Inverse Problems
Classical regularisation theory

Clarice Poon
University of Bath

January 28, 2020

Outline

Regularisation Theory

Spectral regularisation

Other parameter choice rules

Iterative regularisation

We saw that $A^\dagger$ is generally unbounded and so, given noisy data f_δ such that $\|f_\delta - f\| \leqslant \delta$, we cannot expect $A^\dagger f_\delta \to A^\dagger f$ as $\delta \to 0$.
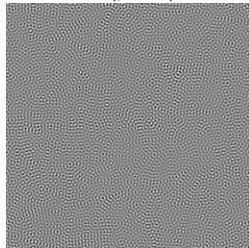
$$y = \mathcal{K}(I) = h \star I \qquad\qquad \mathcal{K}^\dagger(y) \qquad\qquad \mathcal{K}^\dagger(y + \varepsilon)$$

To achieve convergence, replace $A^\dagger$ with a family of well-posed (bounded) operators R_α with $\alpha = \alpha(\delta, f_\delta)$ such that $R_\alpha f_\delta \to A^\dagger f$ for all $f \in \mathcal{D}(A^\dagger)$ and all $f_\delta \in \mathcal{V}$ such that $\|f - f_\delta\|_{\mathcal{V}} < \delta$ as $\delta \to 0$.

Definition 1 (Regularisation)

Let $A \in \mathcal{L}(\mathcal{U}, \mathcal{V})$. A family $\{R_\alpha\}_{\alpha>0}$ of continuous operators is called a regularisation of $A^\dagger$ if $R_\alpha f \to A^\dagger f = u^\dagger$ for all $f \in \mathcal{D}(A^\dagger)$ as $\alpha \to 0$.

$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = $ 1e-4	$\alpha = $ 1e-9

We say that the family $\{R_\alpha\}_\alpha$ is a linear regularisation of $A^\dagger$ if they consist of linear operators.

No uniform boundedness of regularisers

We cannot expect to do better than this definition, i.e. in general, we cannot expect $R_\alpha f$ to converge for $f \notin \mathcal{D}(A^\dagger)$.

Assume: $A \in \mathcal{L}(\mathcal{U}, \mathcal{V})$ and let $\{R_\alpha\}_\alpha$ be a linear regularisation of $A^\dagger$.

Theorem 2

If $A^\dagger$ is not continuous, then $\{R_\alpha\}_\alpha$ cannot be uniformly bounded. In particular, there exists $f \in \mathcal{V}$ such that $\|R_\alpha f\| \to \infty$ as $\alpha \to 0$.

Theorem 3

If $\sup_{\alpha>0} \|AR_\alpha\|_{\mathcal{L}(\mathcal{V}, \mathcal{V})} < \infty$, then $\|R_\alpha f\|_\mathcal{V} \to +\infty$ for all $f \notin \mathcal{D}(A^\dagger)$.

The Banach Steinhaus Theorem

These results are consequences of the Banach Steinhaus theorem:

Theorem 4 (Banach Steinhaus)

Let $\mathcal{U}, \mathcal{V}$ be Hilbert spaces and let $\{A_j\}_{j \in \mathbb{N}} \subset \mathcal{L}(\mathcal{U}, \mathcal{V})$ be a family of pointwise bounded operators such that $\sup_{j \in \mathbb{N}} \|A_j u\|_{\mathcal{V}} \leqslant C(u)$ for all $u \in \mathcal{U}$. Then $\sup_{j \in \mathbb{N}} \|A_j\|_{\mathcal{L}(\mathcal{U}, \mathcal{V})} < \infty$.

A corollary of this is that the following are equivalent

(a) There exists $A \in \mathcal{L}(\mathcal{U}, \mathcal{V})$ such that $Au = \lim_{j \to \infty} A_j u$ for all $u \in \mathcal{U}$.

(b) There is a dense subset $\mathcal{X} \subset \mathcal{U}$ such that $\lim_{j \to \infty} A_j u$ for all $u \in \mathcal{X}$ and $\sup_{j \in \mathbb{N}} \|A_j\|_{\mathcal{L}(\mathcal{U}, \mathcal{V})} < \infty$.

No uniform boundedness of regularisers

Suppose $\{R_\alpha\}$ are linear regularisers. If $A^\dagger$ is not continuous, then $\{R_\alpha\}_\alpha$ cannot be uniformly bounded. In particular, there exists $f \in \mathcal{V}$ such that $\|R_\alpha f\| \to \infty$ as $\alpha \to 0$.

Proof of Theorem 2.

- Suppose that $\{R_\alpha\}_\alpha$ is uniformly bounded, then since $\lim_{\alpha \to 0} R_\alpha(u) = A^\dagger u$ exists for all $u \in \mathcal{D}(A^\dagger)$ which is dense in $\mathcal{V}$, we have $A^\dagger \in \mathcal{L}(\mathcal{U}, \mathcal{V})$, which contradicts the discontinuity of $A^\dagger$.

- If there does not exist f such that $\|R_\alpha f\|$ is unbounded, then Banach Steinhaus says that $\{R_\alpha\}_\alpha$ is uniformly bounded in norm. Contradiction.

$\square$

No uniform boundedness of regularisers

Suppose $\{R_\alpha\}$ are linear regularisers. If $\sup_{\alpha > 0} \|AR_\alpha\|_{\mathcal{L}(\mathcal{U}, \mathcal{V})} < \infty$, then $\|R_\alpha f\|_{\mathcal{V}} \to +\infty$ for all $f \notin \mathcal{D}(A^\dagger)$.

Proof of Theorem 3.

By Banach Steinhaus, since $\sup_{\alpha > 0} \|AR_\alpha\|_{\mathcal{L}(\mathcal{V}, \mathcal{V})} < \infty$ and $\mathcal{D}(A^\dagger)$ is dense in $\mathcal{V}$, there exists $B \in \mathcal{L}(\mathcal{V}, \mathcal{V})$ such that $Bg = \lim_{\alpha \to 0} AR_\alpha g$ for all $g \in \mathcal{V}$.

What can we say about B? For any $g \in \mathcal{D}(A^\dagger)$, $AR_\alpha g \to AA^\dagger g = P_{\overline{\mathcal{R}(A)}} g = Bg$. So, $B = P_{\overline{\mathcal{R}(A)}}$.

No uniform boundedness of regularisers

Suppose $\{R_\alpha\}$ are linear regularisers. If $\sup_{\alpha > 0} \|AR_\alpha\|_{\mathcal{L}(\mathcal{U}, \mathcal{V})} < \infty$, then $\|R_\alpha f\|_{\mathcal{V}} \to +\infty$ for all $f \notin \mathcal{D}(A^\dagger)$.

Proof of Theorem 3.

By Banach Steinhaus, since $\sup_{\alpha > 0} \|AR_\alpha\|_{\mathcal{L}(\mathcal{V}, \mathcal{V})} < \infty$ and $\mathcal{D}(A^\dagger)$ is dense in $\mathcal{V}$, there exists $B \in \mathcal{L}(\mathcal{V}, \mathcal{V})$ such that $Bg = \lim_{\alpha \to 0} AR_\alpha g$ for all $g \in \mathcal{V}$.

What can we say about B? For any $g \in \mathcal{D}(A^\dagger)$, $AR_\alpha g \to AA^\dagger g = P_{\overline{\mathcal{R}(A)}} g = Bg$. So, $B = P_{\overline{\mathcal{R}(A)}}$.

Assume that $f \notin \mathcal{D}(A^\dagger)$. Define $u_\alpha \stackrel{\text{def}}{=} R_\alpha f$. Suppose that $\|u_{\alpha_k}\|_{\mathcal{U}}$ is uniformly bounded, as $\alpha_k \to 0$

No uniform boundedness of regularisers

Suppose $\{R_\alpha\}$ are linear regularisers. If $\sup_{\alpha > 0} \|AR_\alpha\|_{\mathcal{L}(\mathcal{U}, \mathcal{V})} < \infty$, then $\|R_\alpha f\|_{\mathcal{V}} \to +\infty$ for all $f \notin \mathcal{D}(A^\dagger)$.

Proof of Theorem 3.

By Banach Steinhaus, since $\sup_{\alpha > 0} \|AR_\alpha\|_{\mathcal{L}(\mathcal{V}, \mathcal{V})} < \infty$ and $\mathcal{D}(A^\dagger)$ is dense in $\mathcal{V}$, there exists $B \in \mathcal{L}(\mathcal{V}, \mathcal{V})$ such that $Bg = \lim_{\alpha \to 0} AR_\alpha g$ for all $g \in \mathcal{V}$.

What can we say about B? For any $g \in \mathcal{D}(A^\dagger)$, $AR_\alpha g \to AA^\dagger g = P_{\overline{\mathcal{R}(A)}} g = Bg$. So, $B = P_{\overline{\mathcal{R}(A)}}$.

Assume that $f \notin \mathcal{D}(A^\dagger)$. Define $u_\alpha \stackrel{\text{def.}}{=} R_\alpha f$. Suppose that $\|u_{\alpha_k}\|_{\mathcal{U}}$ is uniformly bounded, as $\alpha_k \to 0$

- Since bounded sets in Hilbert spaces are weakly compact, there exists a weakly convergent subsequence $u_{\alpha_{k_n}}$ with weak limit $u \in \mathcal{U}$. Continuous linear operators are weakly continuous, so $Au_{\alpha_{k_n}} \rightharpoonup Au$.

No uniform boundedness of regularisers

Suppose $\{R_\alpha\}$ are linear regularisers. If $\sup_{\alpha>0} \|AR_\alpha\|_{\mathcal{L}(\mathcal{U},\mathcal{V})} < \infty$, then $\|R_\alpha f\|_{\mathcal{V}} \to +\infty$ for all $f \notin \mathcal{D}(A^\dagger)$.

Proof of Theorem 3.

By Banach Steinhaus, since $\sup_{\alpha>0} \|AR_\alpha\|_{\mathcal{L}(\mathcal{V},\mathcal{V})} < \infty$ and $\mathcal{D}(A^\dagger)$ is dense in $\mathcal{V}$, there exists $B \in \mathcal{L}(\mathcal{V},\mathcal{V})$ such that $Bg = \lim_{\alpha \to 0} AR_\alpha g$ for all $g \in \mathcal{V}$.

What can we say about B? For any $g \in \mathcal{D}(A^\dagger)$, $AR_\alpha g \to AA^\dagger g = P_{\overline{\mathcal{R}(A)}}g = Bg$. So, $B = P_{\overline{\mathcal{R}(A)}}$.

Assume that $f \notin \mathcal{D}(A^\dagger)$. Define $u_\alpha \stackrel{\text{def.}}{=} R_\alpha f$. Suppose that $\|u_{\alpha_k}\|_{\mathcal{U}}$ is uniformly bounded, as $\alpha_k \to 0$

- Since bounded sets in Hilbert spaces are weakly compact, there exists a weakly convergent subsequence $u_{\alpha_{k_n}}$ with weak limit $u \in \mathcal{U}$. Continuous linear operators are weakly continuous, so $Au_{\alpha_{k_n}} \rightharpoonup Au$.

- By uniqueness of limits, $Bf = P_{\overline{\mathcal{R}(A)}}f = \lim_{\alpha_{k_n}} AR_{\alpha_{k_n}}f = Au$.

