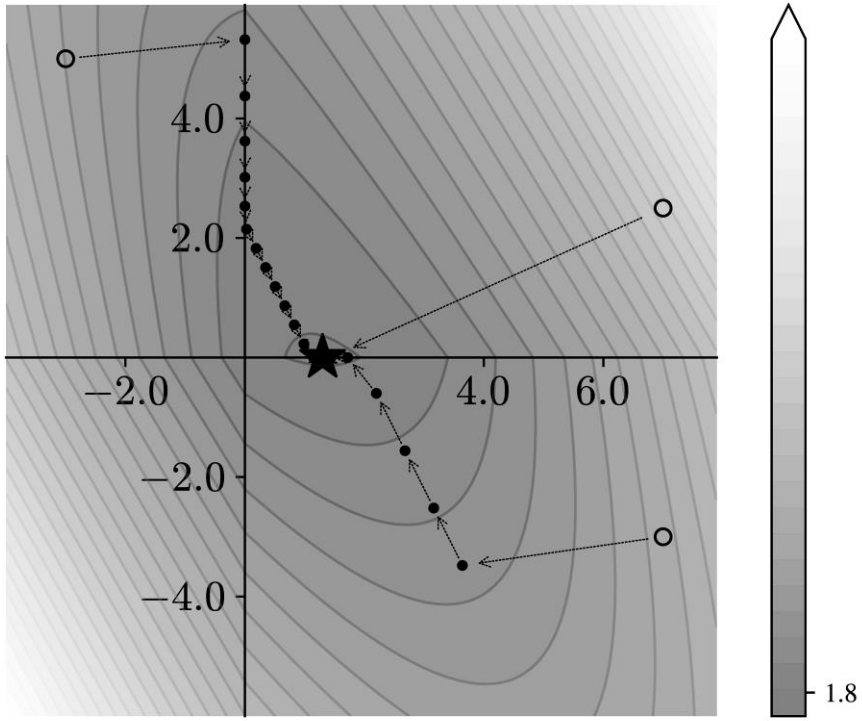
eg:case1

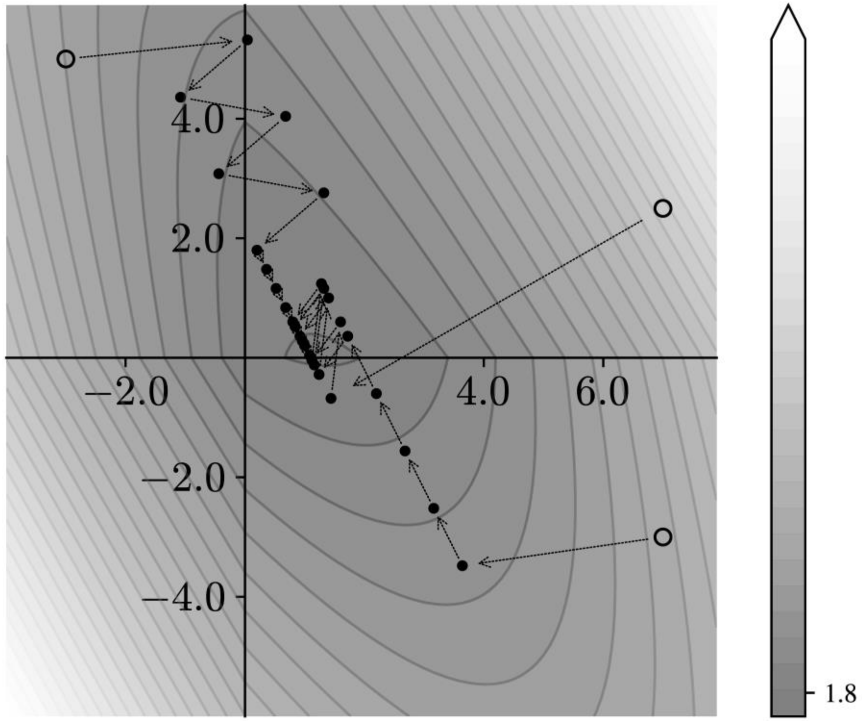
図 3.13 例3.9（$m=1$，$n=2$）の図解．横軸を x_1，縦軸を x_2 として図示している．等高線と $\star$ は，上段：LASSO 回帰，中段：LASSO，下段：BP_δ の目的関数と解を表す．中段の正方形および下段の平行線は，それぞれ LASSO と BP_δ の制約条件が表す領域の境界である．

スパース解をもつ凸緩和問題

- 勾配を使うと，止まれない．
- 解は微分不可能な位置にある．

□ BP: Ba





□ BP_δ: C

$\le \delta$

□ LASS

operator

$\le \tau$

□ LASSO regression / unconstrained BP denoising (BPDN) / ℓ_1-regularized least squares (ℓ_1-LS)

$$\underset{\boldsymbol{x}}{\text{Minimize}} \frac{1}{2}\|\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}\|_2^2+\lambda\|\boldsymbol{x}\|_1$$

