Monograph:

Dodd, T.J. and Harrison, R.F. (2002) Steepest Decsent for Generalised and Regularised Solution of Linear Operator Equations. Research Report. ACSE Research Report 825 . Department of Automatic Control and Systems Engineering

Steepest Descent for Generalised and Regularised Solution of Linear Operator Equations

Tony J. Dodd and Robert F. Harrison
Department of Automatic Control and Systems Engineering
The University of Sheffield, Sheffield S1 3JD, UK
e-mail: {t.j.dodd, r.f.harrison}@shef.ac.uk

Research Report No. 825
June 2002

Abstract

Let H_1, H_2 be Hilbert spaces, T a bounded linear operator on H_1 into H_2 such that the range of T, $R(T)$, is closed. Let T^* denote the adjoint of T. In this paper, we review the convergence of the method of steepest descent to a solution of the equation $T^*Tx = T^*b, b \in H_2$, for any initial approximation $x_0 \in H_1$. The method converges to the unique minimum norm, or generalised, solution if, and only if, x_0 is in the range of T^*. Further, we establish the convergence of the method of steepest descent to the unique regularised solution $(T^*T + \mu I)^{-1}T^*b, b \in H_2$ if x_0 is in the range of T^*.

1 Introduction

Following the discussion in (Nashed 1970), let H_1 and H_2 be Hilbert spaces over the same scalars (real or complex). For any subspace, S, of H_1 or H_2, the orthogonal complement and closure of S are denoted by $S^\perp$ and $\bar{S}$ respectively. We consider a bounded linear operator, T, on H_1 into H_2. Then T^* denotes the adjoint of T, i.e. for all $x \in H_1, y \in H_2$ [1],

$$\langle Tx, y\rangle = \langle x, T^*y\rangle.$$

Let $R(T)$ and $N(T)$ denote, respectively, the range and null spaces of T. The following relations are then well known (Nashed 1970)

$$H_1 = N(T) \oplus N(T)^\perp, \tag{1}$$
$$H_2 = N(T^*) \oplus N(T^*)^\perp, \tag{2}$$
$$\{\overline{R(T)}\}^\perp = N(T^*), \qquad \overline{R(T^*)} = N(T)^\perp, \tag{3}$$
$$R(T) \text{ is closed} \Leftrightarrow R(T^*) \text{ is closed }, \tag{4}$$
$$N(T^*T) = N(T), \qquad \overline{R(T)} = \overline{R(TT^*)}. \tag{5}$$

A vector $u \in H_1$ is called a least squares solution of the operator equation [2]

$$Tx = b, \quad b \in H_2, \tag{6}$$

if $\inf\{\|Tx - b\| : x \in H_1\} = \|Tu - b\|$. It can be shown that u is a least squares solution of Eq. 6 if, and only if, u is a solution of the "normal" equation

$$T^*Tx = T^*b. \tag{7}$$

The following theorem summarises this.

Theorem 1.1 (Groetsch, 1977) *Suppose $T : H_1 \to H_2$ has closed range and $b \in H_2$, then the following conditions on $x \in H_1$ are equivalent:*

(i) $Tx = Pb$;

(ii) $\|Tu - b\| \leq \|Tx - b\|$ *for any* $x \in H_1$; *and*

(iii) $T^*Tx = T^*b$.

where P denotes the projection of b onto $R(T)$.

In the case that H_1 is finite dimensional Eq. 7 always has at least one solution since $R(T^*) = R(T^*T)$. However, when H_1 is infinite dimensional this equation may have no solutions. We therefore assume that $R(T)$ is closed. Then there always exists at least one least squares solution for each $b \in H_2$. For $N(T) \neq \{0\}$

[1] In expressing inner products, $\langle \cdot, \cdot \rangle$, it is assumed the Hilbert space to which the inner product belongs is obvious.

[2] With real applications in mind b will sometimes subsequently be referred to as the data and H_2 as the data space.

there will be infinitely many solutions since if u is a least squares solution so is $u + w$ for any $w \in N(T)$.

By the continuity and linearity of T and T^* the set, S, of all least squares solutions of Eq. 6 is a nonempty closed, convex set. Hence it contains a unique element, v, of minimal norm, i.e.,

$$||Tv - b|| \leq ||Tx - b|| \quad \text{for all } x \in H_1,$$

and

$$||v|| < ||u|| \quad \text{for all } u \in S, \quad u \neq v.$$

We then define the generalised inverse of T as the operator $T^\dagger : H_2 \to H_1$ such that $T^\dagger b = v$, i.e. that operator which assigns, to each $b \in H_2$, the unique least squares solution of minimal norm of Eq. 6. $T^\dagger$ is linear and bounded. Note that if T is invertible we have $T^\dagger = T^{-1}$. The associated least squares solution of minimal norm is the generalised solution, denoted $x^\dagger (= v)$. It can also be shown that $T^\dagger b$ is the unique least squares solution in $R(T^*)$ (Nashed 1970).

The following lemma will be used in the proofs of convergence in Sections 3 and 4.

