Regularization Operators

Aditya Ganeshan

Indian Institute of Technology, Roorkee

December 13, 2016

Table of Content

Regularization

It is the approximation of a well-posed problem by neighboring well-posed problems.
We want to find the best-approximate solution $x^\dagger = T^\dagger y$, but only y^δ is known, with,

$$\|y^\delta - y\| \leqslant \delta$$

Regularization

In Ill-posed problems, $T^{\dagger}y^{\delta}$ is unbounded(It might not even exist!).
Hence, we look for am approximation x_{α}^{δ}, which

- depends continuously on the noisy data y^{δ}.
- tends to $x^{\dagger}$ as noise level decreases to zero (if regularization parameter α is selected appropriately).

Regularization

- As we look not for specific values of y, rather for every $y \in R(T^\dagger)$, we regularize the solution operator $T^\dagger$.
- A simple regularization of $T^\dagger$ is replacement of unbounded operator $T^\dagger$ by a parameter-dependant family $\{R_\alpha\}$, taking $x_\alpha^\delta = R_\alpha y^\delta$.
- This way we define the regularization operator for the whole collection of equations.

$$Tx = y \quad y \in D(T^\dagger)$$

Definition (3.1)

Let $T : X \to Y$ bea a bounded linear operator between Hilbert spaces X and Y, $\alpha_0 \in (0, +\infty)$. for every $\alpha \in (0, \alpha_0)$, let

$$R_\alpha : Y \to X$$

be a continuous(not necessarily linear) operator.The family $\{R_\alpha\}$ is called a regularization or a regularization operator for $T^\dagger$, if , for all $y \in D(T^\dagger)$, there exists a parameter choice rule $\alpha = \alpha(y^\delta, \delta)$ such that

$$\lim_{\delta \to 0} sup\{ \left\| R_{\alpha(y^\delta, \delta)} y^\delta - T^\dagger y \right\| \mid y^\delta \in Y, \|y^\delta - y\| \leqslant \delta\} = 0 \qquad (1)$$

holds.

Definition (continued)

Here,

$$\alpha : R^{+} \times Y \to (0, \alpha_0) \tag{2}$$

is such that

$$\lim_{\delta \to 0} sup\{\alpha(y^{\delta}, \delta) \mid y^{\delta} \in Y, \|y^{\delta} - y\| \leqslant \delta\} = 0 \tag{3}$$

For a specific $y \in D(T^{\dagger})$, a pair (R_{α}, α) is called a convergent regularization method if 1 and 3 holds.

Remark

- We can extend Definition 1 to include perturbations in the operator. For this we assume that only approximation T_η of T is known with

$$\|T - T_\eta\| \leqslant \eta$$

  Then, we model the parameter rule to depend upon δ, η, y^δ, and T_η.
- We do not require the regularization operators $\{R_\alpha\}$ to be a family of linear operators.

Definition (3.2)

Let α be a parameter choice rule according to definition 3.1. If α does not depend on y^δ, but only on δ, then we call α an a-priori parameter choice rule and write $\alpha = \alpha(\delta)$.
Otherwise, we call it a a-posteriori parameter choice rule.

(If $\alpha = \alpha(y^\delta)$, α is called an error-free parameter choice rule)

Theorem (3.3)

Let $T : X \to Y$ be a bounded linear operator and assume that there is a regularization$\{R_\alpha\}$ for $T^\dagger$ with a error free paramter choice rule, such that the regularization method is convergent for every $y \in D(T^\dagger)$. Then $T^\dagger$ is bounded.

Remark

- The Theorem does not say that error-free parameter choice rule cannot behave well for finite noise levels δ.

Proposition (3.4)

Let for all $\alpha > 0$, R_α be a continuous operator. Then, the family $\{R_\alpha\}$ is a regularization for $T^\dagger$ if

$$R_\alpha \to T^\dagger \quad \text{pointwise} \quad \text{on} \quad D(T^\dagger) \quad \text{as} \quad \alpha \to 0. \tag{4}$$

In this case, there exists, for every $y \in D(T^\dagger)$, an a-priori rule α such that (R_α, α) is a convergent regularization method for solving $Tx = y$.