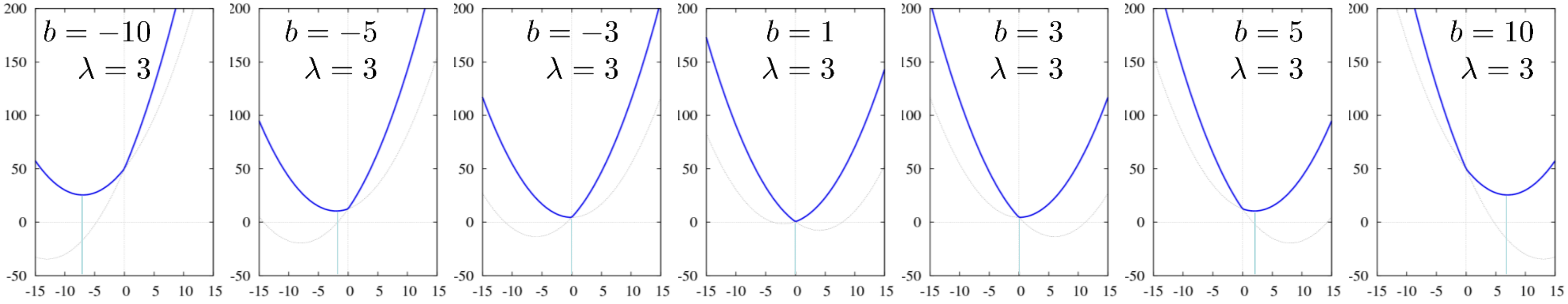
Soft thresholding/shrinkage

$$\boldsymbol{x}^\star = \arg\min_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{x}\|_1 \quad = \mathrm{soft}(\boldsymbol{b}, \lambda) \quad \text{if} \quad \boldsymbol{A} = \boldsymbol{I} \ \text{(identity).}$$

$$x_j^\star = \arg\min_{x_j} \left(\frac{1}{2}(b_j - x_j)^2 + \lambda|x_j|\right) = \begin{cases} b_j - \lambda & (0 < \lambda < b_j) \\ 0 & (-\lambda \le b_j \le \lambda) \\ b_j + \lambda & (b_j < -\lambda < 0) \end{cases}$$

shrink each b_j by λ

$$f_{b,\lambda}(x) = \frac{1}{2}(b - x)^2 + \lambda|x|$$



$$\arg\min_x f_{b,\lambda}(x) = \mathrm{soft}(b, \lambda) = \mathrm{sign}(b)\max(|b| - \lambda, 0)$$

IST

Iterative soft thresholding [Daubechies+, 04]

$$\boldsymbol{x}^\star = \arg\min_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{x}\|_1$$

$$\boxed{\boldsymbol{x}^{(k+1)} := \mathrm{soft}\left(\boldsymbol{x}^{(k)} + \frac{1}{L}\boldsymbol{A}^\top\left(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^{(k)}\right),\ \frac{\lambda}{L}\right)}$$

- The constant L must be larger than the maximum eigenvalue of $\boldsymbol{A}^\top\boldsymbol{A}$.
- $\left\|\boldsymbol{x}^{(k)} - \boldsymbol{x}^\star\right\| = \mathcal{O}(1/k)$
- IST can be derived from locally approximating the objective function with $\boldsymbol{A}^\top\boldsymbol{A} \approx L\boldsymbol{I}$.
- a.k.a. the proximal gradient method.

Ex. $\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 2 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix}$, $\boldsymbol{b} = \begin{bmatrix} 2 \\ -2 \\ -1 \\ -1 \end{bmatrix}$

$\lambda = 0.1,\ L = 5,\ \boldsymbol{x}^{(0)} = \boldsymbol{A}^\top\boldsymbol{b}$

k	x				
0	[2. 00	-3. 00	-4. 00	-1. 00	-1. 00]'
10	[2. 32	-0. 97	-0. 72	0. 19	-0. 50]'
20	[2. 41	-0. 78	-0. 56	0. 23	-0. 39]'
30	[2. 47	-0. 67	-0. 60	0. 19	-0. 33]'
40	[2. 52	-0. 57	-0. 65	0. 15	-0. 28]'
50	[2. 57	-0. 47	-0. 70	0. 10	-0. 23]'
60	[2. 62	-0. 37	-0. 75	0. 05	-0. 18]'
70	[2. 67	-0. 27	-0. 80	0. 00	-0. 13]'
80	[2. 73	-0. 19	-0. 85	0. 00	-0. 07]'
90	[2. 77	-0. 13	-0. 89	0. 00	-0. 03]'
100	[2. 81	-0. 07	-0. 93	0. 00	-0. 00]'
⋯	[3. 00	0. 00	-1. 00	0. 00	0. 00]'

FISTA

Fast iterative soft thresholding algorithm [Beck&Teboulle, 09] [Nesterov, 83]

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{x}\|_1$$

$$\boldsymbol{w}^{(1)} \coloneqq \boldsymbol{x}^{(0)}, \quad t_1 \coloneqq 1$$

$$\boldsymbol{x}^{(k)} \coloneqq \mathrm{soft}\left(\boldsymbol{w}^{(k)} + \frac{1}{L}\boldsymbol{A}^\top\left(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{w}^{(k)}\right), \ \frac{\lambda}{L}\right)$$

$$t_{k+1} \coloneqq \frac{1}{2}\left(1 + \sqrt{1 + 4t_k^2}\right)$$

$$\boldsymbol{w}^{(k+1)} \coloneqq \boldsymbol{x}^{(k)} + \frac{t_k - 1}{t_{k+1}}\left(\boldsymbol{x}^{(k)} - \boldsymbol{x}^{(k-1)}\right)$$