No uniform boundedness of regularisers

Suppose $\{R_\alpha\}$ are linear regularisers. If $\sup_{\alpha>0} \|AR_\alpha\|_{\mathcal{L}(\mathcal{U},\mathcal{V})} < \infty$, then $\|R_\alpha f\|_{\mathcal{V}} \to +\infty$ for all $f \notin \mathcal{D}(A^\dagger)$.

Proof of Theorem 3.

By Banach Steinhaus, since $\sup_{\alpha>0} \|AR_\alpha\|_{\mathcal{L}(\mathcal{V},\mathcal{V})} < \infty$ and $\mathcal{D}(A^\dagger)$ is dense in $\mathcal{V}$, there exists $B \in \mathcal{L}(\mathcal{V},\mathcal{V})$ such that $Bg = \lim_{\alpha \to 0} AR_\alpha g$ for all $g \in \mathcal{V}$.

What can we say about B? For any $g \in \mathcal{D}(A^\dagger)$, $AR_\alpha g \to AA^\dagger g = P_{\overline{\mathcal{R}(A)}} g = Bg$. So, $B = P_{\overline{\mathcal{R}(A)}}$.

Assume that $f \notin \mathcal{D}(A^\dagger)$. Define $u_\alpha \stackrel{\text{def.}}{=} R_\alpha f$. Suppose that $\|u_{\alpha_k}\|_{\mathcal{U}}$ is uniformly bounded, as $\alpha_k \to 0$

- Since bounded sets in Hilbert spaces are weakly compact, there exists a weakly convergent subsequence $u_{\alpha_{k_n}}$ with weak limit $u \in \mathcal{U}$. Continuous linear operators are weakly continuous, so $Au_{\alpha_{k_n}} \rightharpoonup Au$.

- By uniqueness of limits, $Bf = P_{\overline{\mathcal{R}(A)}} f = \lim_{\alpha_{k_n}} AR_{\alpha_{k_n}} f = Au$.

- Since $\mathcal{V} = \overline{\mathcal{R}(A)} \oplus (\overline{\mathcal{R}(A)})^\perp$, we can write $f = f_1 + f_2$ where $f_1 \in \overline{\mathcal{R}(A)}$ and $f_2 \in \overline{\mathcal{R}(A)}^\perp$. So, $Au = P_{\overline{\mathcal{R}(A)}} f = f_1 \in \mathcal{R}(A)$ and hence, $f \in \mathcal{D}(A^\dagger)$. Contradiction. $\qquad\square$
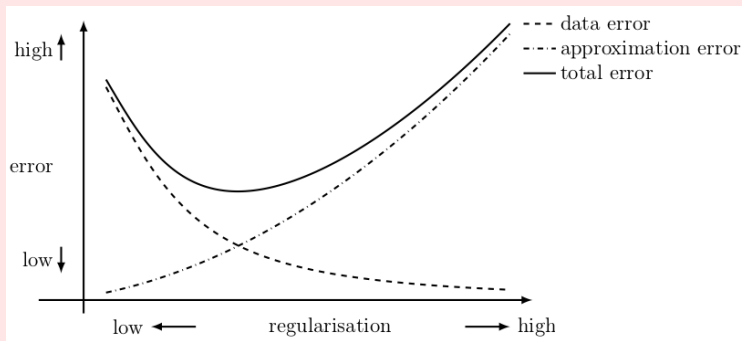
Parameter choice

Want: as $\delta \to 0$, $R_\alpha(f_\delta) \to A^\dagger f$, for all $f \in \mathcal{D}(A^\dagger)$ and f_δ s.t. $\|f - f_\delta\| \leqslant \delta$.

Parameter choice

Want: as $\delta \to 0$, $R_\alpha(f_\delta) \to A^\dagger f$, for all $f \in \mathcal{D}(A^\dagger)$ and f_δ s.t. $\|f - f_\delta\| \leqslant \delta$.

Let $u^\dagger = A^\dagger f$:

$$\left\| R_\alpha f_\delta - u^\dagger \right\|_{\mathcal{U}} \leqslant \left\| R_\alpha f_\delta - R_\alpha f \right\|_{\mathcal{U}} + \left\| R_\alpha f - u^\dagger \right\|_{\mathcal{U}} \leqslant \overbrace{\delta \left\| R_\alpha \right\|_{\mathcal{L}(\mathcal{V},\mathcal{U})}}^{\text{data error}} + \overbrace{\left\| R_\alpha f - A^\dagger f \right\|_{\mathcal{U}}}^{\text{approximation error}}$$

- The data error does not stay bounded as $\alpha \to 0$.
- The approximation error vanishes as $\alpha \to 0$.

Parameter choice

Want: as $\delta \to 0$, $R_\alpha(f_\delta) \to A^\dagger f$, for all $f \in \mathcal{D}(A^\dagger)$ and f_δ s.t. $\|f - f_\delta\| \leqslant \delta$.

Let $u^\dagger = A^\dagger f$:

$$\left\|R_\alpha f_\delta - u^\dagger\right\|_{\mathcal{U}} \leqslant \|R_\alpha f_\delta - R_\alpha f\|_{\mathcal{U}} + \left\|R_\alpha f - u^\dagger\right\|_{\mathcal{U}} \leqslant \overbrace{\delta \|R_\alpha\|_{\mathcal{L}(\mathcal{V},\mathcal{U})}}^{\text{data error}} + \overbrace{\left\|R_\alpha f - A^\dagger f\right\|_{\mathcal{U}}}^{\text{approximation error}}$$
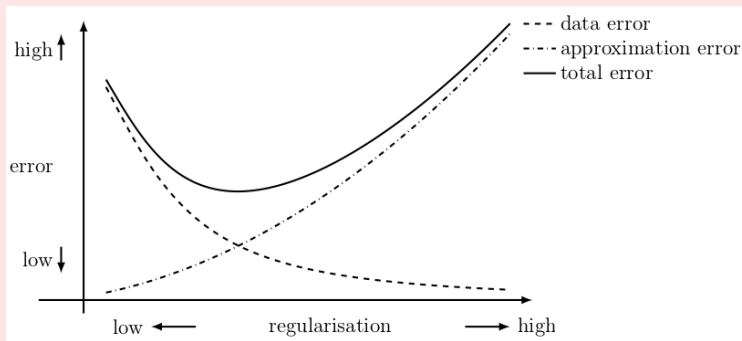

- --- data error
- -·-· approximation error
- —— total error

- The data error does not stay bounded as $\alpha \to 0$.
- The approximation error vanishes as $\alpha \to 0$.

The regularisation parameter need to be chosen to balance the two terms!

Definition 5

A function $\alpha : \mathbb{R}_{>0} \times \mathcal{V} \to \mathbb{R}_{>0}$, $(\delta, f_\delta) \to \alpha(\delta, f_\delta)$ is called a parameter choice rule. We distinguish between:

(a) a priori parameter choice rules which depend only on δ,

(b) a posteriori parameter choice rules which depend on both δ and f_δ.

(c) heuristic parameter choice rules which depend on f_δ only.

Let $\{R_\alpha\}_\alpha$ be a regularisation of $A^\dagger$. If for all $f \in \mathcal{D}(A^\dagger)$, α is a parameter choice rule such that

$$\lim_{\delta \to 0} \sup_{f_\delta : \|f - f_\delta\| \leqslant \delta} \left\| R_\alpha f_\delta - A^\dagger f \right\|_{\mathcal{U}} = 0 \quad \text{and} \quad \lim_{\delta \to 0} \sup_{f_\delta : \|f - f_\delta\| \leqslant \delta} \alpha(\delta, f_\delta) = 0,$$

then (R_α, α) is called a **convergent regularisation**.

A priori parameter choice rules

Theorem 6 (Existence of convergent a-priori parameter choice rules)

Let $\{R_\alpha\}_\alpha$ be a regularisation of $A^\dagger$, for $A \in \mathcal{L}(\mathcal{U}, \mathcal{V})$. Then, there exists an a-priori parameter choice rule $\alpha = \alpha(\delta)$ such that (R_α, α) is a convergent regularisation.

Proof.

A priori parameter choice rules

Theorem 6 (Existence of convergent a-priori parameter choice rules)

Let $\{R_\alpha\}_\alpha$ be a regularisation of $A^\dagger$, for $A \in \mathcal{L}(\mathcal{U}, \mathcal{V})$. Then, there exists an a-priori parameter choice rule $\alpha = \alpha(\delta)$ such that (R_α, α) is a convergent regularisation.

Proof.

- Fix $f \in \mathcal{D}(A^\dagger)$. Since $R_\alpha f \to A^\dagger f$, there exists a monotone increasing function $\gamma : (0, \infty) \to (0, \infty)$ such that $\lim_{\varepsilon \to 0} \gamma(\varepsilon) = 0$ and $\left\| R_{\gamma(\varepsilon)} f - A^\dagger f \right\|_\mathcal{U} \leqslant \frac{\varepsilon}{2}$.

A priori parameter choice rules

Theorem 6 (Existence of convergent a-priori parameter choice rules)

Let $\{R_\alpha\}_\alpha$ be a regularisation of $A^\dagger$, for $A \in \mathcal{L}(\mathcal{U}, \mathcal{V})$. Then, there exists an a-priori parameter choice rule $\alpha = \alpha(\delta)$ such that (R_α, α) is a convergent regularisation.

Proof.

- Fix $f \in \mathcal{D}(A^\dagger)$. Since $R_\alpha f \to A^\dagger f$, there exists a monotone increasing function $\gamma : (0, \infty) \to (0, \infty)$ such that $\lim_{\varepsilon \to 0} \gamma(\varepsilon) = 0$ and $\left\| R_{\gamma(\varepsilon)} f - A^\dagger f \right\|_{\mathcal{U}} \leqslant \frac{\varepsilon}{2}$.

- For fixed ε, $R_{\gamma(\varepsilon)}$ is continuous. So, there exists $\rho(\varepsilon) > 0$ such that $\left\| R_{\gamma(\varepsilon)} f - R_{\gamma(\varepsilon)} g \right\| \leqslant \frac{\varepsilon}{2}$ for all $\|g - f\| \leqslant \rho(\varepsilon)$. WLOG, assume that ρ is a continuous strictly monotone increasing function with $\lim_{\varepsilon \to 0} \rho(\varepsilon) = 0$.

A priori parameter choice rules

Theorem 6 (Existence of convergent a-priori parameter choice rules)

Let $\{R_\alpha\}_\alpha$ be a regularisation of $A^\dagger$, for $A \in \mathcal{L}(\mathcal{U}, \mathcal{V})$. Then, there exists an a-priori parameter choice rule $\alpha = \alpha(\delta)$ such that (R_α, α) is a convergent regularisation.

Proof.

- Fix $f \in \mathcal{D}(A^\dagger)$. Since $R_\alpha f \to A^\dagger f$, there exists a monotone increasing function $\gamma : (0, \infty) \to (0, \infty)$ such that $\lim_{\varepsilon \to 0} \gamma(\varepsilon) = 0$ and $\left\| R_{\gamma(\varepsilon)} f - A^\dagger f \right\|_{\mathcal{U}} \leqslant \frac{\varepsilon}{2}$.