Lemma 1.1 (Petryshyn, 1967) *Suppose T is as described above and $R(T)$ is closed. Then the restriction of T to $N(T)^\perp = R(T^*)$ has a bounded inverse. Equivalently, there exists a number $m > 0$ such that $||Tx|| \geq m||x||$ for all $x \in N(T)^\perp = R(T^*)$.*

In the next section the regularisation method for the solution of the linear operator equation is described. In Section 3 the method of steepest descent is described for generalised solutions, and subsequently extended in Section 4 to the case of regularisation solutions, of the linear operator equation.

2 Regularised Solutions

The problem of solving Eq. 6 is said to be well posed if a unique solutions exists which depends continuously on the data, b. Following (de Mol 1992) we adopt a less restrictive definition since we know that a unique generalised solution exists. The following statements are equivalent:

(i) the problem of solving $Tx = b$ is well posed;

(ii) $x^\dagger$ exists for any $b \in H_2$;

(iii) $T^\dagger$ is continuous (equivalently bounded);

(iv) $R(T)$ is closed; and

(v) $\lambda = 0$ is not an accumulation point of the spectrum of T^*T.

Otherwise, the problem is said to be ill-posed, i.e. the generalised inverse, $T^\dagger$, is unbounded. In order to obtain estimates of $x^\dagger$ which are stable to variations in the data, b, we must seek regularised solutions of the ill-posed problem.

A regularising algorithm, or regulariser, is a family of operators, $\{L_\mu\} : H_2 \to H_1$, depending on a positive parameter μ, which provide bounded approximants of the generalised inverse. The L_μ are therefore continuous linear operators for any $\mu > 0$ which approximate $T^\dagger$ in the sense that

$$\lim_{\mu \to 0+} L_\mu b = T^\dagger b \tag{8}$$

for each $b \in D(T^\dagger)$. A regularisation method is a regularisation algorithm together with a choice $\mu = \mu(\varepsilon, b^\varepsilon)$ which ensures

$$\lim_{\substack{\varepsilon \to 0 \\ b^\varepsilon \to b}} L_{\mu(\varepsilon, b^\varepsilon)} b^\varepsilon = T^\dagger b \tag{9}$$

for any noisy data, $b^\varepsilon \in H_2$, in an ε-neighbourhood of the true (noise-free) data, b (i.e. $||b - b^\varepsilon|| \leq \varepsilon$).

The parameter, μ, called the regularisation parameter, is chosen in such a way that

$$\lim_{\substack{\varepsilon \to 0 \\ b^\varepsilon \to b}} \mu(\varepsilon, b^\varepsilon) = 0. \tag{10}$$

For a given regularisation method, the function $x_\mu^\varepsilon = L_\mu b^\varepsilon$ tends to the generalised solution $x^\dagger$ corresponding to the true data when the noise, ε, tends to zero. Hence, x_μ^ε is a stable and meaningful approximate solution of the linear inverse problem.

We can bound the error in taking $x_\mu^\varepsilon = L_\mu b^\varepsilon$ as an estimate of the exact generalised solution, $x^\dagger = T^\dagger b$, using the triangle inequality as follows

$$||L_\mu b^\varepsilon - T^\dagger b|| \leq ||(L_\mu - T^\dagger)b|| + ||L_\mu(b^\varepsilon - b)||. \tag{11}$$

The first term on the RHS is the error due to the approximation of $T^\dagger$ by the regularising operator L_μ which tends to zero when $\mu \to 0$ (c.f. Eq. 8). The second term is the error due to the presence of noise on the data and is bounded by $\varepsilon||L_\mu||$. When $\mu \to 0$, $||L_\mu||$ tends to $||T^\dagger||$, which is infinite in the case of an ill-posed problem.

We now describe the particular case of the Tikhonov regularisation method. Consider the problem of minimising the functional

$$\Phi_\mu(x) = \frac{1}{2}||Tx - b||^2 + \frac{\mu}{2}||x||^2. \tag{12}$$

The quadratic functional $\Phi_\mu[x]$ has a unique minimum, denoted by x_μ, which is a solution of

$$(T^*T + \mu I)x_\mu = T^*b \tag{13}$$

where I denotes the appropriate identity operator. The operator $(T^*T + \mu I)$ is invertible for $\mu > 0$ and hence

$$x_\mu = L_\mu b, \quad \text{where } L_\mu = (T^*T + \mu I)^{-1}T^* \tag{14}$$

which is known as the Tikhonov regulariser. We can also write the regulariser as follows

$$L_\mu = T^*(TT^* + \mu I)^{-1} \tag{15}$$

from which it is clear that $x_\mu \in R(T^*) \subset N(T)^\perp$ and hence Eq. 14 defines a regularisation method (for a proof see (Groetsch 1984)).

In the sequel we will be interested in the case where $R(T)$ is closed (and hence $T^\dagger$ bounded), for which the problem of solving Eq. 6 is always well posed. However, this does not ensure that the generalised solution will be numerically stable. The relative error in the generalised solution corresponding to noisy data is bounded as follows:

$$\frac{||\delta x^\dagger||}{||x^\dagger||} \leq C(T)\frac{||\delta b^\dagger||}{||b^\dagger||} \tag{16}$$

where $C(T)$ is the condition number of the operator T, given by

$$C(T) = ||T||||T^\dagger|| = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}}. \tag{17}$$

λ_{min} and λ_{max} are, respectively, the lower and upper limits of the positive part of the spectrum of LL^*. For a large condition number, the amplification of the data error may cause the generalised solution, $x^\dagger$, to be very unstable. The problem is then said to be ill-conditioned and, even though well-posed, we still need to apply regularisation to achieve a stable solution. This is the reason for discussing regularisation for the linear operator equation, Eq. 6, even though $R(T)$ is closed.