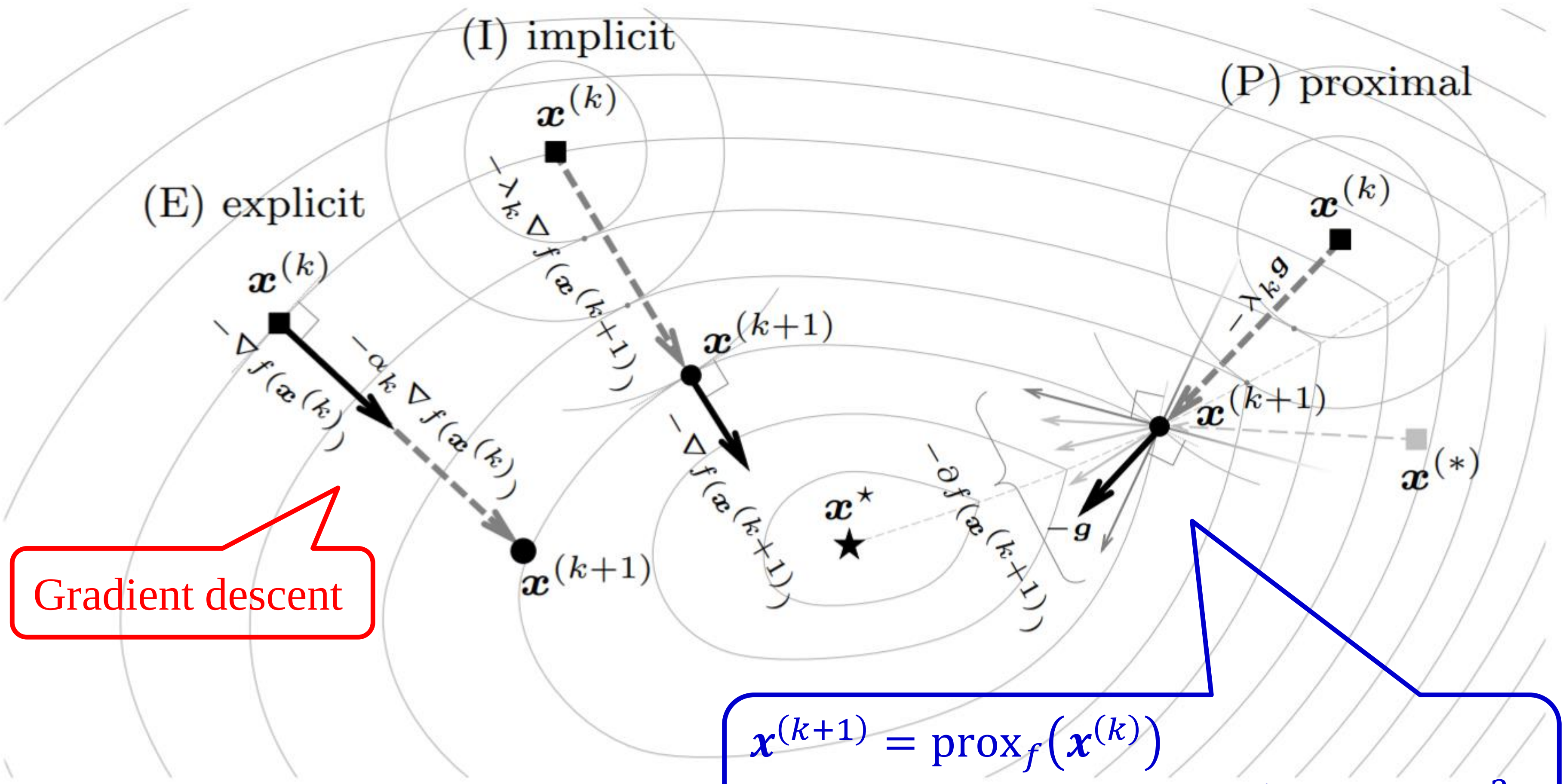
- $\left\|\boldsymbol{x}^{(k)} - \boldsymbol{x}^\star\right\| = \mathcal{O}(1/k^2)$
- IST with the Nesterov acceleration

Ex. $\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 2 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix}, \ \boldsymbol{b} = \begin{bmatrix} 2 \\ -2 \\ -1 \\ -1 \end{bmatrix}$

$\lambda = 0.1, \ L = 10, \ \boldsymbol{x}^{(0)} = \boldsymbol{A}^\top\boldsymbol{b}$

k	x				
0	[2.00	−3.00	−4.00	−1.00	−1.00]'
10	[2.32	−0.87	−0.48	0.26	−0.49]'
20	[2.49	−0.68	−0.61	0.22	−0.31]'
30	[2.65	−0.31	−0.78	0.02	−0.15]'
40	[2.90	0.00	−0.98	−0.00	−0.00]'
50	[2.86	−0.00	−0.97	−0.00	−0.00]'
60	[2.85	−0.02	−0.95	−0.00	−0.00]'
70	[2.87	−0.00	−0.97	−0.00	−0.00]'
80	[2.87	−0.00	−0.97	−0.00	−0.00]'
90	[2.87	−0.00	−0.97	−0.00	−0.00]'
100	[2.87	−0.00	−0.97	−0.00	−0.00]'
⋯	[3.00	0.00	−1.00	0.00	0.00]'

近接点法 ＝ 陰的(劣)勾配降下法



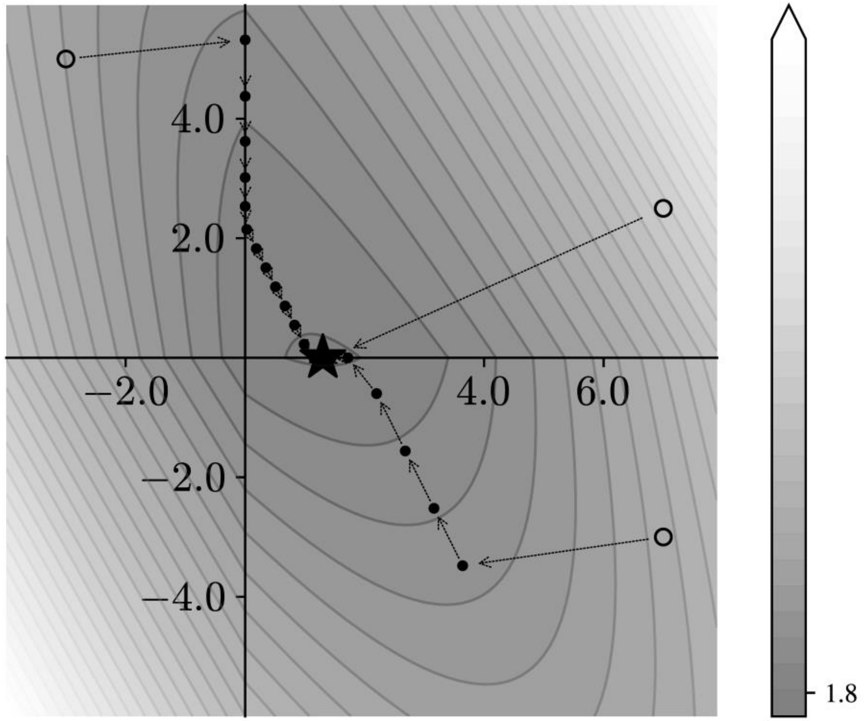
$$
\begin{aligned}
\boldsymbol{x}^{(k+1)} &= \mathrm{prox}_f\left(\boldsymbol{x}^{(k)}\right) \\
&\coloneqq \arg\min_{\boldsymbol{x}} \; f(\boldsymbol{x}) + \frac{1}{2}\left\|\boldsymbol{x} - \boldsymbol{x}^{(k)}\right\|_2^2
\end{aligned}
$$