Proposition (3.5)

Let $\{R_\alpha\}$ be a regularization, α be a linear be as defined in definition 3.1 for all $y \in Y$. Then

$$\{x_\alpha\} \text{ converges to } T^\dagger y \text{ as } \alpha \to 0 \text{ for } y \in D(T^\dagger) \tag{5}$$

and if,

$$\sup\{\|TR_\alpha\| \mid \alpha > 0\} < \infty \tag{6}$$

then

$$\|x_\alpha\| \to +\infty \text{ as } \alpha \to 0 \text{ for } y \notin D(T^\dagger) \tag{7}$$

Proposition (3.6)

Let $\{R_\alpha\}$ be a linear regularization; for every $y \in D(T^\dagger)$, let $\alpha : R^+ \to R^+$ be an a-priori choice rule. Then (R_α, α) is a convergent regularization method if and only if,

$$\lim_{\delta \to 0} \alpha(\delta) = 0 \tag{8}$$

and

$$\lim_{\delta \to 0} \delta \left\| R_{\alpha(\delta)} \right\| = 0 \tag{9}$$

hold.

Table of Content

Order optimality

The rate at which

$$\left\| x_\alpha - x^\dagger \right\| \to 0 \quad as \quad \alpha \to 0. \tag{10}$$

or

$$\left\| x_{\alpha(\delta, y^\delta)} - x^\dagger \right\| \to 0 \quad as \quad \delta \to 0. \tag{11}$$

Definition (3.7)

For $M \subseteq X$, $\delta > 0$, let

$$\Omega(\delta, M) = \sup\left\{\|x\| \mid x \in M, \|Tx\| \leqslant \delta\right\} \tag{12}$$

Ω is known as the modulus of continuity.

Remark

- In general $\Omega(\delta, M)$ will be infinite
- Let $M \bigcap N(T) = \{0\}$, then $\Omega(\delta, M)$ is finite if and only if $T^\dagger$ is continuous on TM.

Definition (3.8)

The worst-case error under the information that $\|y^\delta - y\| \leq \delta$ and a-priori information that $x^\dagger \in M$ is given by

$$\triangle(\delta, M, R) = sup\{\|Ry^\delta - x\| \mid x \in M, y^\delta \in Y, \|Tx - y^\delta\| \leq \delta\} \tag{13}$$

Remark

An "optimal method" R_0 in a class of methods R would be one for which

$$\triangle(\delta, M, R_0) = inf\{\triangle(\delta, M, R) \mid R \in R\} \tag{14}$$

Proposition (3.9)

Let $M \subseteq X$, $\delta > 0$, $R : Y \to X$ *be an arbitrary map with* $R(0) = 0$. *Then*

$$\triangle(\delta, M, R) \geqslant \Omega(\delta, M) \tag{15}$$

Proposition (3.10)

Let $R(T)$ be a non-closed, $\{R_\alpha\}$ be an regularization operator for $T^\dagger$, with $R_\alpha(0) = 0$, $\alpha = \alpha(\delta, y^\delta)$ be a parameter choice rule.Then there can be no function $f : R^+ \to R^+$ with $\lim_{\delta\to 0} f(\delta) = 0$ such that,

$$\left\| R_{\alpha(\delta, y^\delta)} y^\delta - T^\dagger y \right\| \leqslant f(\delta) \tag{16}$$

holds for all $y \in D(T^\dagger)$ with $\|y\| \leqslant 1$ and all $\delta > 0$.