3 The Method of Steepest Descent

Let T be a bounded linear operator on H_1 into H_2 and assume that $R(T)$ is closed. Suppose that $f : H_1 \to \mathbb{R}$ is the non-negative functional

$$f(x) = \frac{1}{2}||Tx - b||^2. \tag{18}$$

We seek a point $x^* \in H_1$ such that

$$f(x^*) = \inf\{f(x) : x \in H_1\}.$$

We assume that f is Fréchet differentiable at each point of H_1. Given an initial approximation, x_0, the method of steepest descent for minimising $f(x)$ is given by

$$x_{n+1} = x_n - \alpha_n \nabla f(x_n) \tag{19}$$

where $\nabla f(x_n)$ is the gradient of f at x_n (Groetsch 1977). That is, we move in the direction of most rapid decrease of f. The α_n are then chosen to minimise $f(x_{n+1})$ at each step.

Now, it can be shown that (Groetsch 1977)

$$\nabla f(x) = T^*Tx - T^*b \tag{20}$$

which we denote by R, and therefore

$$x_{n+1} = x_n - \alpha_n R_n. \tag{21}$$

Choosing α_n to minimise $f(x_{n+1})$

$$\begin{aligned} f(x_{n+1}) &= \frac{1}{2}\|Tx_{n+1} - b\|^2 = \frac{1}{2}\langle Tx_{n+1} - b, Tx_{n+1} - b\rangle \\ &= \frac{1}{2}\langle T(x_n - \alpha_n R_n) - b, T(x_n - \alpha_n R_n) - b\rangle. \end{aligned} \tag{22}$$

But $Tx_n - b = r_n$, the residual, and therefore

$$\begin{aligned} f(x_{n+1}) &= \frac{1}{2}\langle r_n - \alpha_n TR_n, r_n - \alpha_n TR_n\rangle \\ &= \frac{1}{2}\langle r_n, r_n\rangle - \alpha_n\langle r_n, TR_n\rangle + \frac{1}{2}\alpha_n^2\langle TR_n, TR_n\rangle. \end{aligned}$$

This is minimised when

$$\frac{\partial f(x_{n+1})}{d\alpha_n} = 0.$$

Therefore

$$-\langle r_n, TR_n\rangle + \alpha_n\langle TR_n, TR_n\rangle = 0,$$

from which

$$\alpha_n = \frac{\langle r_n, TR_n\rangle}{\|TR_n\|^2} = \frac{\langle T^* r_n, R_n\rangle}{\|TR_n\|^2} = \frac{\|R_n\|^2}{\|TR_n\|^2}. \tag{23}$$

Checking the second partial derivative condition for a minimum

$$\frac{\partial^2 f(x_{n+1})}{\partial\alpha_n^2} = \langle TR_n, TR_n\rangle = \|TR_n\|^2 \geq 0$$

and therefore the choice, Eq. 23, does in fact minimise $f(x_{n+1})$.

Theorem 3.1 *Let H_1 and H_2 be Hilbert spaces and T be a bounded linear operator on H_1 into H_2 such that its range, $R(T)$, is closed. The sequence of steepest descent defined by*

$$x_{n+1} = x_n - \alpha_n R_n, \tag{24}$$

$$R_n = T^*Tx_n - T^*b, \tag{25}$$

$$\alpha_n = \frac{\|R_n\|^2}{\|TR_n\|^2} \tag{26}$$

converges to a least squares solution of $Tx = b$ for any $x_0 \in H_1$. The sequence $\{x_n\}$ converges to the unique element $T^\dagger b$ if and only if $x_0 \in R(T^*)$.

Proof.

By Eq. 22

$$\begin{aligned} f(x_{n+1}) &= \frac{1}{2}\langle Tx_n - \alpha_n TR_n - b, Tx_n - \alpha_n TR_n - b\rangle \\ &= \frac{1}{2}\|Tx_n - b\|^2 - \alpha_n\langle Tx_n - b, TR_n\rangle + \frac{1}{2}\alpha_n^2\|TR_n\|^2. \end{aligned}$$

Substituting for α_n, Eq. 26,

$$\begin{aligned} f(x_{n+1}) &= f(x_n) - \frac{\|R_n\|^2}{\|TR_n\|^2}\langle T^*Tx_n - T^*b, R_n\rangle + \frac{1}{2}\frac{\|R_n\|^4}{\|TR_n\|^4}\|TR_n\|^2 \\ &= f(x_n) - \frac{\|R_n\|^2}{\|TR_n\|^2}\|R_n\|^2 + \frac{1}{2}\frac{\|R_n\|^4}{\|TR_n\|^2}. \end{aligned}$$

Combining the last two terms

$$f(x_{n+1}) = f(x_n) - \frac{1}{2}\frac{\|R_n\|^4}{\|TR_n\|^2}.$$

Thus $f(x_{n+1}) \leq f(x_n)$ for all n, with equality holding when $R_n = 0$.