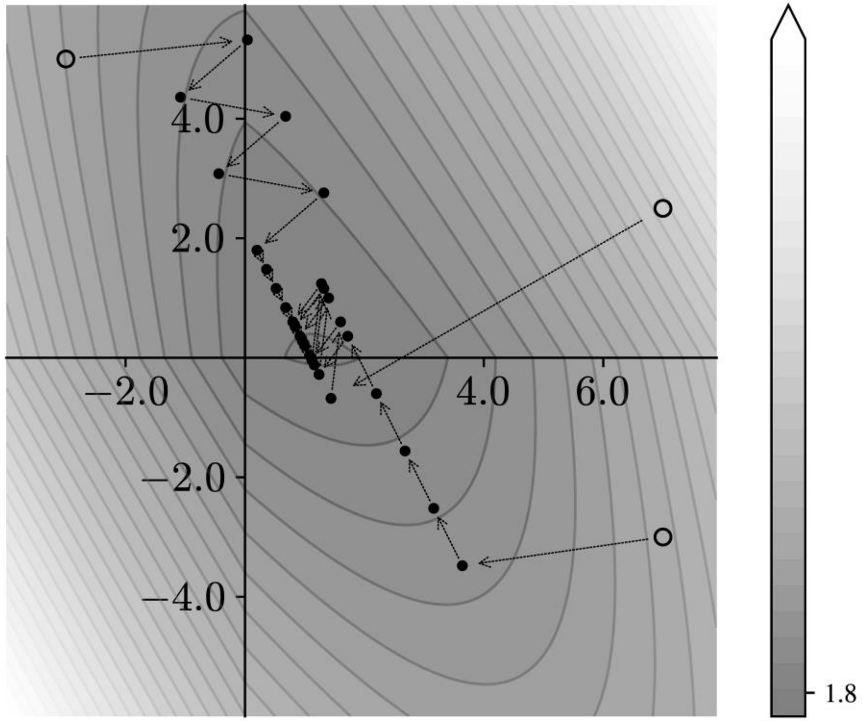
スパース解をもつ凸緩和問題

- 勾配を使うと，止まれない．
- 解は微分不可能な位置にある．

- BP: Ba

  M





- BP$_\delta$: C

  M $\le \delta$

- LASS operator

  M $\le \tau$

- LASSO regression / unconstrained BP denoising (BPDN) / ℓ_1-regularized least squares (ℓ_1-LS)

  $$\underset{\boldsymbol{x}}{\text{Minimize}} \frac{1}{2}\|\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x}\|_2^2+\lambda\|\boldsymbol{x}\|_1$$

近接写像の反復法

- 凸最小化問題 $\underset{\boldsymbol{x}\in\mathbb{R}^{d_x}}{\text{Minimize}} \ \ f_1(\boldsymbol{x}) + f_2(\boldsymbol{x}) + \cdots + f_n(\boldsymbol{x})$

  $f_1, \ldots, f_n$ は凸関数で，そのいくつかは微分不可能

- 近接作用素 $\boxed{\operatorname{prox}_f(\boldsymbol{q}) = \underset{\boldsymbol{x}}{\arg\min} \ f(\boldsymbol{x}) + \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{q}\|_2^2}$ 近似解の更新に使う

  - $\boldsymbol{p} = \operatorname{prox}_f(\boldsymbol{q})$ とすると $f(\boldsymbol{p}) \le f(\boldsymbol{q})$.
  - $\boldsymbol{x}^{(k+1)} = \operatorname{prox}_f\left(\boldsymbol{x}^{(k)}\right) \ \Leftrightarrow \ f\left(\boldsymbol{x}^{(k+1)}\right) \le f\left(\boldsymbol{x}^{(k)}\right) \quad \because \ -\left(\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)}\right) \in \partial f\left(\boldsymbol{x}^{(k+1)}\right)$ 劣勾配降下
  - 最小解 $\boldsymbol{x}^\star$ は不動点： $\boldsymbol{x}^\star = \operatorname{prox}_f(\boldsymbol{x}^\star) = \underset{\boldsymbol{x}}{\arg\min} \ f(\boldsymbol{x}).$

- 算法
  - 反復ソフト閾値処理（ISTA）/ alternating projections / averaged projections / ADMM, etc.