- For fixed ε, $R_{\gamma(\varepsilon)}$ is continuous. So, there exists $\rho(\varepsilon) > 0$ such that $\left\| R_{\gamma(\varepsilon)} f - R_{\gamma(\varepsilon)} g \right\| \leqslant \frac{\varepsilon}{2}$ for all $\|g - f\| \leqslant \rho(\varepsilon)$. WLOG, assume that ρ is a continuous strictly monotone increasing function with $\lim_{\varepsilon \to 0} \rho(\varepsilon) = 0$.

- By the inverse function theorem, there exists ρ^{-1} which is also strictly monotone and continuous on the range of ρ such that $\lim_{\delta \to 0} \rho^{-1}(\delta) = 0$. Continuously extend ρ^{-1} to $(0, \infty)$ and define $\alpha(\delta) \overset{\text{def.}}{=} \gamma(\rho^{-1}(\delta))$. Then, $\lim_{\delta \to 0} \alpha(\delta) = 0$.

A priori parameter choice rules

Theorem 6 (Existence of convergent a-priori parameter choice rules)

Let $\{R_\alpha\}_\alpha$ be a regularisation of $A^\dagger$, for $A \in \mathcal{L}(\mathcal{U}, \mathcal{V})$. Then, there exists an a-priori parameter choice rule $\alpha = \alpha(\delta)$ such that (R_α, α) is a convergent regularisation.

Proof.

- Fix $f \in \mathcal{D}(A^\dagger)$. Since $R_\alpha f \to A^\dagger f$, there exists a monotone increasing function $\gamma : (0, \infty) \to (0, \infty)$ such that $\lim_{\varepsilon \to 0} \gamma(\varepsilon) = 0$ and $\left\|R_{\gamma(\varepsilon)}f - A^\dagger f\right\|_\mathcal{U} \leqslant \frac{\varepsilon}{2}$.

- For fixed ε, $R_{\gamma(\varepsilon)}$ is continuous. So, there exists $\rho(\varepsilon) > 0$ such that $\left\|R_{\gamma(\varepsilon)}f - R_{\gamma(\varepsilon)}g\right\| \leqslant \frac{\varepsilon}{2}$ for all $\|g - f\| \leqslant \rho(\varepsilon)$. WLOG, assume that ρ is a continuous strictly monotone increasing function with $\lim_{\varepsilon \to 0} \rho(\varepsilon) = 0$.

- By the inverse function theorem, there exists ρ^{-1} which is also strictly monotone and continuous on the range of ρ such that $\lim_{\delta \to 0} \rho^{-1}(\delta) = 0$. Continuously extend ρ^{-1} to $(0, \infty)$ and define $\alpha(\delta) \stackrel{\text{def.}}{=} \gamma(\rho^{-1}(\delta))$. Then, $\lim_{\delta \to 0} \alpha(\delta) = 0$.

- For all $\varepsilon > 0$, there exists $\delta \stackrel{\text{def.}}{=} \rho(\varepsilon)$ such that with $\alpha(\delta) = \gamma(\varepsilon)$,

$$\left\|R_{\alpha(\delta)}f_\delta - A^\dagger f\right\|_\mathcal{U} \leqslant \left\|R_{\alpha(\delta)}f_\delta - R_{\alpha(\delta)}f\right\| + \left\|R_{\gamma(\varepsilon)}f - A^\dagger f\right\|_\mathcal{U} \leqslant \varepsilon,$$

  for all f_δ with $\|f - f_\delta\| \leqslant \delta$.

$\square$

A priori parameter choice rules for linear regularisers

Theorem 7

Let $\{R_\alpha\}_\alpha$ be a linear regularisation and α be an a priori parameter choice rule. Then, (R_α, α) is a convergent regularisation if and only if

(a) $\alpha(\delta) \to 0$ as $\delta \to 0$.

(b) $\lim_{\delta \to 0} \delta \left\| R_{\alpha(\delta)} \right\| = 0$.

A priori parameter choice rules for linear regularisers

Theorem 7

Let $\{R_\alpha\}_\alpha$ be a linear regularisation and α be an a priori parameter choice rule. Then, (R_α, α) is a convergent regularisation if and only if

(a) $\alpha(\delta) \to 0$ as $\delta \to 0$.

(b) $\lim_{\delta \to 0} \delta \left\| R_{\alpha(\delta)} \right\| = 0$.

Clearly, if (a) and (b) are satisfied, then for all $\|f_\delta - f\| \leqslant \delta$,
$\left\| R_{\alpha(\delta)} f_\delta - A^\dagger f \right\| \leqslant \delta \left\| R_{\alpha(\delta)} \right\| + \left\| R_{\alpha(\delta)} A - A^\dagger f \right\| \to 0$ as $\delta \to 0$.

A priori parameter choice rules for linear regularisers

Theorem 7

Let $\{R_\alpha\}_\alpha$ be a linear regularisation and α be an a priori parameter choice rule. Then, (R_α, α) is a convergent regularisation if and only if

(a) $\alpha(\delta) \to 0$ as $\delta \to 0$.

(b) $\lim_{\delta \to 0} \delta \left\| R_{\alpha(\delta)} \right\| = 0$.

Clearly, if (a) and (b) are satisfied, then for all $\|f_\delta - f\| \leqslant \delta$,
$\left\| R_{\alpha(\delta)} f_\delta - A^\dagger f \right\| \leqslant \delta \left\| R_{\alpha(\delta)} \right\| + \left\| R_{\alpha(\delta)} A - A^\dagger f \right\| \to 0$ as $\delta \to 0$.

If (R_α, α) is a convergent regularisation, then (a) holds by definition. Suppose $\lim_{\delta \to 0} \delta \left\| R_{\alpha(\delta)} \right\| \neq 0$.

A priori parameter choice rules for linear regularisers

Theorem 7

Let $\{R_\alpha\}_\alpha$ be a linear regularisation and α be an a priori parameter choice rule. Then, (R_α, α) is a convergent regularisation if and only if

(a) $\alpha(\delta) \to 0$ as $\delta \to 0$.

(b) $\lim_{\delta \to 0} \delta \left\| R_{\alpha(\delta)} \right\| = 0$.

Clearly, if (a) and (b) are satisfied, then for all $\|f_\delta - f\| \leqslant \delta$,
$\left\| R_{\alpha(\delta)} f_\delta - A^\dagger f \right\| \leqslant \delta \left\| R_{\alpha(\delta)} \right\| + \left\| R_{\alpha(\delta)} A - A^\dagger f \right\| \to 0$ as $\delta \to 0$.

If (R_α, α) is a convergent regularisation, then (a) holds by definition. Suppose $\lim_{\delta \to 0} \delta \left\| R_{\alpha(\delta)} \right\| \neq 0$.

This means: taking a sequence $\delta_k \to 0$, for each δ_k, there exists $c > 0$ and f_{δ_k} with $\left\| f_{\delta_k} - f \right\| \leqslant \delta_k$ such that $\left\| R_{\alpha(\delta_k)}(f_{\delta_k} - f) \right\| \geqslant c$.

A priori parameter choice rules for linear regularisers

Theorem 7

Let $\{R_\alpha\}_\alpha$ be a linear regularisation and α be an a priori parameter choice rule. Then, (R_α, α) is a convergent regularisation if and only if

(a) $\alpha(\delta) \to 0$ as $\delta \to 0$.

(b) $\lim_{\delta \to 0} \delta \left\| R_{\alpha(\delta)} \right\| = 0$.

Clearly, if (a) and (b) are satisfied, then for all $\|f_\delta - f\| \leqslant \delta$,
$$\left\| R_{\alpha(\delta)} f_\delta - A^\dagger f \right\| \leqslant \delta \left\| R_{\alpha(\delta)} \right\| + \left\| R_{\alpha(\delta)} A - A^\dagger f \right\| \to 0 \text{ as } \delta \to 0.$$

If (R_α, α) is a convergent regularisation, then (a) holds by definition. Suppose $\lim_{\delta \to 0} \delta \left\| R_{\alpha(\delta)} \right\| \neq 0$.

This means: taking a sequence $\delta_k \to 0$, for each δ_k, there exists $c > 0$ and f_{δ_k} with $\left\| f_{\delta_k} - f \right\| \leqslant \delta_k$ such that $\left\| R_{\alpha(\delta_k)}(f_{\delta_k} - f) \right\| \geqslant c$.

So:
$$\left\| R_{\alpha(\delta_k)} f_{\delta_k} - A^\dagger f \right\| \geqslant \left\| R_{\alpha(\delta_k)}(f_{\delta_k} - f) \right\| - \left\| R_{\alpha(\delta_k)} - A^\dagger f \right\| \geqslant c - \left\| R_{\alpha(\delta_k)} - A^\dagger f \right\|.$$

Contradiction, since the LHS converges to 0, and the RHS converges to c.

Outline

Recall that the spectral representation of $A^\dagger$:

$$A^\dagger f = \sum_{j \in \mathbb{N}} \frac{1}{\sigma_j} \langle f, v_j \rangle u_j$$

The source of ill-posedness of $A^\dagger$ is that the eigenvalues $1/\sigma_j$ explode as $j \to \infty$.

Spectral regularisation modifies the eigenvalues

$$R_\alpha f = \sum_{j \in \mathbb{N}} g_\alpha(\sigma_j) \langle f, v_j \rangle u_j$$

where $g_\alpha : (0, \infty) \to (0, \infty)$ such that

1. $\lim_{\alpha \to 0} g_\alpha(\sigma) = \frac{1}{\sigma}$
2. for all $\sigma > 0$, $g_\alpha(\sigma) \leqslant C_\alpha$.

Spectral regularisation

Theorem 8 (Mild growth of g_α will ensures regularisation of $A^\dagger$)

Assume that $\sup_{\alpha,\sigma} \sigma g_\alpha(\sigma) \leqslant \gamma$ for some $\gamma > 0$.

- Then, $R_\alpha f \to A^\dagger f$ as $\alpha \to 0$ for all $f \in \mathcal{D}(A^\dagger)$.
- If moreover, $\alpha = \alpha(\delta)$ is an a-priori parameter choice rule such that $\lim_{\delta \to 0} \delta C_{\alpha(\delta)} = 0$, then $(R_{\alpha(\delta)}, \alpha(\delta))$ is a convergent regularisation.

Spectral regularisation

Theorem 8 (Mild growth of g_α will ensures regularisation of $A^\dagger$)

Assume that $\sup_{\alpha,\sigma} \sigma g_\alpha(\sigma) \leqslant \gamma$ for some $\gamma > 0$.

- Then, $R_\alpha f \to A^\dagger f$ as $\alpha \to 0$ for all $f \in \mathcal{D}(A^\dagger)$.
- If moreover, $\alpha = \alpha(\delta)$ is an a-priori parameter choice rule such that $\lim_{\delta \to 0} \delta C_{\alpha(\delta)} = 0$, then $(R_{\alpha(\delta)}, \alpha(\delta))$ is a convergent regularisation.

From the SVD of $A^\dagger$ and definition of R_α,

$$R_\alpha f - A^\dagger f = \sum_{j=1}^\infty \left(g_\alpha(\sigma_j) - \frac{1}{\sigma_j} \right) \langle f, v_j \rangle u_j = \sum_{j=1}^\infty \left(\sigma_j g_\alpha(\sigma_j) - 1 \right) \langle A^\dagger f, u_j \rangle u_j$$

By assumption, $\left| \sigma_j g_\alpha(\sigma_j) - 1 \right| \leqslant 1 + \gamma$. So, $\left\| R_\alpha f - A^\dagger f \right\| \leqslant (1+\gamma) \left\| A^\dagger f \right\| < \infty$.