Recursively

$$f(x_{n+1}) = f(x_0) - \frac{1}{2}\sum_{i=0}^{n}\frac{\|R_i\|^4}{\|TR_i\|^2}. \tag{27}$$

Since $f(x)$ $(= \frac{1}{2}\|Tx - b\|^2)$ is bounded below by zero

$$\sum_{i=0}^{\infty}\frac{\|R_i\|^4}{\|TR_i\|^2} < \infty. \tag{28}$$

Moreover, by Schwarz's inequality $\|TR_i\|^2 \leq \|T\|^2\|R_i\|^2$ and therefore

$$\sum_{i=0}^{\infty}\frac{\|R_i\|^4}{\|T\|^2\|R_i\|^2} = \frac{1}{\|T\|^2}\sum_{i=0}^{\infty}\|R_i\|^2 \leq \sum_{i=0}^{\infty}\frac{\|R_i\|^4}{\|TR_i\|^2}. \tag{29}$$

Hence, combining Eq. 28 and 29

$$\sum_{i=0}^{\infty}\|R_i\|^2 < \infty \tag{30}$$

Therefore

$$R_n = T^*Tx_n - T^*b \to 0 \text{ as } n \to \infty. \tag{31}$$

All that remains is to show strong convergence of $\{x_n\}$. Given Eq. 24 we have, by recursion,

$$x_{n+1} = x_0 - \sum_{i=0}^{n}\alpha_i R_i. \tag{32}$$

Therefore

$$x_m = x_0 - \sum_{i=0}^{m-1} \alpha_i R_i, \quad x_n = x_0 - \sum_{i=0}^{n-1} \alpha_i R_i. \tag{33}$$

Consequently, for $m > n$,

$$x_m - x_n = -\sum_{i=n}^{m-1} \alpha_i R_i. \tag{34}$$

Since $R_i = T^*Tx_i - T^*b \in R(T^*)$ for all i we must have $x_m - x_n \in R(T^*)$ for all m, n. By Lemma 1.1 there exists a positive number δ such that

$$\begin{aligned} \delta^2 \|x_m - x_n\|^2 \leq \|T(x_m - x_n)\|^2 &= \langle T(x_m - x_n), T(x_m - x_n)\rangle \\ &= \langle T^*T(x_m - x_n), x_m - x_n\rangle. \end{aligned} \tag{35}$$

But

$$\begin{aligned} &\langle T^*T(x_m - x_n), x_m - x_n\rangle \\ &= \langle T^*T(x_m - x_n) - T^*b + T^*b, x_m - x_n\rangle \\ &= \langle T^*Tx_m - T^*b, x_m - x_n\rangle - \langle T^*Tx_n - T^*b, x_m - x_n\rangle. \end{aligned}$$

Hence

$$\begin{aligned} &\langle T^*T(x_m - x_n), x_m - x_n\rangle \\ &\leq |\langle T^*Tx_m - T^*b, x_m - x_n\rangle| + |\langle T^*Tx_n - T^*b, x_m - x_n\rangle| \\ &\leq \|T^*Tx_m - T^*b\|\|x_m - x_n\| + \|T^*Tx_n - T^*b\|\|x_m - x_n\|. \end{aligned}$$

Combining with Eq. 35 we have, finally,

$$\begin{aligned} &\langle T^*T(x_m - x_n), x_m - x_n\rangle \\ &\leq (1/\delta)(\|T^*Tx_m - T^*b\| + \|T^*Tx_n - T^*b\|)\|T(x_m - x_n)\|. \end{aligned} \tag{36}$$

But $\{T(x_m - x_n)\}$ is bounded, say $\|T(x_m - x_n)\| \leq M$. Thus combining this with Eqs. 35 and 36

$$\begin{aligned} \delta^2\|x_m - x_n\|^2 &\leq \langle T^*T(x_m - x_n), x_m - x_n\rangle \\ &\leq (1/\delta)(\|T^*Tx_m - T^*b\| + \|T^*Tx_n - T^*b\|)\|T(x_m - x_n)\| \\ &\leq \frac{M}{\delta}(\|T^*Tx_m - T^*b\| + \|T^*Tx_n - T^*b\|) \end{aligned}$$

Recall, $T^*Tx_m - T^*b = R_m$ and $T^*Tx_n - T^*b = R_n$, therefore

$$\delta^2\|x_m - x_n\|^2 \leq \frac{M}{\delta}(R_m + R_n) \to 0 \text{ as } m, n \to \infty.$$

Hence $\{x_n\}$ is a Cauchy sequence and therefore converges to an element $u \in H_1$ and

$$\lim_{n\to\infty} f(x_n) = f(u) = \inf\{f(x) : x \in H_1\}.$$

Since

$$T^*Tx_n - T^*b = R_n \to 0$$

then

$$T^*Tu = T^*b, \quad u = T^\dagger b$$

that is, u is a least squares solution.