近接作用素の例

$$\mathrm{prox}_f(\boldsymbol{q}) = \arg\min_{\boldsymbol{x}}\ f(\boldsymbol{x}) + \frac{1}{2}\|\boldsymbol{x}-\boldsymbol{q}\|_2^2$$

- ℓ_p ノルムのp乗： $f(\boldsymbol{x}) = \lambda\,\|\boldsymbol{x}\|_p^p$
  - $p=2$:　$\mathrm{prox}_{\lambda\|\cdot\|_2^2}(\boldsymbol{q}) = \boldsymbol{q}/(2\lambda+1)$　　$\boldsymbol{q}$ を縮小する処理
  - $p=1$:　$\mathrm{prox}_{\lambda\|\cdot\|}(\boldsymbol{q}) = \mathrm{soft}(\boldsymbol{q},\lambda)$　　$\boldsymbol{q}$ の各成分をλだけ切り詰める処理

- 核ノルム $\|\boldsymbol{A}\|_* =$（行列$\boldsymbol{A}$の特異値 $\boldsymbol{\kappa} = [\kappa_1,\dots,\kappa_r]^\top$の和）$= \|\boldsymbol{\kappa}\|_1$　ただし $\boldsymbol{A} = \boldsymbol{U}\mathrm{diag}(\boldsymbol{\kappa})\boldsymbol{V}^\top$
  $\boldsymbol{A}$ の低ランク性（行または列の線形従属性）を表す凸関数
  - $f(\boldsymbol{x}) = \lambda\|\boldsymbol{A}\|_*$　$\mathrm{prox}_{\lambda\|\cdot\|_*}(\boldsymbol{A}) = \boldsymbol{U}\mathrm{diag}\big(\mathrm{soft}(\boldsymbol{\kappa},\lambda)\big)\boldsymbol{V}^\top \coloneqq \mathrm{svt}(\boldsymbol{A},\lambda)$ (singular value thresholding)

- 指示関数 $\iota_{\mathcal{C}}(\boldsymbol{x}) = \begin{cases} 0 & \text{if } \boldsymbol{x}\in\mathcal{C} \\ \infty & \text{otherwise}\end{cases}$　制約条件を満たすか否かを表す凸関数
  - $\mathcal{C}_{\boldsymbol{c},\varepsilon} = \{\boldsymbol{x} \mid \|\boldsymbol{x}-\boldsymbol{c}\|_2 \le \varepsilon\}$:　$\mathrm{prox}_{\iota_{\mathcal{C}_{\boldsymbol{c},\varepsilon}}}(\boldsymbol{q}) = \boldsymbol{c} + \varepsilon\frac{\boldsymbol{q}-\boldsymbol{c}}{\|\boldsymbol{q}-\boldsymbol{c}\|_2}$
  - $\mathcal{C}_{\boldsymbol{A},\boldsymbol{b}} = \{\boldsymbol{x} \mid \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\}$:　$\mathrm{prox}_{\iota_{\mathcal{C}_{\boldsymbol{A},\boldsymbol{b}}}}(\boldsymbol{q}) = (\boldsymbol{I} - \boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{A}^\top)^{-1}\boldsymbol{A})\boldsymbol{q} + \boldsymbol{A}^\top(\boldsymbol{A}\boldsymbol{A}^\top)^{-1}\boldsymbol{b}$