Source Sets

Convergence rates can only be on subsets of $D(T^{\dagger})$. i.e. under a-priori assumptions on the exact data.Hence, we consider subsets of the form

$$\{x \in X \mid x = Bw, \|w\| \leqslant \rho\}$$

where B is a linear operator from some Hilbert space into X. For the choice of B,

$$B = (T^*T)^{\mu}$$

for some $\mu > 0$, we denote the set formed by

$$X_{\mu,\rho} := \{x \in X \mid x = (T^*T)^{\mu}w, \|w\| < \rho\} \tag{17}$$

Source sets

We use further the notation,

$$X_\mu := \bigcup_{\rho>0} X_{\mu,\rho} = R\left((T^\perp T)^\mu\right) \tag{18}$$

These are usually called Source sets, $x \in X_{\mu,\rho}$ is said to have a source representation.
This requirement can be considered as a smoothness condition.

Proposition (3.11)

Let K be compact with singular system $(\sigma_n; v_n, u_n)$*. Then for* $\mu > 0$

$$K^\dagger y \in R((K^*K)^\mu) \tag{19}$$

if and only if

$$\sum_{n=1}^{\infty} \frac{|\langle y, u_n \rangle|^2}{\sigma_n^{2+4\mu}} < \infty \tag{20}$$

Compared to Picard's Criterion, this can be seen as a condition on the decay rate of $\{\langle y, u_n \rangle\}$.

Proposition (3.11)

Let K be compact with singular system $(\sigma_n; v_n, u_n)$. Then for $\mu > 0$

$$K^{\dagger}y \in R((K^*K)^{\mu}) \tag{19}$$

if and only if

$$\sum_{n=1}^{\infty} \frac{|\langle y, u_n\rangle|^2}{\sigma_n^{2+4\mu}} < \infty \tag{20}$$

Compared to Picard's Criterion, this can be seen as a condition on the decay rate of $\{\langle y, u_n\rangle\}$.

Proposition (3.12)

For any $\mu, \rho > 0$,*let* $X_{\mu,\rho}$ *be as earlier defined.Then for any* $\delta > 0$

$$\Omega(\delta, X_{\mu,\rho}) \leqslant \delta^{\frac{2\mu}{2\mu+1}} \rho^{\frac{1}{2\mu+1}} \tag{21}$$

holds.

Proposition

Let K be compact with non-closed range. Then, for any $\mu, \rho > 0$, *there is a sequence* $\{\delta_k\}$ *converging to 0 such that*

$$\Omega(\delta_k, X_{\mu,\rho}) = \delta_k^{\frac{2\mu}{2\mu+1}} \rho^{\frac{1}{2\mu+1}} \tag{22}$$

This tells us that there is a sequence $\{\delta_k\}$ converging to 0 such that $\Omega(\delta_k, X_{\mu,\rho})$ does not go faster to 0 than $\delta_k^{\frac{2\mu}{2\mu+1}}$

Definition (3.13)

Let $R(T)$ be non-closed, $\{R_\alpha\}$ be a regularization operator for $T_\dagger$. For $\mu, \rho > 0$ and $y \in TX_{\mu,\rho}$, let α be a parameter choice rule. We call (R_α, α) optimal in $X_{\mu,\rho}$ if

$$\triangle(\delta, X_{\mu,\rho}, R_\alpha) = \delta^{\frac{2\mu}{2\mu+1}} \rho^{\frac{1}{2\mu+1}} \tag{23}$$

holds for all $\delta > 0$. We call (R_α, α) of optimal order in $X_{\mu,\rho}$ if there exist a constant $c \geqslant 1$ such that

$$\triangle(\delta, X_{\mu,\rho}, R_\alpha) \leqslant c\delta^{\frac{2\mu}{2\mu+1}} \rho^{\frac{1}{2\mu+1}} \tag{24}$$

holds for all $\delta > 0$

Theorem (3.14)

If, for all $\tau > \tau_0 \geqslant 1$, the regularization method (R_α, α_τ) is of optimal order in $X_{\mu,\rho}$ for some $\mu > 0$ and all $\rho > 0$,with

$$\alpha_\tau = \alpha(y^\delta, \tau\delta), \quad \tau > 1 \tag{25}$$

then all regularization methods (R_α, α_τ), with $\tau > \tau_0 \geqslant 1$ are convergent for $y \in R(T)$, and they are of optimal order for all $X_{v,\rho}$ with $0 < v \leqslant \mu$ and $\rho > 0$.