Now, if $x_0 \in R(T^*)$

$$x_{n+1} = x_0 - \sum_{i=0}^{n} \alpha_i R_i \in R(T^*)$$

since $R_i \in R(T^*)$ for all i. But $R(T^*)$ is closed and therefore $u \in R(T^*)$. But $T^\dagger b$ is the unique least squares solution in $R(T^*)$. Thus we must have $\{x_n\}$ converges to $u = T^\dagger b$.

For $x_0 \notin R(T^*)$, $x_0 = x_0' + x_0''$ where $x_0' \in R(T^*) = N(T)^\perp$ (since $R(T^*)$ is closed) and $x_0'' \in R(T^*)^\perp = N(T)$. Hence

$$\begin{aligned} x_{n+1} &= x_0 - \sum_{i=0}^{n} \alpha_i R_i = x_0' - \sum_{i=0}^{n} \alpha_i R_i + x_0'' \\ &= x_0' - \sum_{i=0}^{n} \alpha_i R_i + P_{N(T)} x_0 \end{aligned}$$

where $P_{N(T)}$ denotes the projection on the null space, $N(T)$. Since $x_0' \in R(T^*)$

$$x_0' - \sum_{i=0}^{n} \alpha_i R_i \to T^\dagger b, \text{ as } n \to \infty.$$

Thus, for $x_0 \notin R(T^*)$,

$$x_n \to T^\dagger b + P_{N(T)} x_0, \text{ as } n \to \infty$$

i.e. $\{x_n\}$ converges to a, not necessarily unique, least squares solution. □

4 The Regularised Case

Again, let T be a bounded linear operator on H_1 into H_2 and assume that $R(T)$ is closed. Suppose now that instead of $f(x)$ we have the non-negative functional (which is equal to $\Phi_\mu(x)$)

$$f_{reg}(x) = \frac{1}{2}\|Tx - b\|^2 + \frac{\mu}{2}\|x\|^2. \tag{37}$$

We then seek a point $x^* \in H_1$ such that

$$f_{reg}(x^*) = \inf\{f_{reg}(x) : x \in H_1\}. \tag{38}$$

Assuming that f_{reg} is Fréchet differentiable at each point of H_1, and given an initial approximation, x_0, the method of steepest descent for minimising $f_{reg}(x)$ is now given by

$$x_{n+1} = x_n - \alpha_n^{reg} \nabla f_{reg}(x_n) \tag{39}$$

where $\nabla f_{reg}(x_n)$ is the gradient of f_{reg} at x_n and the α_n^{reg} are chosen to minimise $f_{reg}(x_{n+1})$ at each step.

The gradient is given by

$$\nabla f_{reg}(x) = T^*Tx - T^*b + \mu x = R^{reg} \tag{40}$$

and therefore the steepest descent algorithm becomes

$$x_{n+1} = x_n - \alpha_n^{reg} R_n^{reg}. \tag{41}$$

Choosing α_n^{reg} to minimise $f_{reg}(x_{n+1})$

$$f_{reg}(x_{n+1}) = \frac{1}{2}\|Tx_{n+1} - b\|^2 + \frac{\mu}{2}\|x_{n+1}\|^2$$

but $x_{n+1} = x_n - \alpha_n^{reg} R_n^{reg}$, therefore

$$f_{reg}(x_{n+1}) = \frac{1}{2}\|T(x_n - \alpha_n^{reg} R_n^{reg}) - b\|^2 + \frac{\mu}{2}\|x_n - \alpha_n^{reg} R_n^{reg}\|^2.$$

Substituting $r_n = Tx_n - b$ and expanding

$$\begin{aligned}
f_{reg}(x_{n+1}) &= \frac{1}{2}\langle r_n - \alpha_n^{reg} TR_n^{reg}, r_n - \alpha_n^{reg} TR_n^{reg}\rangle + \\
&\quad \frac{\mu}{2}\langle x_n - \alpha_n^{reg} R_n^{reg}, x_n - \alpha_n^{reg} R_n^{reg}\rangle \\
&= \frac{1}{2}\langle r_n, r_n\rangle - \alpha_n^{reg}\langle r_n, TR_n^{reg}\rangle + \frac{(\alpha_n^{reg})^2}{2}\langle TR_n^{reg}, TR_n^{reg}\rangle + \\
&\quad \frac{\mu}{2}\langle x_n, x_n\rangle - \alpha_n^{reg}\mu\langle x_n, R_n^{reg}\rangle + \frac{(\alpha_n^{reg})^2}{2}\langle R_n^{reg}, R_n^{reg}\rangle.
\end{aligned}$$

This is minimised for

$$\frac{\partial f_{reg}(x_{n+1})}{\partial \alpha_n^{reg}} = 0.$$

Therefore

$$-\langle r_n, TR_n^{reg}\rangle + \alpha_n^{reg}\langle TR_n^{reg}, TR_n^{reg}\rangle - \mu\langle x_n, R_n^{reg}\rangle + \alpha_n^{reg}\langle R_n^{reg}, R_n^{reg}\rangle = 0$$

from which

$$\alpha_n^{reg} = \frac{\langle r_n, TR_n^{reg}\rangle + \mu\langle x_n, R_n^{reg}\rangle}{\langle TR_n^{reg}, TR_n^{reg}\rangle + \mu\langle R_n^{reg}, R_n^{reg}\rangle}.$$