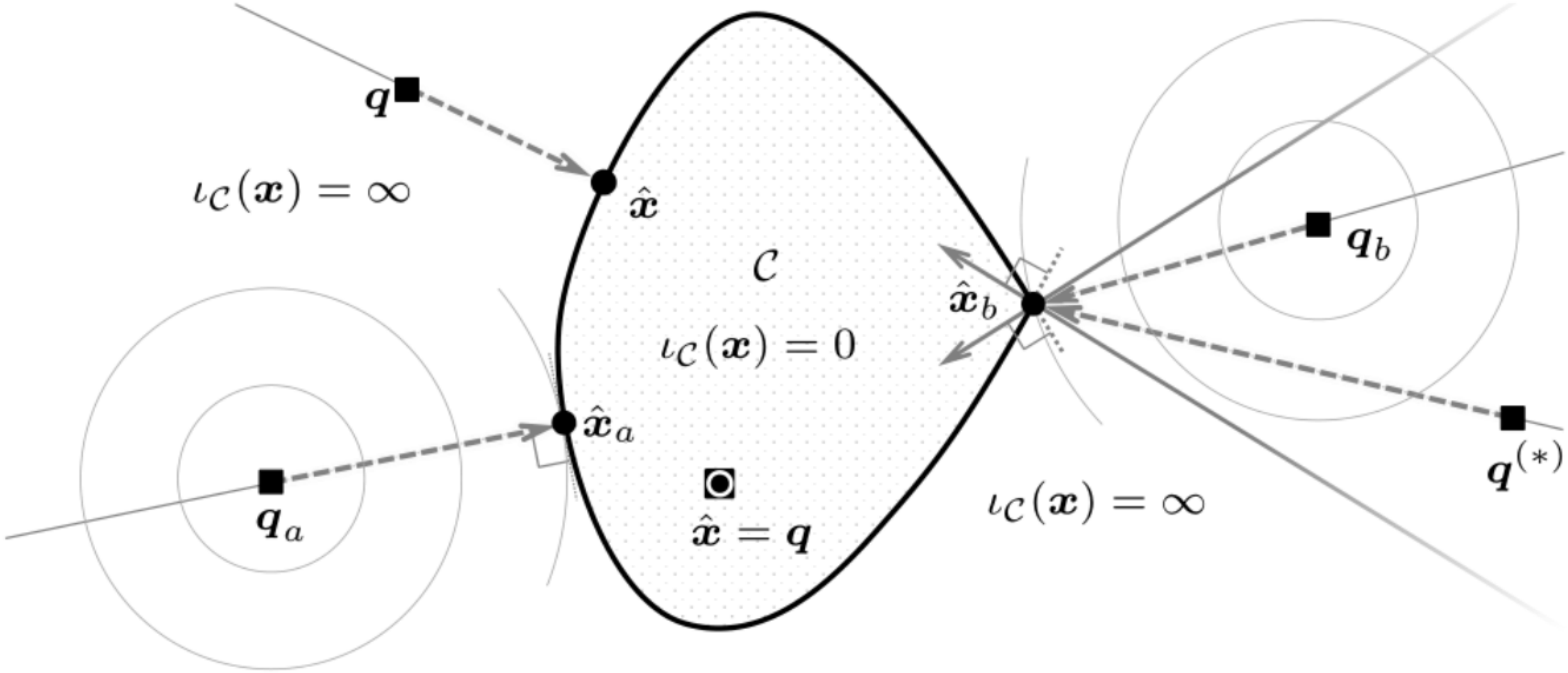


図 4.3　閉凸集合 $\mathcal{C}$ の標示関数 $\iota_{\mathcal{C}}$ と $\mathcal{C}$ への近接写像．$\mathcal{C}$ の外部の点 $\boldsymbol{q} \notin \mathcal{C}$ の近接点 $\hat{\boldsymbol{x}}$ は，$\boldsymbol{q}$ から最寄りの $\mathcal{C}$ の境界に位置する．

$$\begin{aligned}\mathrm{prox}_{\iota_{\mathcal{C}}}(\boldsymbol{q}) &= \arg\min_{\boldsymbol{x}} \iota_{\mathcal{C}}(\boldsymbol{x}) + \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{q}\|_2^2 \\ &= \arg\min_{\boldsymbol{x}\in\mathcal{C}} \|\boldsymbol{x} - \boldsymbol{q}\|_2\end{aligned}$$

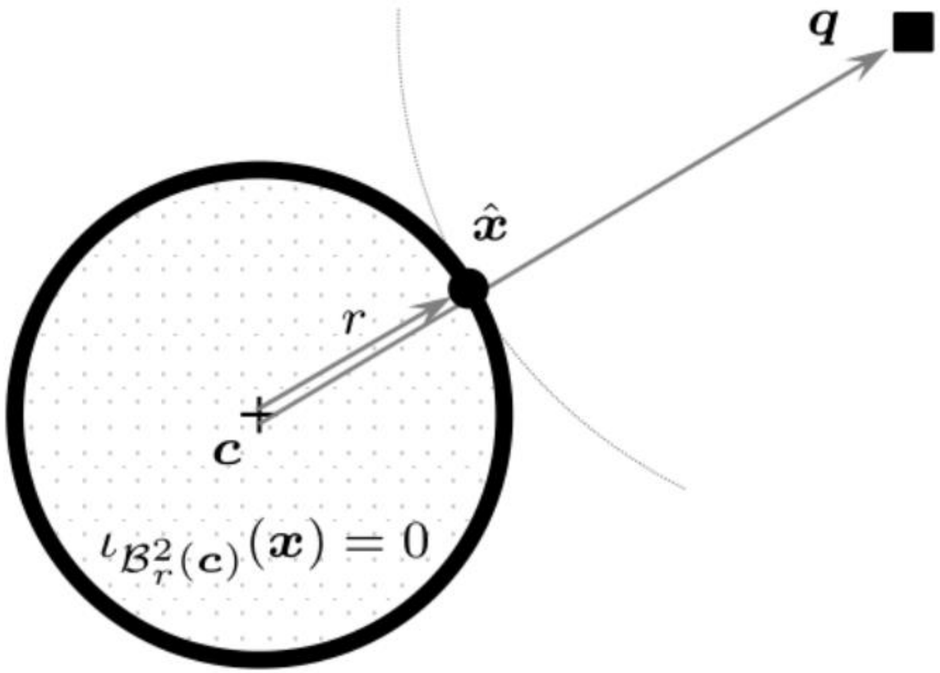


図 4.7　閉 ℓ_2 超球体 $\mathcal{B}_r^2(\boldsymbol{c})$ への近接写像．閉 ℓ_2 超球体外部の点 $\boldsymbol{q} \notin \mathcal{B}_r^2(\boldsymbol{c})$ の近接点 $\hat{\boldsymbol{x}}$ は，中心 $\boldsymbol{c}$ から $\boldsymbol{q} - \boldsymbol{c}$ の方向に距離 r だけ移動した超球面 $\partial\mathcal{B}_r^2(\boldsymbol{c})$ の上に位置する．

$$
\begin{aligned}
\mathrm{prox}_{\iota_{\mathcal{C}}}(\boldsymbol{q}) &= \arg\min_{\boldsymbol{x}} \iota_{\mathcal{C}}(\boldsymbol{x}) + \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{q}\|_2^2 \\
&= \arg\min_{\boldsymbol{x}\in\mathcal{C}} \|\boldsymbol{x} - \boldsymbol{q}\|_2
\end{aligned}
$$