Spectral regularisation

Theorem 8 (Mild growth of g_α will ensures regularisation of $A^\dagger$)

Assume that $\sup_{\alpha,\sigma} \sigma g_\alpha(\sigma) \leqslant \gamma$ for some $\gamma > 0$.

- *Then, $R_\alpha f \to A^\dagger f$ as $\alpha \to 0$ for all $f \in \mathcal{D}(A^\dagger)$.*
- *If moreover, $\alpha = \alpha(\delta)$ is an a-priori parameter choice rule such that $\lim_{\delta \to 0} \delta C_{\alpha(\delta)} = 0$, then $(R_{\alpha(\delta)}, \alpha(\delta))$ is a convergent regularisation.*

From the SVD of $A^\dagger$ and definition of R_α,

$$R_\alpha f - A^\dagger f = \sum_{j=1}^{\infty} \left(g_\alpha(\sigma_j) - \frac{1}{\sigma_j} \right) \langle f, v_j \rangle u_j = \sum_{j=1}^{\infty} \left(\sigma_j g_\alpha(\sigma_j) - 1 \right) \langle A^\dagger f, u_j \rangle u_j$$

By assumption, $\left| \sigma_j g_\alpha(\sigma_j) - 1 \right| \leqslant 1 + \gamma$. So, $\left\| R_\alpha f - A^\dagger f \right\| \leqslant (1+\gamma) \left\| A^\dagger f \right\| < \infty$.

By the reverse Fatou's lemma

$$\limsup_{\alpha \to 0} \left\| R_\alpha f - A^\dagger f \right\|^2 \leqslant \limsup_{\alpha \to 0} \sum_j \left(\sigma_j g_\alpha(\sigma_j) - 1 \right)^2 \left| \langle A^\dagger f, u_j \rangle \right|^2$$

$$\leqslant \sum_j \limsup_{\alpha \to 0} \left(\sigma_j g_\alpha(\sigma_j) - 1 \right)^2 \left| \langle A^\dagger f, u_j \rangle \right|^2 = 0.$$

Spectral regularisation

Theorem 8 (Mild growth of g_α will ensures regularisation of $A^\dagger$)

Assume that $\sup_{\alpha,\sigma} \sigma g_\alpha(\sigma) \leqslant \gamma$ for some $\gamma > 0$.

- Then, $R_\alpha f \to A^\dagger f$ as $\alpha \to 0$ for all $f \in \mathcal{D}(A^\dagger)$.
- If moreover, $\alpha = \alpha(\delta)$ is an a-priori parameter choice rule such that $\lim_{\delta \to 0} \delta C_{\alpha(\delta)} = 0$, then $(R_{\alpha(\delta)}, \alpha(\delta))$ is a convergent regularisation.

From the SVD of $A^\dagger$ and definition of R_α,

$$R_\alpha f - A^\dagger f = \sum_{j=1}^{\infty} \left(g_\alpha(\sigma_j) - \frac{1}{\sigma_j} \right) \langle f, v_j \rangle u_j = \sum_{j=1}^{\infty} \left(\sigma_j g_\alpha(\sigma_j) - 1 \right) \langle A^\dagger f, u_j \rangle u_j$$

By assumption, $\left| \sigma_j g_\alpha(\sigma_j) - 1 \right| \leqslant 1 + \gamma$. So, $\left\| R_\alpha f - A^\dagger f \right\| \leqslant (1 + \gamma) \left\| A^\dagger f \right\| < \infty$.

By the reverse Fatou's lemma

$$\limsup_{\alpha \to 0} \left\| R_\alpha f - A^\dagger f \right\|^2 \leqslant \limsup_{\alpha \to 0} \sum_j \left(\sigma_j g_\alpha(\sigma_j) - 1 \right)^2 \left| \langle A^\dagger f, u_j \rangle \right|^2$$

$$\leqslant \sum_j \limsup_{\alpha \to 0} \left(\sigma_j g_\alpha(\sigma_j) - 1 \right)^2 \left| \langle A^\dagger f, u_j \rangle \right|^2 = 0.$$

The claim on convergent regularisation is from Theorem 7 since $\left\| R_{\alpha(\delta)} \right\| \leqslant C_{\alpha(\delta)}$.

Truncated singular value decomposition

Truncated SVD: discard all singular values below threshold α:

$$g_\alpha(\sigma) = \begin{cases} \sigma^{-1} & \sigma \geqslant \alpha \\ 0 & \text{else.} \end{cases}$$

We then have

$$R_\alpha f = \sum_{\sigma_j \geqslant \alpha} \frac{1}{\sigma_j} \langle f, v_j \rangle u_j$$

This is always well-defined for compact operators (zero is the only accumulation point of singular values).

Is this a convergent regularisation of $A^\dagger$?
For all $\sigma > 0$, we naturally have $\lim_{\alpha \to 0} g_\alpha(\sigma) = \frac{1}{\sigma}$.
We have $\sup_{\sigma, \alpha} \sigma g_\alpha(\sigma) = 1$ and $C_\alpha = \alpha^{-1}$.
So, the truncated SVD is a convergent regularisation if $\lim_{\delta \to 0} \frac{\delta}{\alpha(\delta)} = 0$.

Disadvantage: requires the knowledge of the singular vectors of A (only finitely many, but this number might be large).

The idea is to shift the eigenvalues of A^*A by a constant factor: Let $g_\alpha(\sigma) = \frac{\sigma}{\sigma^2+\alpha}$ and the corresponding Tikhonov regularisation is

$$R_\alpha f = \sum_{j=1}^{\infty} \frac{\sigma_j}{\sigma_j^2 + \alpha} \langle f, \, v_j \rangle u_j.$$

Tikhonov regularisation

The idea is to shift the eigenvalues of A^*A by a constant factor: Let $g_\alpha(\sigma) = \frac{\sigma}{\sigma^2 + \alpha}$ and the corresponding Tikhonov regularisation is

$$R_\alpha f = \sum_{j=1}^\infty \frac{\sigma_j}{\sigma_j^2 + \alpha} \langle f, \, v_j \rangle u_j.$$

Is this a convergent regularisation of $A^\dagger$?
For all $\sigma > 0$, $\lim_{\alpha \to 0} g(\sigma) = \frac{1}{\sigma}$. We again have $\sup_{\alpha, \sigma} \sigma g_\alpha \leqslant 1$ and since $0 \leqslant (\sigma - \sqrt{\alpha})^2 = \sigma^2 - 2\sigma\sqrt{\alpha} + \alpha$, we get $\sigma^2 + \alpha \geqslant 2\sigma\sqrt{\alpha}$ which implies that

$$\frac{\sigma}{\sigma^2 + \alpha} \leqslant \frac{1}{2\sqrt{\alpha}}.$$

So, $g_\alpha(\sigma) \leqslant C_\alpha = \frac{1}{2\sqrt{\alpha}}$. We have a convergent regularisation if

$$\lim_{\delta \to 0} \delta/\sqrt{\alpha} = 0.$$

Tikhonov regularisation

The idea is to shift the eigenvalues of A^*A by a constant factor: Let $g_\alpha(\sigma) = \frac{\sigma}{\sigma^2 + \alpha}$ and the corresponding Tikhonov regularisation is

$$R_\alpha f = \sum_{j=1}^{\infty} \frac{\sigma_j}{\sigma_j^2 + \alpha} \langle f, \ v_j \rangle u_j.$$

This is easy to compute:
Note that σ_j^2 are the eigenvalues of A^*A and $\sigma_j^2 + \alpha$ are the eigenvalues of $A^*A + \alpha\mathrm{Id}$. So, for $u_\alpha = R_\alpha f$, we have

$$(A^*A + \alpha\mathrm{Id})u_\alpha = \sum_{j=1}^{\infty} \sigma_j \langle f, \ v_j \rangle u_j = A^*f.$$

So, we just need to invert this (well-posed) linear system. Knowledge of σ_j's not needed!

Convergence rates

Let's look at the error between $u_\alpha = R_\alpha f$ and $u_\alpha^\delta = R_\alpha f_\delta$, where $f \in \mathcal{D}(A^\dagger)$ and $f^\delta \in \mathcal{V}$ satisfies $\left\| f - f^\delta \right\| \leqslant \delta$ for some $\delta > 0$:

Theorem 9

Assume that $\sup_{\alpha,\sigma} \sigma g_\alpha(\sigma) \leqslant \gamma$. Then

$$\left\| Au_\alpha - Au_\alpha^\delta \right\|_{\mathcal{V}} \leqslant \gamma\delta \quad \text{and} \quad \left\| u_\alpha - u_\alpha^\delta \right\|_{\mathcal{U}} \leqslant C_\alpha\delta.$$

Convergence rates

Let's look at the error between $u_\alpha = R_\alpha f$ and $u_\alpha^\delta = R_\alpha f_\delta$, where $f \in \mathcal{D}(A^\dagger)$ and $f^\delta \in \mathcal{V}$ satisfies $\left\| f - f^\delta \right\| \leqslant \delta$ for some $\delta > 0$:

Theorem 9

Assume that $\sup_{\alpha,\sigma} \sigma g_\alpha(\sigma) \leqslant \gamma$. Then

$$\left\| Au_\alpha - Au_\alpha^\delta \right\|_{\mathcal{V}} \leqslant \gamma\delta \quad and \quad \left\| u_\alpha - u_\alpha^\delta \right\|_{\mathcal{U}} \leqslant C_\alpha\delta.$$

NB: the error on $\left\| Au_\alpha - Au_\alpha^\delta \right\|_{\mathcal{V}}$ is linear wrt δ, but since α depends on δ, the error on $\left\| u_\alpha - u_\alpha^\delta \right\|_{\mathcal{U}}$ will be slower than $\mathcal{O}(\delta)$.

Convergence rates

Let's look at the error between $u_\alpha = R_\alpha f$ and $u_\alpha^\delta = R_\alpha f_\delta$, where $f \in \mathcal{D}(A^\dagger)$ and $f^\delta \in \mathcal{V}$ satisfies $\left\| f - f^\delta \right\| \leqslant \delta$ for some $\delta > 0$:

Theorem 9

Assume that $\sup_{\alpha, \sigma} \sigma g_\alpha(\sigma) \leqslant \gamma$. Then

$$\left\| A u_\alpha - A u_\alpha^\delta \right\|_\mathcal{V} \leqslant \gamma \delta \quad and \quad \left\| u_\alpha - u_\alpha^\delta \right\|_\mathcal{U} \leqslant C_\alpha \delta.$$

Proof:

$$A u_\alpha - A u_\alpha^\delta = \sum_j \sigma_j \langle u_\alpha - u_\alpha^\delta,\, u_j \rangle v_j = \sum_j \sigma_j g_\alpha(\sigma_j) \langle f - f_\delta,\, v_j \rangle v_j$$

So, $\left\| A u_\alpha - A u_\alpha^\delta \right\| \leqslant \sup_{\sigma, \alpha} \sigma g_\alpha(\sigma) \left\| f - f_\delta \right\| \leqslant \gamma \delta$.

Similarly, recall that $|g_\alpha(\sigma)| \leqslant C_\alpha$ for all $\sigma > 0$, so

$$u_\alpha - u_\alpha^\delta = \sum_j g_\alpha(\sigma_j) \langle f - f_\delta,\, v_j \rangle u_j$$

implies $\left\| u_\alpha - u_\alpha^\delta \right\| \leqslant C_\alpha \delta$.