But

$$\begin{aligned}
\langle r_n, TR_n^{reg}\rangle = \langle T^* r_n, R_n^{reg}\rangle &= \langle R_n^{reg} - \mu x_n, R_n^{reg}\rangle \\
&= \langle R_n^{reg}, R_n^{reg}\rangle - \mu\langle x_n, R_n^{reg}\rangle
\end{aligned}$$

and therefore, finally,

$$\alpha_n^{reg} = \frac{||R_n^{reg}||^2}{||TR_n^{reg}||^2 + \mu||R_n^{reg}||^2}. \tag{42}$$

As a check we have

$$\begin{aligned}\frac{\partial^2 f_{reg}(x_{n+1})}{\partial(\alpha_n^{reg})^2} &= \langle TR_n^{reg}, TR_n^{reg}\rangle + \mu\langle R_n^{reg}, R_n^{reg}\rangle \\ &= ||TR_n^{reg}||^2 + \mu||R_n^{reg}||^2 \geq 0\end{aligned}$$

for $\mu \geq 0$, therefore the particular choice of α_n^{reg}, Eq. 42, does, in fact, minimise $f_{reg}(x_{n+1})$.

Theorem 4.1 *Let H_1 and H_2 be Hilbert spaces and T be a bounded linear operator on H_1 into H_2 such that its range, $R(T)$, is closed. The sequence of regularised steepest descent defined by*

$$x_{n+1} = x_n - \alpha_n^{reg} R_n^{reg}, \tag{43}$$

$$R_n^{reg} = T^*Tx_n - T^*b + \mu x_n, \tag{44}$$

$$\alpha_n = \frac{||R_n^{reg}||^2}{||TR_n^{reg}||^2 + \mu||R_n^{reg}||^2} \tag{45}$$

converges to the unique regularised least squares solution, $(T^*T + \mu I)^{-1}T^*b$, of $Tx = b$ for any $x_0 \in R(T^*)$.

Proof.

$$\begin{aligned}f_{reg}(x_{n+1}) &= \frac{1}{2}\langle Tx_n - b - \alpha_n^{reg}TR_n^{reg}, Tx_n - b - \alpha_n^{reg}TR_n^{reg}\rangle + \\ &\quad \frac{\mu}{2}\langle x_n - \alpha_n^{reg}R_n^{reg}, x_n - \alpha_n^{reg}R_n^{reg}\rangle.\end{aligned}$$

Expanding

$$\begin{aligned}f_{reg}(x_{n+1}) &= \frac{1}{2}||Tx_n - b||^2 - \alpha_n^{reg}\langle Tx_n - b, TR_n^{reg}\rangle + \frac{(\alpha_n^{reg})^2}{2}||TR_n^{reg}||^2 + \\ &\quad \frac{\mu}{2}||x_n||^2 - \alpha_n^{reg}\mu\langle x_n, R_n^{reg}\rangle + \frac{(\alpha_n^{reg})^2}{2}\mu||R_n^{reg}||^2.\end{aligned}$$

But $f_{reg}(x_n) = \frac{1}{2}||Tx_n - b||^2 + \frac{\mu}{2}||x_n||^2$, thus

$$\begin{aligned}f_{reg}(x_{n+1}) &= f_{reg}(x_n) - \alpha_n^{reg}\langle Tx_n - b, TR_n^{reg}\rangle + \frac{(\alpha_n^{reg})^2}{2}||TR_n^{reg}||^2 - \\ &\quad \alpha_n^{reg}\mu\langle x_n, R_n^{reg}\rangle + \frac{(\alpha_n^{reg})^2}{2}\mu||R_n^{reg}||^2\end{aligned}$$

which can also be written as

$$\begin{aligned}f_{reg}(x_{n+1}) &= f_{reg}(x_n) - \alpha_n^{reg}\langle T^*Tx_n - T^*b + \mu x_n - \mu x_n, R_n^{reg}\rangle + \\ &\quad \frac{(\alpha_n^{reg})^2}{2}||TR_n^{reg}||^2 - \alpha_n^{reg}\mu\langle x_n, R_n^{reg}\rangle + \frac{(\alpha_n^{reg})^2}{2}\mu||R_n^{reg}||^2.\end{aligned}$$

Now, $T^*Tx_n - T^*b + \mu x_n = R_n^{reg}$, and thus

$$
\begin{aligned}
f_{reg}(x_{n+1}) &= f_{reg}(x_n) - \alpha_n^{reg}\langle R_n^{reg} - \mu x_n, R_n^{reg}\rangle + \frac{(\alpha_n^{reg})^2}{2}\|TR_n^{reg}\|^2 - \\
&\quad \alpha_n^{reg}\mu\langle x_n, R_n^{reg}\rangle + \frac{(\alpha_n^{reg})^2}{2}\mu\|R_n^{reg}\|^2.
\end{aligned}
$$