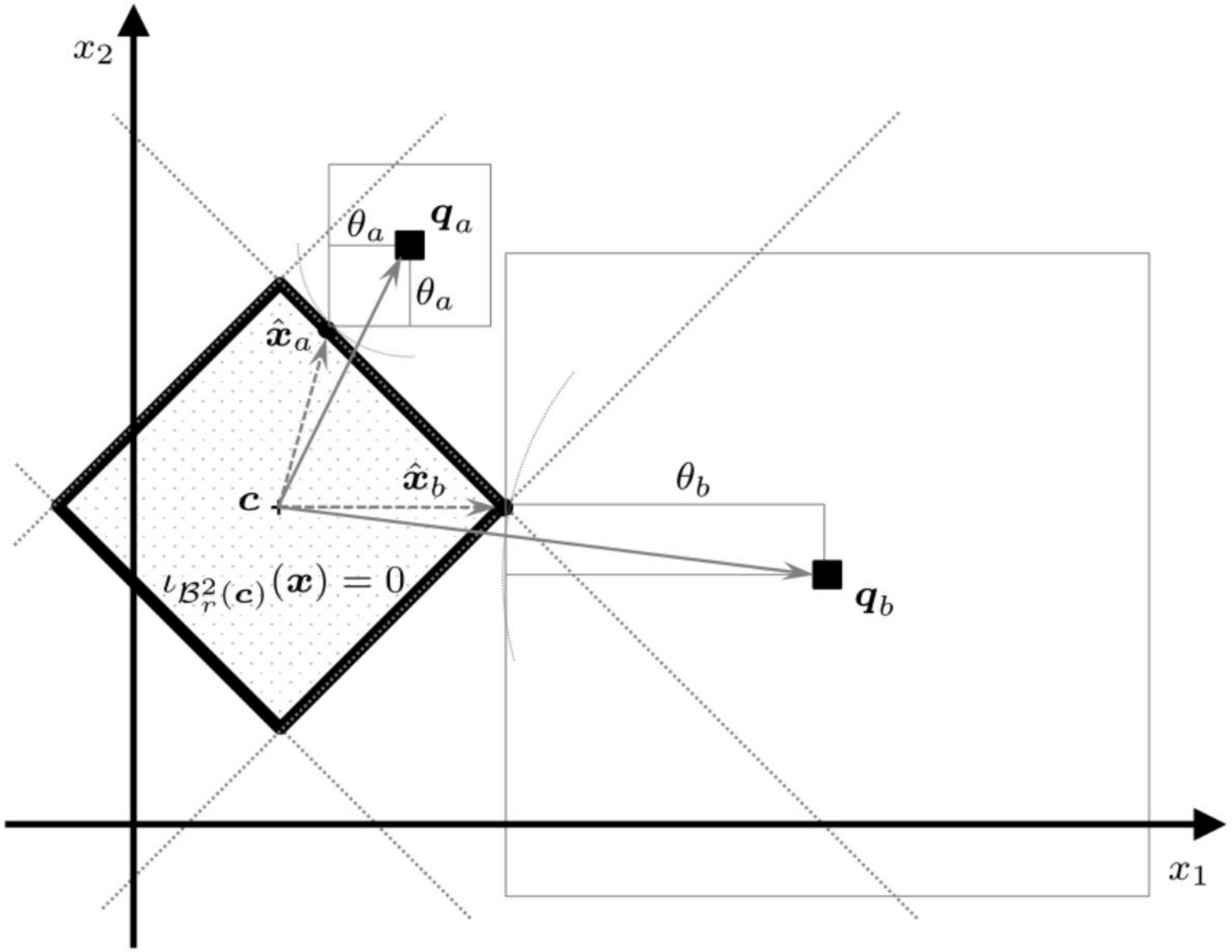


図 **4.8**　閉 ℓ_1 超球体 $\mathcal{B}_r^1(\boldsymbol{c})$ への近接写像．$\mathcal{B}_r^1(\boldsymbol{c})$ 外部の点 $\boldsymbol{q}_a$ と $\boldsymbol{q}_b$ の近接点はそれぞれ $\hat{\boldsymbol{x}}_a = \boldsymbol{c} + \mathrm{soft}(\boldsymbol{q}_a - \boldsymbol{c}, \theta_a)$，$\hat{\boldsymbol{x}}_b = \boldsymbol{c} + \mathrm{soft}(\boldsymbol{q}_b - \boldsymbol{c}, \theta_b)$ である．

$$\hat{\boldsymbol{x}} = \mathrm{prox}_{\iota_{\mathcal{B}_r^1(\boldsymbol{c})}}(\boldsymbol{q}) = \boldsymbol{c} + \mathrm{soft}(\boldsymbol{q} - \boldsymbol{c}, \theta)$$

$$\theta \leftarrow \mathrm{FindThresh}(\boldsymbol{q} - \boldsymbol{c}, r)$$

アルゴリズム 11 正軸体へ近接写像するための閾値（並べ替え法）[126]：$\theta \leftarrow \textsc{FindThreshS}(\boldsymbol{v}, r)$

入力： $\boldsymbol{v} = [v_1, \ldots, v_n]^\top \in \mathbb{R}^n$，正軸体の半径 $r \geqq 0$;
出力： θ: 閾値;

```
v の成分の絶対値を降順に並べたものを u_1, ..., u_n とする;
K ← max_{k∈{1,...,n}} {k | (Σ_{j=1}^{k} u_j − r)/k < u_k}: 不等式を満たす最大の k を K とする;
θ = (Σ_{j=1}^{K} u_j − r)/K;
```

アルゴリズム 12 正軸体へ近接写像するための閾値（アクティブセット法）[127]：$\theta \leftarrow \textsc{FindThreshA}(\boldsymbol{v}, r)$

入力： $\boldsymbol{v} = [v_1, \ldots, v_n]^\top \in \mathbb{R}^n$，正軸体の半径 $r \geqq 0$;
出力： θ: 閾値;

```
台の初期値を T ← {1, ..., n} とする;
初期値 θ ← (‖v‖_1 − r)/n を計算する;
k ← 0;
while |T| ≠ k do
    k ← |T|;
    T ← {j | |v_j| > θ, j ∈ T}: 閾値 θ を超える v の成分の番号を集合 T に残す;
    v の台 T の成分のみから θ ← (‖v_T‖_1 − r)/|T| を計算する;
end while
```

Proximal gradient and IST

Bruck, Lions, Mercier (1970s)