We have bounded the data error, but for the approximation error $\left\| u^\dagger - u_\alpha \right\|$ where $u_\alpha = R_\alpha f$ and $u^\dagger = A^\dagger f$, this depends on additional properties of $u^\dagger$:

The source condition: There exists $w \in \mathcal{U}$ and $\mu > 0$ such that $u^\dagger = (K^*K)^\mu w$. For arbitrary $\mu > 0$, this is interpreted as

$$(A^*A)^\mu w = \sum_{j=1}^\infty \sigma_j^{2\mu} \langle w,\, u_j \rangle u_j.$$

We have bounded the data error, but for the approximation error $\|u^\dagger - u_\alpha\|$ where $u_\alpha = R_\alpha f$ and $u^\dagger = A^\dagger f$, this depends on additional properties of $u^\dagger$:

The source condition: There exists $w \in \mathcal{U}$ and $\mu > 0$ such that $u^\dagger = (K^*K)^\mu w$. For arbitrary $\mu > 0$, this is interpreted as

$$(A^*A)^\mu w = \sum_{j=1}^\infty \sigma_j^{2\mu} \langle w, u_j \rangle u_j.$$

Example (Differentiation): Let $(Au)(y) = \int_0^y u(x)\mathrm{d}x$. For $\mu = 1$, the source condition says that

$$u^\dagger(x) = \int_x^1 \int_0^y w(z)\mathrm{d}x\mathrm{d}y$$

i.e. $u^\dagger$ is twice differentiable.

The source condition

Assume the source condition and $\sigma^{2\mu} |\sigma g_\alpha(\sigma) - 1| \leqslant \omega_\mu(\alpha)$ for all $\sigma > 0$.

Then

$$\left\| R_\alpha f - A^\dagger f \right\|^2 \leqslant \sum_j \left| \sigma_j g_\alpha(\sigma_j) - 1 \right|^2 \underbrace{\left| \langle u^\dagger, u_j \rangle \right|^2}_{\sigma_j^{4\mu} |\langle w, u_j \rangle|^2}$$

The source condition

Assume the source condition and $\sigma^{2\mu} |\sigma g_\alpha(\sigma) - 1| \leqslant \omega_\mu(\alpha)$ for all $\sigma > 0$.

Then

$$\left\| R_\alpha f - A^\dagger f \right\|^2 \leqslant \sum_j \left| \sigma_j g_\alpha(\sigma_j) - 1 \right|^2 \underbrace{\left| \langle u^\dagger, u_j \rangle \right|^2}_{\sigma_j^{4\mu} |\langle w, u_j \rangle|^2}$$

So, $\left\| u_\alpha - u^\dagger \right\| \leqslant \omega_\mu(\alpha) \|w\|$ and

$$\left\| u_\alpha^\delta - u^\dagger \right\| \leqslant \omega_\mu(\alpha) \|w\| + C_\alpha \delta.$$

The source condition

Assume the source condition and $\sigma^{2\mu} |\sigma g_\alpha(\sigma) - 1| \leqslant \omega_\mu(\alpha)$ for all $\sigma > 0$.

Then

$$\left\| R_\alpha f - A^\dagger f \right\|^2 \leqslant \sum_j |\sigma_j g_\alpha(\sigma_j) - 1|^2 \underbrace{\left| \langle u^\dagger, u_j \rangle \right|^2}_{\sigma_j^{4\mu} |\langle w, u_j \rangle|^2}$$

So, $\left\| u_\alpha - u^\dagger \right\| \leqslant \omega_\mu(\alpha) \|w\|$ and

$$\left\| u_\alpha^\delta - u^\dagger \right\| \leqslant \omega_\mu(\alpha) \|w\| + C_\alpha \delta.$$

Example of Truncated SVD.
Recall $C_\alpha = 1/\alpha$ and $g_\alpha(\sigma) = 0$ for all $\sigma < \alpha$ and $g_\alpha(\sigma) = 1/\sigma$ for $\sigma \geqslant \alpha$. So,

$$\sigma^{2\mu} |\sigma g_\alpha(\sigma) - 1| \leqslant \alpha^{2\mu}.$$

Let $\omega_\mu(\alpha) = \alpha^{2\mu}$. To minimise $\alpha^{2\mu} \|w\| + \delta/\alpha$, choose $\alpha = \left(\frac{\delta}{2\mu \|w\|} \right)^{1/(2\mu+1)}$

This yields $\left\| u_\alpha^\delta - u^\dagger \right\| \leqslant \delta^{\frac{2\mu}{2\mu+1}}$.

The source condition

Assume the source condition and $\sigma^{2\mu} |\sigma g_\alpha(\sigma) - 1| \leqslant \omega_\mu(\alpha)$ for all $\sigma > 0$.

Then

$$\left\| R_\alpha f - A^\dagger f \right\|^2 \leqslant \sum_j \left| \sigma_j g_\alpha(\sigma_j) - 1 \right|^2 \underbrace{\left| \langle u^\dagger, u_j \rangle \right|^2}_{\sigma_j^{4\mu} |\langle w, u_j \rangle|^2}$$

So, $\left\| u_\alpha - u^\dagger \right\| \leqslant \omega_\mu(\alpha) \|w\|$ and

$$\left\| u_\alpha^\delta - u^\dagger \right\| \leqslant \omega_\mu(\alpha) \|w\| + C_\alpha \delta.$$

Example of Truncated SVD.
Recall $C_\alpha = 1/\alpha$ and $g_\alpha(\sigma) = 0$ for all $\sigma < \alpha$ and $g_\alpha(\sigma) = 1/\sigma$ for $\sigma \geqslant \alpha$. So,

$$\sigma^{2\mu} |\sigma g_\alpha(\sigma) - 1| \leqslant \alpha^{2\mu}.$$

Let $\omega_\mu(\alpha) = \alpha^{2\mu}$. To minimise $\alpha^{2\mu} \|w\| + \delta/\alpha$, choose $\alpha = \left(\frac{\delta}{2\mu \|w\|} \right)^{1/(2\mu+1)}$

This yields $\left\| u_\alpha^\delta - u^\dagger \right\| \leqslant \delta^{\frac{2\mu}{2\mu+1}}$.

Whatever the choice of μ, the convergence rate is always slower than $\mathcal{O}(\delta)$. One can show that this rate is optimal.

A posteriori parameter choice rules

We may want a parameter choice rule which takes the approximate data f_δ into account. One way of approaching this is via the Morozov's discrepancy principle:

Definition 10

Let $u_\alpha = R_\alpha f_\delta$ with $\alpha(\delta, f_\delta)$ chosen as follows:

$$\alpha(\delta, f_\delta) = \sup \left\{ \alpha > 0 \; ; \; \left\| A u_{\alpha(\delta, f_\delta)} - f_\delta \right\| \leqslant \eta \delta \right\}$$

where $\eta > 1$. Then, $u_{\alpha(\delta, f_\delta)}$ is said to satisfy Morozov's discrepancy principle.

A posteriori parameter choice rules

We may want a parameter choice rule which takes the approximate data f_δ into account. One way of approaching this is via the Morozov's discrepancy principle:

Definition 10

Let $u_\alpha = R_\alpha f_\delta$ with $\alpha(\delta, f_\delta)$ chosen as follows:

$$\alpha(\delta, f_\delta) = \sup \left\{ \alpha > 0 \; ; \; \left\| A u_{\alpha(\delta, f_\delta)} - f_\delta \right\| \leqslant \eta \delta \right\}$$

where $\eta > 1$. Then, $u_{\alpha(\delta, f_\delta)}$ is said to satisfy Morozov's discrepancy principle.

- It can be shown that the a-posteriori parameter choice rule indeed yields a convergent regularisation method.

A posteriori parameter choice rules

We may want a parameter choice rule which takes the approximate data f_δ into account. One way of approaching this is via the Morozov's discrepancy principle:

Definition 10

Let $u_\alpha = R_\alpha f_\delta$ with $\alpha(\delta, f_\delta)$ chosen as follows:

$$\alpha(\delta, f_\delta) = \sup \left\{ \alpha > 0 \; ; \; \left\| A u_{\alpha(\delta, f_\delta)} - f_\delta \right\| \leqslant \eta \delta \right\}$$

where $\eta > 1$. Then, $u_{\alpha(\delta, f_\delta)}$ is said to satisfy Morozov's discrepancy principle.

- It can be shown that the a-posteriori parameter choice rule indeed yields a convergent regularisation method.

- Given $f \in \mathcal{D}(A^\dagger)$ and $f_\delta \in \mathcal{V}$ such that $\|f_\delta - f\| \leqslant \delta$, let $u^\dagger$ be the minimal norm solution to data f and define $\mu \overset{\text{def.}}{=} \left\| A u^\dagger - f \right\|$. Then,

$$\left\| A u^\dagger - f_\delta \right\| \leqslant \left\| A u^\dagger - f \right\| + \|f_\delta - f\| \leqslant \mu + \delta.$$

It may be hard to estimate μ in practice, but if $\mathcal{R}(A)$ is dense in $\mathcal{V}$, then $\mu = 0$.

A posteriori parameter choice rules

We may want a parameter choice rule which takes the approximate data f_δ into account. One way of approaching this is via the Morozov's discrepancy principle:

Definition 10

Let $u_\alpha = R_\alpha f_\delta$ with $\alpha(\delta, f_\delta)$ chosen as follows:

$$\alpha(\delta, f_\delta) = \sup \left\{ \alpha > 0 \; ; \; \left\| Au_{\alpha(\delta, f_\delta)} - f_\delta \right\| \leqslant \eta\delta \right\}$$

where $\eta > 1$. Then, $u_{\alpha(\delta, f_\delta)}$ is said to satisfy Morozov's discrepancy principle.

- It can be shown that the a-posteriori parameter choice rule indeed yields a convergent regularisation method.

- Given $f \in \mathcal{D}(A^\dagger)$ and $f_\delta \in \mathcal{V}$ such that $\|f_\delta - f\| \leqslant \delta$, let $u^\dagger$ be the minimal norm solution to data f and define $\mu \stackrel{\text{def}}{=} \left\| Au^\dagger - f \right\|$. Then,

$$\left\| Au^\dagger - f_\delta \right\| \leqslant \left\| Au^\dagger - f \right\| + \|f_\delta - f\| \leqslant \mu + \delta.$$

  It may be hard to estimate μ in practice, but if $\mathcal{R}(A)$ is dense in $\mathcal{V}$, then $\mu = 0$.

- In practice, pick a null sequence $\{\alpha_j\}_j$ and iteratively compute $u_{\alpha_j} \stackrel{\text{def}}{=} R_{\alpha_j} f_\delta$ for $j = 1, \ldots, j^*$, until $u_{\alpha_{j^*}}$ satisfies Morozov's discrepancy principle.

Heuristic parameter choice rules

Estimating δ can be hard in practice. One might try to define a rule based on only f_δ.

Estimating δ can be hard in practice. One might try to define a rule based on only f_δ.