Therefore

$$
\begin{aligned}
f_{reg}(x_{n+1}) &= f_{reg}(x_n) - \alpha_n^{reg}\|R_n^{reg}\|^2 + \mu\alpha_n^{reg}\langle x_n, R_n^{reg}\rangle + \frac{(\alpha_n^{reg})^2}{2}\|TR_n^{reg}\|^2 \\
&\quad -\mu\alpha_n^{reg}\langle x_n, R_n^{reg}\rangle + \frac{(\alpha_n^{reg})^2}{2}\mu\|R_n^{reg}\|^2
\end{aligned}
$$

and finally

$$
f_{reg}(x_{n+1}) = f_{reg}(x_n) - \alpha_n^{reg}\|R_n^{reg}\|^2 + \frac{(\alpha_n^{reg})^2}{2}\|TR_n^{reg}\|^2 + \frac{(\alpha_n^{reg})^2}{2}\mu\|R_n^{reg}\|^2.
$$

Substituting for α_n^{reg}, Eq. 42,

$$
\begin{aligned}
f_{reg}(x_{n+1}) &= f_{reg}(x_n) - \frac{\|R_n^{reg}\|^4}{\|TR_n^{reg}\|^2 + \mu\|R_n^{reg}\|^2} + \\
&\quad \frac{1}{2}\frac{\|R_n^{reg}\|^4}{(\|TR_n^{reg}\|^2 + \mu\|R_n^{reg}\|^2)^2}\|TR_n^{reg}\|^2 + \\
&\quad \frac{\mu}{2}\frac{\|R_n^{reg}\|^4}{(\|TR_n^{reg}\|^2 + \mu\|R_n^{reg}\|^2)^2}\|R_n^{reg}\|^2 \\
&= f_{reg}(x_n) - \frac{\|R_n^{reg}\|^4}{\|TR_n^{reg}\|^2 + \mu\|R_n^{reg}\|^2} + \\
&\quad \frac{1}{2}\frac{\|R_n^{reg}\|^4}{(\|TR_n^{reg}\|^2 + \mu\|R_n^{reg}\|^2)^2}(\|TR_n^{reg}\|^2 + \mu\|R_n^{reg}\|^2)
\end{aligned}
$$

and finally

$$
f_{reg}(x_{n+1}) = f_{reg}(x_n) - \frac{1}{2}\frac{\|R_n^{reg}\|^4}{\|TR_n^{reg}\|^2 + \mu\|R_n^{reg}\|^2}.
$$

Therefore $f_{reg}(x_{n+1}) \le f_{reg}(x_n)$ for all n, with equality holding when $R_n^{reg} = 0$.

Recursively

$$
f_{reg}(x_{n+1}) = f_{reg}(x_0) - \frac{1}{2}\sum_{i=0}^{n}\frac{\|R_i^{reg}\|^4}{\|TR_i^{reg}\|^2 + \mu\|R_i^{reg}\|^2}.
$$

Again, since $f_{reg}(x)$ $(= \frac{1}{2}\|Tx - b\|^2 + \frac{\mu}{2}\|x\|^2)$ is bounded below by zero

$$
\sum_{i=0}^{\infty}\frac{\|R_i^{reg}\|^4}{\|TR_i^{reg}\|^2 + \mu\|R_i^{reg}\|^2} < \infty. \tag{46}
$$

Moreover, by Schwarz's inequality, $\|TR_i^{reg}\|^2 \leq \|T\|^2\|R_i^{reg}\|^2$ and therefore

$$\begin{aligned} \|TR_i^{reg}\|^2 + \mu\|R_i\|^2 &\leq \|T\|^2\|R_i^{reg}\|^2 + \mu\|R_i^{reg}\|^2 \\ &= \|R_i^{reg}\|^2(\|T\|^2 + \mu). \end{aligned}$$

From which

$$\begin{aligned} &\sum_{i=0}^{\infty} \frac{\|R_i^{reg}\|^4}{\|T\|^2\|R_i^{reg}\|^2 + \mu\|R_i^{reg}\|^2} = \sum_{i=0}^{\infty} \frac{\|R_i^{reg}\|^4}{\|R_i^{reg}\|^2(\|T\|^2 + \mu)} \\ &= \frac{1}{\|T\|^2 + \mu} \sum_{i=0}^{\infty} \|R_i^{reg}\|^2 \leq \sum_{i=0}^{\infty} \frac{\|R_i^{reg}\|^4}{\|TR_i^{reg}\|^2 + \mu\|R_i^{reg}\|^2}. \end{aligned} \tag{47}$$

Combining Eqs. 46 and 47

$$\sum_{i=0}^{\infty} \|R_i^{reg}\|^2 < \infty$$

and therefore $R_n^{reg} = T^*Tx_n - T^*b + \mu x_n \to 0$ as $n \to \infty$.