$\underset{\boldsymbol{x}}{\text{Minimize}}\ f(\boldsymbol{x}) + g(\boldsymbol{x})$ for smooth f

$$\boxed{\boldsymbol{x}^{(k+1)} = \operatorname{prox}_{\gamma g}\left(\boldsymbol{x}^{(k)} - \gamma \nabla f(\boldsymbol{x}^{(k)}) \right)}$$

$$= \arg\min_{\boldsymbol{x}}\ \gamma \left(g(\boldsymbol{x}) + \tfrac{1}{2\gamma} \|\boldsymbol{x} - (\boldsymbol{x}^{(k)} - \gamma \nabla f(\boldsymbol{x}^{(k)}) \|_2^2 \right)$$

$$= \arg\min_{\boldsymbol{x}} \left(g(\boldsymbol{x}) + \nabla f\left(\boldsymbol{x}^{(k)}\right)^\top \left(\boldsymbol{x} - \boldsymbol{x}^{(k)}\right) + \tfrac{1}{2\gamma} \left\|\boldsymbol{x} - \boldsymbol{x}^{(k)}\right\|_2^2 + \tfrac{1}{2\gamma} \|\nabla f(\boldsymbol{x}^{(k)})\|_2^2 \quad + f\left(\boldsymbol{x}^{(k)}\right) \right)$$

$$= \arg\min_{\boldsymbol{x}} \left(g(\boldsymbol{x}) + f\left(\boldsymbol{x}^{(k)}\right) + \nabla f\left(\boldsymbol{x}^{(k)}\right)^\top \left(\boldsymbol{x} - \boldsymbol{x}^{(k)}\right) + \tfrac{1}{2\gamma} \left\|\boldsymbol{x} - \boldsymbol{x}^{(k)}\right\|_2^2 \right)$$

- $\boldsymbol{x}^{(k+1)}$ minimizes $g(\boldsymbol{x})$ plus the 2nd order Taylor approximation of $f(\boldsymbol{x})$ around $\boldsymbol{x}^{(k)}$.
- If $g(\boldsymbol{x}) = \lambda \|\boldsymbol{x}\|_1$, we have $\boxed{\text{IST:} \quad \boldsymbol{x}^{(k+1)} = \operatorname{soft}\left(\boldsymbol{x}^{(k)} - \gamma \nabla f\left(\boldsymbol{x}^{(k)}\right),\ \gamma\lambda\right).}$
  - Special case: linear model fitting, ℓ_2 loss $f(\boldsymbol{x}) = \frac{1}{2} \|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|_2^2$ and $\gamma = \frac{1}{L}$;

$$\boldsymbol{x}^{(k+1)} = \operatorname{soft}\left(\boldsymbol{x}^{(k)} + \tfrac{1}{L} \boldsymbol{A}^\top \left(\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}^{(k)}\right),\ \tfrac{\lambda}{L} \right)$$

ADMM

Alternating direction method of multipliers [Gabay&Mercier, 76]

$$\underset{(\boldsymbol{x},\boldsymbol{z})}{\text{Minimize}} \quad f(\boldsymbol{x}) + g(\boldsymbol{z}) \quad \text{subject to} \quad \boldsymbol{z} = \boldsymbol{G}\boldsymbol{x}$$

$$\boldsymbol{x}^{(k+1)} := \underset{\boldsymbol{x}}{\arg\min}\, f(\boldsymbol{x}) + \frac{\rho}{2}\left\|\boldsymbol{G}\boldsymbol{x} - (\boldsymbol{z}^{(k)} - \boldsymbol{u}^{(k)})\right\|_2^2$$

$$\boldsymbol{z}^{(k+1)} := \underset{\boldsymbol{z}}{\arg\min}\, g(\boldsymbol{z}) + \frac{\rho}{2}\left\|\boldsymbol{z} - (\boldsymbol{G}\boldsymbol{x}^{(k+1)} + \boldsymbol{u}^{(k)})\right\|_2^2$$

$$\boldsymbol{u}^{(k+1)} := \boldsymbol{u}^{(k)} + \boldsymbol{G}\boldsymbol{x}^{(k+1)} - \boldsymbol{z}^{(k+1)}$$

SALSA

Split augmented Lagrangian shrinkage algorithm [Afonso+, 10]

$$\boldsymbol{x}^\star = \arg\min_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{x}\|_1$$

$$\begin{aligned}
\boldsymbol{v} &:= \mathrm{soft}\left(\boldsymbol{x}^{(k)} + \boldsymbol{y}^{(k)}, \frac{\lambda}{\rho}\right) - \boldsymbol{y}^{(k)} \\
\boldsymbol{x}^{(k+1)} &:= (\boldsymbol{A}^\top\boldsymbol{A} + \rho\boldsymbol{I})^{-1}(\boldsymbol{A}^\top\boldsymbol{b} + \rho\boldsymbol{v}) \\
\boldsymbol{y}^{(k+1)} &:= \boldsymbol{x}^{(k+1)} - \boldsymbol{v}
\end{aligned}$$

- $\left\|\boldsymbol{x}^{(k)} - \boldsymbol{x}^\star\right\| = \mathcal{O}(1/k)$
- converges $\forall\rho > 0$
- derived from ADMM

Ex. $\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 2 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 2 \\ -2 \\ -1 \\ -1 \end{bmatrix}$

$\lambda = 0.1, \ \rho = 5, \ \boldsymbol{x}^{(0)} = \boldsymbol{A}^\top\boldsymbol{b}$

k	x				
0	[2.00	−3.00	−4.00	−1.00	−1.00]’
10	[2.52	−0.57	−0.65	0.15	−0.28]’
20	[2.77	−0.12	−0.89	−0.01	−0.03]’
30	[2.86	−0.00	−0.96	−0.00	0.00]’
40	[2.87	−0.00	−0.97	−0.00	0.00]’
50	[2.87	−0.00	−0.97	−0.00	0.00]’
60	[2.87	−0.00	−0.97	−0.00	0.00]’
70	[2.87	−0.00	−0.97	−0.00	0.00]’
80	[2.87	−0.00	−0.97	−0.00	0.00]’
90	[2.87	−0.00	−0.97	−0.00	0.00]’
100	[2.87	0.00	−0.97	0.00	0.00]’
⋯	[3.00	0.00	−1.00	0.00	0.00]’

今日は
ここまで