Heuristic rules yield convergent regularisation only for well-posed problems:

Theorem 11 (The Bakushinskii veto)

Let $A \in \mathcal{L}(\mathcal{U}, \mathcal{V})$ and $\{R_\alpha\}_\alpha$ be a regularisation for $A^\dagger$. Let $\alpha = \alpha(f_\delta)$ be such that (R_α, α) is a convergent regularisaton. Then, $A^\dagger$ is continuous from $\mathcal{V} \to \mathcal{U}$.

Estimating δ can be hard in practice. One might try to define a rule based on only f_δ.

Heuristic rules yield convergent regularisation only for well-posed problems:

Theorem 11 (The Bakushinskii veto)

Let $A \in \mathcal{L}(\mathcal{U}, \mathcal{V})$ and $\{R_\alpha\}_\alpha$ be a regularisation for $A^\dagger$. Let $\alpha = \alpha(f_\delta)$ be such that (R_α, α) is a convergent regularisaton. Then, $A^\dagger$ is continuous from $\mathcal{V} \to \mathcal{U}$.

If we have a convergent regularisation, then

$$\lim_{\delta \to 0} \sup \left\{ \left\| R_{\alpha(f_\delta)} f_\delta - A^\dagger f \right\| \; ; \; f \in \mathcal{D}(A^\dagger), \; \|f_\delta - f\| \leqslant \delta \right\} = 0$$

i.e. $R_{\alpha(f)} f = A^\dagger f$ for all $f \in \mathcal{D}(A^\dagger)$.

Estimating δ can be hard in practice. One might try to define a rule based on only f_δ.

Heuristic rules yield convergent regularisation only for well-posed problems:

Theorem 11 (The Bakushinskii veto)

Let $A \in \mathcal{L}(\mathcal{U}, \mathcal{V})$ and $\{R_\alpha\}_\alpha$ be a regularisation for $A^\dagger$. Let $\alpha = \alpha(f_\delta)$ be such that (R_α, α) is a convergent regularisaton. Then, $A^\dagger$ is continuous from $\mathcal{V} \to \mathcal{U}$.

If we have a convergent regularisation, then

$$\lim_{\delta \to 0} \sup \left\{ \left\| R_{\alpha(f_\delta)} f_\delta - A^\dagger f \right\| \; ; \; f \in \mathcal{D}(A^\dagger), \; \|f_\delta - f\| \leqslant \delta \right\} = 0$$

i.e. $R_{\alpha(f)} f = A^\dagger f$ for all $f \in \mathcal{D}(A^\dagger)$.

Taking any sequence $f_j \in \mathcal{D}(A^\dagger)$ which converges to $f \in \mathcal{D}(A^\dagger)$,
$\lim_{j \to \infty} A^\dagger f_j = \lim_{j \to \infty} R_{\alpha(f_j)} f_j = A^\dagger f$. Therefore, $A^\dagger$ is continuous on $\mathcal{D}(A^\dagger)$.
Since $\mathcal{D}(A^\dagger)$ is dense in $\mathcal{V}$, there exists a continuous extension of $A^\dagger$ on $\mathcal{V}$.

Heuristic parameter choice rules

Despite this negative result of Bakushinskii, heuristic rules are still employed in practices because

- This only applies to infinite dimensional operators.
- This is an asymptotic and a worst-case result. For fixed noise levels or restricted noise values, heuristic rules can still give good performance.

Hanke-Raus rule. Choose $\alpha(f^\delta)$ as

$$\alpha(f^\delta) \stackrel{\text{def.}}{=} \text{argmin}_\alpha \frac{1}{\sqrt{\alpha}} \left\| A u_\alpha^\delta - f^\delta \right\|_{\mathcal{V}}.$$

Heuristic parameter choice rules

Despite this negative result of Bakushinskii, heuristic rules are still employed in practices because

- This only applies to infinite dimensional operators.
- This is an asymptotic and a worst-case result. For fixed noise levels or restricted noise values, heuristic rules can still give good performance.

Hanke-Raus rule. Choose $\alpha(f^\delta)$ as

$$\alpha(f^\delta) \overset{\text{def.}}{=} \operatorname{argmin}_\alpha \frac{1}{\sqrt{\alpha}} \left\| A u_\alpha^\delta - f^\delta \right\|_{\mathcal{V}}.$$

What's with the $\alpha^{-\frac{1}{2}}$?

Heuristic parameter choice rules

Despite this negative result of Bakushinskii, heuristic rules are still employed in practices because

- This only applies to infinite dimensional operators.
- This is an asymptotic and a worst-case result. For fixed noise levels or restricted noise values, heuristic rules can still give good performance.

Hanke-Raus rule. Choose $\alpha(f^\delta)$ as

$$\alpha(f^\delta) \stackrel{\text{def.}}{=} \operatorname{argmin}_\alpha \frac{1}{\sqrt{\alpha}} \left\| A u_\alpha^\delta - f^\delta \right\|_{\mathcal{V}}.$$

What's with the $\alpha^{-\frac{1}{2}}$?

Recall that $\left\| A u_\alpha - A u_\alpha^\delta \right\| \leqslant \gamma\delta$ where $\gamma \geqslant \sup_{\alpha,\sigma} \sigma g_\alpha(\sigma)$. We also have

$$\left\| A u^\dagger - A u_\alpha \right\|^2 = \sum_j \left| g_\alpha(\sigma_j)\sigma_j - 1 \right|^2 \sigma_j^2 \left| \langle u^\dagger,\, u_j \rangle \right|^2 \leqslant \alpha^2 \left\| u^\dagger \right\|^2$$

for both truncated SVD and Tikhonov regularisation (check this!).

Heuristic parameter choice rules

Despite this negative result of Bakushinskii, heuristic rules are still employed in practices because

- This only applies to infinite dimensional operators.
- This is an asymptotic and a worst-case result. For fixed noise levels or restricted noise values, heuristic rules can still give good performance.

Hanke-Raus rule. Choose $\alpha(f^\delta)$ as

$$\alpha(f^\delta) \stackrel{\text{def.}}{=} \operatorname{argmin}_\alpha \frac{1}{\sqrt{\alpha}} \left\| A u_\alpha^\delta - f^\delta \right\|_{\mathcal{V}}.$$

What's with the $\alpha^{-\frac{1}{2}}$?

Recall that $\left\| A u_\alpha - A u_\alpha^\delta \right\| \leqslant \gamma \delta$ where $\gamma \geqslant \sup_{\alpha,\sigma} \sigma g_\alpha(\sigma)$. We also have

$$\left\| A u^\dagger - A u_\alpha \right\|^2 = \sum_j \left| g_\alpha(\sigma_j)\sigma_j - 1 \right|^2 \sigma_j^2 \left| \langle u^\dagger, \, u_j \rangle \right|^2 \leqslant \alpha^2 \left\| u^\dagger \right\|^2$$

for both truncated SVD and Tikhonov regularisation (check this!).

So, $\left\| A u^\dagger - A u_\alpha^\delta \right\| \lesssim \alpha + \delta$.

The optimal choice is $\alpha \sim \delta$. Note that $\delta = \operatorname{argmin}_\alpha \frac{(\alpha+\delta)}{\sqrt{\alpha}}$.

The L-Curve

To motivate the L-Curve, recall our error bounds with $\gamma = \sup_{\alpha, \sigma} g_\alpha(\sigma)$.

- $\left\| A u_\alpha^\delta - A u^\dagger \right\| \leqslant \sum_j \left| g_\alpha(\sigma_j) \sigma_j - 1 \right|^2 \sigma_j^2 \left| \langle u^\dagger, u_j \rangle \right|^2 + \gamma \delta$

- $\left\| u_\alpha^\delta - u^\dagger \right\| \leqslant \sum_j \left| g_\alpha(\sigma_j) \sigma_j - 1 \right|^2 \left| \langle u^\dagger, u_j \rangle \right|^2 + C_\alpha \delta.$

To motivate the L-Curve, recall our error bounds with $\gamma = \sup_{\alpha, \sigma} g_\alpha(\sigma)$.

- $\left\| Au_\alpha^\delta - Au^\dagger \right\| \leqslant \sum_j \left| g_\alpha(\sigma_j)\sigma_j - 1 \right|^2 \sigma_j^2 \left| \langle u^\dagger, u_j \rangle \right|^2 + \gamma\delta$

- $\left\| u_\alpha^\delta - u^\dagger \right\| \leqslant \sum_j \left| g_\alpha(\sigma_j)\sigma_j - 1 \right|^2 \left| \langle u^\dagger, u_j \rangle \right|^2 + C_\alpha\delta.$

As we decrease α, the data error increases, so $\left\| u_\alpha^\delta \right\|$ grows, but $\left\| Au_\alpha^\delta - Au^\dagger \right\|$ remains roughly constant.

The L-Curve

To motivate the L-Curve, recall our error bounds with $\gamma = \sup_{\alpha,\sigma} g_\alpha(\sigma)$.

- $\left\| Au_\alpha^\delta - Au^\dagger \right\| \leqslant \sum_j \left| g_\alpha(\sigma_j)\sigma_j - 1 \right|^2 \sigma_j^2 \left| \langle u^\dagger, u_j \rangle \right|^2 + \gamma\delta$

- $\left\| u_\alpha^\delta - u^\dagger \right\| \leqslant \sum_j \left| g_\alpha(\sigma_j)\sigma_j - 1 \right|^2 \left| \langle u^\dagger, u_j \rangle \right|^2 + C_\alpha\delta.$

As we decrease α, the data error increases, so $\left\| u_\alpha^\delta \right\|$ grows, but $\left\| Au_\alpha^\delta - Au^\dagger \right\|$ remains roughly constant.

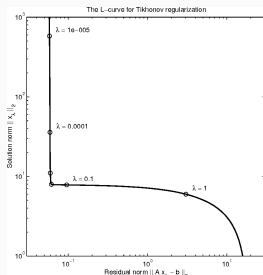As we increase α, the approximation error grows, but $\|u_\alpha\|$ remains roughly constant.

The L-Curve

To motivate the L-Curve, recall our error bounds with $\gamma = \sup_{\alpha,\sigma} g_\alpha(\sigma)$.

- $\left\| Au_\alpha^\delta - Au^\dagger \right\| \leqslant \sum_j \left| g_\alpha(\sigma_j)\sigma_j - 1 \right|^2 \sigma_j^2 \left| \langle u^\dagger,\, u_j \rangle \right|^2 + \gamma\delta$

- $\left\| u_\alpha^\delta - u^\dagger \right\| \leqslant \sum_j \left| g_\alpha(\sigma_j)\sigma_j - 1 \right|^2 \left| \langle u^\dagger,\, u_j \rangle \right|^2 + C_\alpha\delta.$

As we decrease α, the data error increases, so $\left\| u_\alpha^\delta \right\|$ grows, but $\left\| Au_\alpha^\delta - Au^\dagger \right\|$ remains roughly constant.

As we increase α, the approximation error grows, but $\|u_\alpha\|$ remains roughly constant.

Plotting $\log \|Au_\alpha - f_\delta\|$ against $\log(\|u_\alpha\|)$ for varying α gives an L-curve.

L-Curve. Choose $\alpha(f^\delta) = \mathrm{argmin}_{\alpha>0} \|u_\alpha\|_{\mathcal{U}} \left\| Au_\alpha - f^\delta \right\|_{\mathcal{V}}.$

Idea: withhold parts of f, and choose α such that we can predict this withheld data.

Idea: withhold parts of f, and choose α such that we can predict this withheld data.