Table of Content

Regularization by Projection

In regularization by projection, regularization is achieved by a finite-dimensional approximation.

In our first approach, we find the minimum-norm solution of $Tx = y$ in a finite-dimensional subspace of X.

Regularization by Projection

In regularization by projection, regularization is achieved by a finite-dimensional approximation.

In our first approach, we find the minimum-norm solution of $Tx = y$ in a finite-dimensional subspace of X.

Regularization by Projection

That is, given a sequence

$$X_1 \subset X_2 \subset X_3 \subset \ldots$$

of finite dimensional subspaces of X whose union is dense in X, x_n is the least-squares solution of minimal norm in the space X_n.

$$x_n = T_n^\dagger y, \tag{26}$$

$$T_n = TP_n \tag{27}$$

where P_n is the orthogonal projector onto X_n.

Remark

Without additional assumptions it cannot be guaranteed that x_n converges $x^\dagger$.

Theorem (3.15)

Let $y \in D(T^\dagger)$ *and let* x_n *be as discussed.*

$$x_n \to x^\dagger \text{ if and only if } \{\|x_n\|\} \text{ is bounded.} \tag{28}$$

$$x_n \to x^\dagger \text{ if and only if } \lim_{n\to\infty} \sup \|x_n\| \leqslant \|x^\dagger\| \tag{29}$$

Proposition (3.16)

Let $y \in D(T^\dagger)$ and x_n as discussed. If

$$\lim_{n\to\infty} sup \left\| \left(T_n^\dagger\right)^* x_n \right\| = \lim_{n\to\infty} sup \left\| \left(T_n^*\right)^\dagger x_n \right\| < \infty \tag{30}$$

holds, then $x_n \to x^\dagger$.

Proposition (3.17)

If T is compact and (30) is satisfied, then $x^\dagger \in R(T^*)$

Theorem (3.18)

Let $y \in D(T^\dagger)$ and x_n as discussed.If T is compact and (30) holds, then

$$\left\|x_n - x^\dagger\right\| = O(\|(I - P_n)T^*\|) \tag{31}$$

Theorem (3.19)

Let $y \in D(T^\dagger)$ and x_n as discussed. Then $x_n = P_n x^\dagger$, where P_n is a orthogonal projector onto $X_n = T^* Y_n$. Moreover,

$$x_n \to x^\dagger \quad as \quad n \to \infty \tag{32}$$

Stability analysis

Theorem (3.20)

Let $y \in D(T^\dagger)$ and let

$$\|Q_n(y - y^\delta)\| \leqslant \delta \tag{33}$$

If $\delta/\mu_n \to 0$ as $\delta \to 0$ and $n \to \infty$, where μ_n is the smallest singular value of T_n, then

$$x_n \to x^\dagger \quad as \quad \delta \to 0, n \to \infty. \tag{34}$$

Proposition (3.21)

Let T be compact with singular system $(\sigma_n; v_n, u_n)$ and let Y_n be such that $\dim(Y_n) = n$. Then

$$\mu_n \leqslant \sigma_n \tag{35}$$

Proposition (3.22)

Let T be compact with singular system $(\sigma_n; v_n, u_n)$ and let Y_n be such that $dim(Y_n) = n$. Then

$$\|(I - P_n)T^*\| \geqslant \sigma_{n+1} \tag{36}$$

If $Y_n = U_n$, then equality holds.

Thus the convergence rate of

$$\left\| x_n^\delta - x^\dagger \right\| = O(\sigma_{n+1} + \frac{\delta}{\sigma_n}) \tag{37}$$

is the best possible rate that one can expect for compact operator T, and $x^\dagger \in R(T^*)$.