Cross Validation: Let $A \in \mathbb{R}^{m \times m}$ and let $A^{(i)}$ be the matrix A with ith row removed, and $f^{(i)}$ be the vector with the ith entry removed. Let $u_\alpha^{(i)}$ be the regularised solution to $A^{(i)} u = f^{(i)}$.

Generalized Cross Validation

Idea: withhold parts of f, and choose α such that we can predict this withheld data.

Cross Validation: Let $A \in \mathbb{R}^{m \times m}$ and let $A^{(i)}$ be the matrix A with ith row removed, and $f^{(i)}$ be the vector with the ith entry removed. Let $u_\alpha^{(i)}$ be the regularised solution to $A^{(i)} u = f^{(i)}$.

$$\alpha_* = \operatorname{argmin}_\alpha \frac{1}{m} \sum_{i=1}^{m} \left(A(i,:) u_\alpha^{(i)} - f_i \right)^2$$

Generalized Cross Validation

Idea: withhold parts of f, and choose α such that we can predict this withheld data.

Cross Validation: Let $A \in \mathbb{R}^{m \times m}$ and let $A^{(i)}$ be the matrix A with ith row removed, and $f^{(i)}$ be the vector with the ith entry removed. Let $u_\alpha^{(i)}$ be the regularised solution to $A^{(i)} u = f^{(i)}$.

$$\alpha_* = \text{argmin}_\alpha \frac{1}{m} \sum_{i=1}^m \left(A(i,:) u_\alpha^{(i)} - f_i \right)^2$$

For Tikhonov regularisation, this is the same as

$$\alpha_* = \text{argmin}_\alpha \frac{1}{m} \sum_{i=1}^m \left(\frac{A(i,:) u_\alpha - f_i}{1 - h_{ii}} \right)^2$$

where h_{ii} are diagonal elements of $A(A^\top A + \alpha \text{Id})^{-1} A^\top$.

Generalized Cross Validation

Idea: withhold parts of f, and choose α such that we can predict this withheld data.

Cross Validation: Let $A \in \mathbb{R}^{m \times m}$ and let $A^{(i)}$ be the matrix A with ith row removed, and $f^{(i)}$ be the vector with the ith entry removed. Let $u_\alpha^{(i)}$ be the regularised solution to $A^{(i)} u = f^{(i)}$.

$$\alpha_* = \text{argmin}_\alpha \frac{1}{m} \sum_{i=1}^{m} \left(A(i,:) u_\alpha^{(i)} - f_i \right)^2$$

For Tikhonov regularisation, this is the same as

$$\alpha_* = \text{argmin}_\alpha \frac{1}{m} \sum_{i=1}^{m} \left(\frac{A(i,:) u_\alpha - f_i}{1 - h_{ii}} \right)^2$$

where h_{ii} are diagonal elements of $A(A^\top A + \alpha \text{Id})^{-1} A^\top$.

Generalised Cross Validation: h_{ii} depends on the way you order the rows of A. To remove this dependence, take the average of h_{ii}'s:

$\text{tr}\left(A(A^\top A + \alpha \text{Id})^{-1} A^\top \right) = \sum_{i=1}^{n} \lambda_i$ where $\lambda_i = \frac{\sigma_i^2}{\sigma_i^2 + \alpha}$.

$$\alpha_* = \text{argmin}_\alpha \frac{\|A u_\alpha - f\|^2}{\left(m - \sum_{i=1}^{n} \lambda_i \right)^2}$$

Landweber iteration

Let us consider computing a least squares solution via gradient descent on

$$F(u) = \frac{1}{2} \|Au - f\|_{\mathcal{V}}^2.$$

We have $\nabla F(u) = A^*(Au - f)$, and gradient descent is know as

The Landweber iterations

$$u^{k+1} = (\mathrm{Id} - \tau A^* A)u^k + \tau A^* f$$
$$u^0 = 0$$

Let us consider computing a least squares solution via gradient descent on

$$F(u) = \frac{1}{2} \|Au - f\|_{\mathcal{V}}^2 \,.$$

We have $\nabla F(u) = A^*(Au - f)$, and gradient descent is know as

The Landweber iterations

$$u^{k+1} = (\mathrm{Id} - \tau A^* A)u^k + \tau A^* f$$
$$u^0 = 0$$

We don't want to compute $A^\dagger f$ if $f \notin \mathcal{D}(A^\dagger)$, we will see that k corresponds to a regularisation parameter and stopping early is a form of regularisation!

Lemma 4.1

Let $\tau \in (0, \frac{2}{\|A\|^2})$. Then, $\left\| Au^{k+1} - f \right\|_{\mathcal{V}} \leqslant \left\| Au^k - f \right\|_{\mathcal{V}}$, with equality only if $A^*(Au^k - f) = 0$.

Lemma 4.1

Let $\tau \in (0, \frac{2}{\|A\|^2})$. Then, $\left\|Au^{k+1} - f\right\|_{\mathcal{V}} \leqslant \left\|Au^k - f\right\|_{\mathcal{V}}$, with equality only if $A^*(Au^k - f) = 0$.

$$
\begin{aligned}
\left\|Au^{k+1} - f\right\|^2 &= \left\|A(\mathrm{Id} - \tau A^* A)u^k + \tau AA^* f - f\right\|^2 \\
&= \left\|Au^k - f - \tau AA^* Au^k + \tau AA^* f\right\|^2 \\
&= \left\|Au^k - f\right\|^2 + \tau^2 \left\|AA^*(Au^k - f)\right\|^2 - 2\tau\langle A^*(Au^k - f), A^*(Au^k - f)\rangle \\
&\leqslant \left\|Au^k - f\right\|^2 + (\tau^2 \|A\|^2 - 2\tau) \left\|A^*(Au^k - f)\right\|^2
\end{aligned}
$$

with equality only if $A^*(Au^k - f) = 0$ since $(\tau^2 \|A\|^2 - 2\tau)$ is negative by our choice of τ.

Landweber iteration is a form of spectral regularisation

By induction we have

$$u^k = \tau \sum_{\ell=0}^{k-1} (\mathrm{Id} - \tau A^* A)^\ell A^* f \stackrel{\text{def.}}{=} R_k f.$$

Since $A^* f = \sum_{j=1}^{\infty} \sigma_j \langle f, v_j \rangle u_j$, we have

$$
\begin{aligned}
R_k f &= \tau \sum_{j=1}^{\infty} \sum_{\ell=0}^{k-1} \sigma_j \langle f, v_j \rangle (\mathrm{Id} - \tau A^* A)^\ell u_j \\
&= \tau \sum_{j=1}^{\infty} \sigma_j \langle f, v_j \rangle \sum_{\ell=0}^{k-1} (1 - \tau \sigma_j^2)^\ell u_j = \sum_{j=1}^{\infty} \frac{1 - (1 - \tau \sigma_j^2)^k}{\sigma_j} \langle f, v_j \rangle u_j.
\end{aligned}
$$

Let $\tau \in (0, \frac{2}{\|A\|^2})$ and $g_k(\sigma) \stackrel{\text{def.}}{=} \frac{1 - (1 - \tau \sigma^2)^k}{\sigma}$. Then

$$R_k f = \sum_{j=1}^{\infty} g_k(\sigma) \langle f, v_j \rangle u_j.$$

Regularisation and convergence

Regularisation: Note that $\left|1 - \tau\sigma_j\right| < 1$ for $\tau \in (0, 2/\left\|A\right\|^2)$, so

$$g_k(\sigma) \stackrel{\text{def.}}{=} \frac{1 - (1 - \tau\sigma^2)^k}{\sigma} \to \frac{1}{\sigma}, \quad \text{as } k \to \infty.$$

Since $\sigma g_k(\sigma) \leqslant 2$ uniformly, we have a regularisation.

Regularisation and convergence

Regularisation: Note that $\left| 1 - \tau \sigma_j \right| < 1$ for $\tau \in (0, 2/\|A\|^2)$, so

$$g_k(\sigma) \stackrel{\text{def.}}{=} \frac{1 - (1 - \tau \sigma^2)^k}{\sigma} \to \frac{1}{\sigma}, \quad \text{as } k \to \infty.$$

Since $\sigma g_k(\sigma) \leqslant 2$ uniformly, we have a regularisation.

Note that R_k is a linear regulariser and

$$AR_k f = \sum_{j=1}^{\infty} \left(1 - (1 - \tau \sigma_j^2)^k \right) \langle f, v_j \rangle v_j \implies \|AR_k\|_{\mathcal{L}(\mathcal{V}, \mathcal{V})} \leqslant 2,$$

so we have $\|R_k f\| \to +\infty$ for all $f \notin \mathcal{D}(A^\dagger)$.

Regularisation and convergence

Regularisation: Note that $\left|1 - \tau\sigma_j\right| < 1$ for $\tau \in (0, 2/\left\|A\right\|^2)$, so

$$g_k(\sigma) \overset{\text{def.}}{=} \frac{1 - (1 - \tau\sigma^2)^k}{\sigma} \to \frac{1}{\sigma}, \quad \text{as } k \to \infty.$$

Since $\sigma g_k(\sigma) \leqslant 2$ uniformly, we have a regularisation.

Note that R_k is a linear regulariser and

$$AR_k f = \sum_{j=1}^{\infty} \left(1 - (1 - \tau\sigma_j^2)^k\right) \langle f, v_j\rangle v_j \implies \left\|AR_k\right\|_{\mathcal{L}(\mathcal{V},\mathcal{V})} \leqslant 2,$$

so we have $\left\|R_k f\right\| \to +\infty$ for all $f \notin \mathcal{D}(A^\dagger)$.

How to choose k to ensure convergence?

Regularisation and convergence

Regularisation: Note that $\left|1 - \tau\sigma_j\right| < 1$ for $\tau \in (0, 2/\|A\|^2)$, so

$$g_k(\sigma) \overset{\text{def.}}{=} \frac{1 - (1 - \tau\sigma^2)^k}{\sigma} \to \frac{1}{\sigma}, \quad \text{as } k \to \infty.$$

Since $\sigma g_k(\sigma) \leqslant 2$ uniformly, we have a regularisation.

Note that R_k is a linear regulariser and

$$AR_k f = \sum_{j=1}^{\infty} \left(1 - (1 - \tau\sigma_j^2)^k\right) \langle f, v_j \rangle v_j \implies \|AR_k\|_{\mathcal{L}(\mathcal{V},\mathcal{V})} \leqslant 2,$$

so we have $\|R_k f\| \to +\infty$ for all $f \notin \mathcal{D}(A^\dagger)$.

How to choose k to ensure convergence?
Recall that we need $g_k(\sigma) \leqslant C_k$ and $\lim_{\delta \to 0} \delta C_k = 0$. From applying $e^{-x} \geqslant 1 - x$ twice, we have

$$g_k(\sigma) \leqslant \frac{1 - e^{-\tau\sigma^2 k}}{\sigma} \leqslant \frac{\tau\sigma^2 k}{\sigma} = \tau k\sigma \leqslant \|A\| \, k\tau,$$

we need a stopping criteria of $k_*(\delta)$ such that $\lim_{\delta \to 0} k_*(\delta)\delta = 0$.

To summarise:

Lemma 4.2

Let $\tau \in (0, 2/\|A\|^2)$.

(i) Let $f \in \mathcal{D}(A^\dagger)$ and $u^\dagger = A^\dagger f$, then $\|u^k - u^\dagger\| \to 0$.

(ii) If $f \notin \mathcal{D}(A^\dagger)$, then $\|u^k\| \to \infty$.

(iii) If $\lim_{\delta \to 0} k_*(\delta)\delta = 0$, then $\|u^{k_*(\delta)} - u^\dagger\| \to 0$ as $\delta \to 0$, when doing Landweber iteration on f_δ such that $\|f_\delta - f\| \leqslant \delta$ and $f \in \mathcal{D}(A^\dagger)$.

Interpret $\alpha \stackrel{\text{def.}}{=} 1/k$, then

$$u_\alpha = R_\alpha f = \sum_{j=1}^{\infty} \left(1 - (1 - \tau \sigma_j^2)^{1/\alpha}\right) \frac{1}{\sigma_j} \langle f, v_j \rangle u_j$$

Theorem 12

Let $\tau \in (0, 2/\|A\|^2)$. Assume that there exists $w \in \mathcal{V}$ such that $u^\dagger \stackrel{\text{def.}}{=} A^\dagger f = A^* w$. Then,

(i) letting $f = Au^\dagger$,

$$\left\|u^k - u^\dagger\right\|_{\mathcal{U}} = \mathcal{O}\left(\frac{1}{\sqrt{k}}\right) = \mathcal{O}(\sqrt{\alpha})$$

(ii) letting $f = Au^\dagger$ and $f^\delta \in \mathcal{V}$ such that $\left\|f^\delta - f\right\| \leqslant \delta$,

$$\left\|u_\delta^k - u^\dagger\right\|_{\mathcal{U}} \leqslant \sqrt{\tau k}\delta + \frac{\|w\|}{\sqrt{\tau(2k-1)}}$$

Note the trade-off between approximation and data error. We need to stop early!

Let $f = Au^{\dagger}$. Recall that

$$u^{\dagger} - u^k = u^{\dagger} - \tau \sum_{j=0}^{k-1} (\mathrm{Id} - \tau A^*A)^j A^* A u^{\dagger}.$$

One can check that $\sum_{j=0}^{k-1} (\mathrm{Id} - \tau A^*A)^j A^* A = A^* A \sum_{j=0}^{k-1} (\mathrm{Id} - \tau A^*A)^j$ and show by induction that

$$\mathrm{Id} - (\mathrm{Id} - \tau A^*A)^k = \tau A^* A \sum_{j=0}^{k-1} (\mathrm{Id} - \tau A^*A)^j$$

Therefore, $u^{\dagger} - u^k = (\mathrm{Id} - \tau A^*A)^k u^{\dagger}$.

Using the source condition that $u^{\dagger} = A^* w$, we have

$$\langle u^{\dagger} - u^k, u_j \rangle = \langle A^* w, (1 - \tau \sigma_j^2)^k u_j \rangle = (1 - \tau \sigma_j^2)^k \sigma_j \langle w, v_j \rangle.$$

Therefore,

$$\left\| u^{\dagger} - u^k \right\| \leqslant \|w\| \max_{\sigma}(1 - \tau \sigma^2)^k \sigma \leqslant \|w\| \left((1 - \tau \sigma^2)^k \sigma \right) \leqslant \frac{\|w\|}{\sqrt{(2k+1)\tau}}$$

where the maximum is achieved at $\tau \sigma^2 = 1/(2k+1)$.

For (ii), recall that $u_k = R_k(f) \overset{\text{def.}}{=} \tau \sum_{j=0}^{k-1}(\text{Id} - \tau A^*A)^j A^* f$ and
$\text{Id} - (\text{Id} - \tau A^*A)^k = \tau A^*A \sum_{\ell=0}^{k-1}(\text{Id} - \tau A^*A)^\ell$. So, $u_k - u_k^\delta = R_k(f - f_\delta)$. Now,

$$\|R_k\|^2 = \|R_k R_k^*\| = \tau^2 \left\| \sum_{j=0}^{k-1}(\text{Id} - \tau A^*A)^j A^* A \sum_{\ell=0}^{k-1}(\text{Id} - \tau A^*A)^\ell \right\|$$

$$= \tau \left\| \sum_{j=0}^{k-1}(\text{Id} - \tau A^*A)^j \left(\text{Id} - (\text{Id} - \tau A^*A)^k\right) \right\|$$

$$\leqslant \tau \left\| \sum_{j=0}^{k-1}(\text{Id} - \tau A^*A)^j \right\| \leqslant k\tau.$$

Therefore,

$$\left\| u_k - u_k^\delta \right\| \leqslant \sqrt{k\tau} \, \|f - f_\delta\| \leqslant \sqrt{k\tau}\,\delta.$$

Mozorov's discrepancy principle

Given noisy data $\|f - f_\delta\| \leqslant \delta$, consider Mozorov's discrepancy principle as a stopping criteria: Stop when

$$\left\| A u_\delta^k - f_\delta \right\| \leqslant \eta \delta, \quad \text{where} \quad \eta > 1.$$

Lemma 4.3

Let $\tau \in (0, \frac{2}{\|A\|})$. Then, for all $k \leqslant k^*$ and $f = Au^\dagger$ and $\|f^\delta - f\| \leqslant \delta$, we have

$$\left\| u_\delta^{k+1} - u^\dagger \right\| \leqslant \left\| u_\delta^k - u^\dagger \right\|_{\mathcal{U}}$$

where k^* is chosen in accordance to the discrepancy principle with
$\eta = \frac{2}{2 - \tau \|K\|^2} > 1$. Equality is attained only for $\delta = 0$ and $A^*(Au_\delta^k - f_\delta) = 0$.

i.e. We move closer to $u^\dagger$ as long as the discrepancy principle is violated. One can in fact show that $\left\| u_\delta^{k^*} - u^\dagger \right\| = \mathcal{O}(\delta^{1/2})$ under the source condition of $u^\dagger = A^* w$.

Assume that $\left\|Au_\delta^k - f_\delta\right\| > \eta\delta$.

Assume that $\left\| A u_\delta^k - f_\delta \right\| > \eta \delta$.

Plug in the definition $u_\delta^{k+1} = u_\delta^k - \tau A^* A u_\delta^k + \tau A^* f_\delta$ and rearrange:

$$\left\| u_\delta^{k+1} - u^\dagger \right\|^2 - \left\| u_\delta^k - u^\dagger \right\|^2 = \left\| u_\delta^k - \tau A^* A u_\delta^k + \tau A^* f_\delta - u^\dagger \right\|^2 - \left\| u_\delta^k - u^\dagger \right\|^2$$

$$= \left\| -\tau A^* A u_\delta^k + \tau A^* f_\delta \right\|^2 + 2\langle -\tau A^* A u_\delta^k + \tau A^* f_\delta,\, u_\delta^k - u^\dagger \rangle$$

$$= \left\| -\tau A^* A u_\delta^k + \tau A^* f_\delta \right\|^2 + 2\tau \langle -A u_\delta^k + f_\delta,\, A u_\delta^k - A u^\dagger \rangle$$

Assume that $\left\|Au_\delta^k - f_\delta\right\| > \eta\delta$.

Plug in the definition $u_\delta^{k+1} = u_\delta^k - \tau A^* Au_\delta^k + \tau A^* f_\delta$ and rearrange:

$$\left\|u_\delta^{k+1} - u^\dagger\right\|^2 - \left\|u_\delta^k - u^\dagger\right\|^2 = \left\|u_\delta^k - \tau A^* Au_\delta^k + \tau A^* f_\delta - u^\dagger\right\|^2 - \left\|u_\delta^k - u^\dagger\right\|^2$$

$$= \left\|-\tau A^* Au_\delta^k + \tau A^* f_\delta\right\|^2 + 2\langle -\tau A^* Au_\delta^k + \tau A^* f_\delta, \, u_\delta^k - u^\dagger\rangle$$

$$= \left\|-\tau A^* Au_\delta^k + \tau A^* f_\delta\right\|^2 + 2\tau\langle -Au_\delta^k + f_\delta, \, Au_\delta^k - Au^\dagger\rangle$$

Note that

- $\left\|-\tau A^* Au_\delta^k + \tau A^* f_\delta\right\|^2 \leqslant \tau^2 \left\|A\right\|^2 \left\|Au_\delta^k - f_\delta\right\|^2$

- $2\tau\langle -Au_\delta^k + f_\delta, \, Au_\delta^k - f\rangle = -2\tau\left\|Au_\delta^k + f_\delta\right\|^2 + 2\tau\langle -Au_\delta^k + f_\delta, \, f_\delta - f\rangle$

Assume that $\left\| Au_\delta^k - f_\delta \right\| > \eta\delta$.

Plug in the definition $u_\delta^{k+1} = u_\delta^k - \tau A^* A u_\delta^k + \tau A^* f_\delta$ and rearrange:

$$\left\| u_\delta^{k+1} - u^\dagger \right\|^2 - \left\| u_\delta^k - u^\dagger \right\|^2 = \left\| u_\delta^k - \tau A^* A u_\delta^k + \tau A^* f_\delta - u^\dagger \right\|^2 - \left\| u_\delta^k - u^\dagger \right\|^2$$

$$= \left\| -\tau A^* A u_\delta^k + \tau A^* f_\delta \right\|^2 + 2\langle -\tau A^* A u_\delta^k + \tau A^* f_\delta,\ u_\delta^k - u^\dagger \rangle$$

$$= \left\| -\tau A^* A u_\delta^k + \tau A^* f_\delta \right\|^2 + 2\tau\langle -A u_\delta^k + f_\delta,\ A u_\delta^k - A u^\dagger \rangle$$

Note that

- $\left\| -\tau A^* A u_\delta^k + \tau A^* f_\delta \right\|^2 \leqslant \tau^2 \left\| A \right\|^2 \left\| A u_\delta^k - f_\delta \right\|^2$
- $2\tau\langle -A u_\delta^k + f_\delta,\ A u_\delta^k - f \rangle = -2\tau \left\| A u_\delta^k + f_\delta \right\|^2 + 2\tau\langle -A u_\delta^k + f_\delta,\ f_\delta - f \rangle$

Therefore, unless $A^*(A u_\delta^k - A^* f_\delta) = 0$,

$$\left\| u_\delta^{k+1} - u^\dagger \right\|^2 - \left\| u_\delta^k - u^\dagger \right\|^2$$

$$\leqslant \tau^2 \left\| A \right\|^2 \left\| A u_\delta^k - f_\delta \right\|^2 - 2\tau \left\| A u_\delta^k + f_\delta \right\|^2 + 2\tau\delta \left\| -A u_\delta^k + f_\delta \right\|$$

$$= \tau \left\| A u_\delta^k - f_\delta \right\| \left((\tau \left\| A \right\|^2 - 2) \left\| A u_\delta^k - f_\delta \right\| + 2\delta \right)$$

$$< \tau \left\| A u_\delta^k - f_\delta \right\| \left((\tau \left\| A \right\|^2 - 2)\eta\delta + 2\delta \right) < 0$$

Summary

- We defined the notion of convergent regularisations: there is a trade-off between data error and approximation error, so parameters need to be chosen carefully.
- We looked at various forms of spectral (linear) regularisation
- Tikhonov and Landweber iteration are special forms of spectral regularisation which do not require explicit knowledge of the spectrum.
- Convergence rates were obtained under source conditions.