Again, all that remains is to show strong convergence of $\{x_n\}$. By recursion

$$x_{n+1} = x_0 - \sum_{i=0}^{n} \alpha_i^{reg} R_i^{reg} \tag{48}$$

Hence, for $m > n$,

$$x_m - x_n = -\sum_{i=n}^{m-1} \alpha_i^{reg} R_i^{reg}. \tag{49}$$

Now

$$R_i^{reg} = T^*Tx_i - T^*b + \mu x_i = (T^*T + I)x_i - T^*b$$

Therefore, if $x_0 \in R(T^*)$ we must have $R_i^{reg} \in R(T^*)$ for all i and therefore $x_m - x_n \in R(T^*)$ for all m, n. Then, by Lemma 1.1

$$\delta^2\|x_m - x_n\|^2 \leq \|T(x_m - x_n)\|^2 = \langle T^*T(x_m - x_n), x_m - x_n\rangle. \tag{50}$$

But

$$\begin{aligned} &\langle T^*T(x_m - x_n), x_m - x_n\rangle \\ &= \langle T^*T(x_m - x_n) - T^*b + T^*b - \mu x_m + \mu x_m - \mu x_n + \mu x_n, x_m - x_n\rangle \\ &= \langle T^*Tx_m - T^*b + \mu x_m, x_m - x_n\rangle - \langle T^*Tx_n - T^*b + \mu x_n, x_m - x_n\rangle - \\ &\quad \mu\langle x_m - x_n, x_m - x_n\rangle \end{aligned}$$

and therefore

$$\begin{aligned} \delta^2\|x_m - x_n\|^2 \leq\; &\langle T^*Tx_m - T^*b + \mu x_m, x_m - x_n\rangle - \\ &\langle T^*Tx_n - T^*b + \mu x_n, x_m - x_n\rangle - \mu\|x_m - x_n\|^2. \end{aligned}$$

Since $\mu \geq 0$ we have $\delta^2 + \mu > 0$ and therefore

$$\begin{aligned}
&(\delta^2 + \mu)\|x_m - x_n\|^2 \\
&\leq \langle T^*Tx_m - T^*b + \mu x_m, x_m - x_n\rangle - \langle T^*Tx_n - T^*b + \mu x_n, x_m - x_n\rangle \\
&\leq |\langle T^*Tx_m - T^*b + \mu x_m, x_m - x_n\rangle| + |\langle T^*Tx_n - T^*b + \mu x_n, x_m - x_n\rangle| \\
&\leq \|T^*Tx_m - T^*b + \mu x_m\|\|x_m - x_n\| + \|T^*Tx_n - T^*b + \mu x_n\|\|x_m - x_n\|
\end{aligned}$$

But, $\|x_m - x_n\| \leq 1/\delta\|T(x_m - x_n)\|$, and therefore

$$\begin{aligned}
&(\delta^2 + \mu)\|x_m - x_n\|^2 \leq \\
&\quad (1/\delta)(\|T^*Tx_m - T^*b + \mu x_m\| + \|T^*Tx_n - T^*b + \mu x_n\|)\|T(x_m - x_n)\|.
\end{aligned}$$

$\{T(x_m - x_n)\}$ is bounded, say $\|T(x_m - x_n)\| \leq M$ and hence

$$\begin{aligned}
&(\delta^2 + \mu)\|x_m - x_n\|^2 \\
&\leq \frac{M}{\delta}(\|T^*Tx_m - T^*b + \mu x_m\| + \|T^*Tx_n - T^*b + \mu x_n\|) \\
&= \frac{M}{\delta}(R_m^{reg} + R_n^{reg}) \to 0 \text{ as } m, n \to \infty.
\end{aligned}$$

Thus $\{x_n\}$ is a Cauchy sequence and therefore converges to an element, $u_{reg} \in H_1$, and

$$\lim_{n\to\infty} f_{reg}(x_n) = f_{reg}(u_{reg}) = \inf\{f_{reg}(x) : x \in H_1\} \tag{51}$$

Since

$$T^*Tx_n - T^*b + \mu x_n = R_n^{reg} \to 0 \tag{52}$$

then

$$(T^*T + \mu I)u_{reg} = T^*b \tag{53}$$

or

$$u_{reg} = (T^*T + \mu I)^{-1}T^*b \tag{54}$$

i.e. u_{reg} is a regularised least squares solution.

Now, we have already required that $x_0 \in R(T^*)$ (to ensure $R_i^{reg} \in R(T^*)$) and therefore

$$x_{n+1} = x_0 - \sum_{i=0}^{n} \alpha_i^{reg} R_i^{reg} \in R(T^*) \tag{55}$$

since $R_i^{reg} \in R(T^*)$ for all i. Since $R(T^*)$ is closed then $u_{reg} \in R(T^*)$. Now $(T^*T + \mu I)^{-1}T^*b$ is the unique regularised least squares solution (in $R(T^*)$) and therefore we must have that $\{x_n\}$ converges to $u_{reg} = (T^*T + \mu I)^{-1}T^*b$. □

5 Concluding Remarks

In the case of a bounded linear operator between Hilbert spaces with closed range unique generalised and regularised solutions exist. In this report convergence of the method of steepest descent to these solutions has been shown. In the case of the generalised solution, convergence is to the unique solution if, and only if, the initial approximation is in the range of the adjoint linear operator. If the initial approximation is not in the range of the adjoint, the method of steepest descent converges to a not necessarily unique solution. In the regularised case convergence is guaranteed if the initial approximation is in the range of the adjoint and convergence is to the unique solution. The case where the initial approximation is not in the range of the adjoint was not considered for the regularised case.

Acknowledgements

The authors would like to thank the UK EPSRC for their financial support under Grant No. GR/R15726/01.

References

de Mol, C. (1992). A critical survey of regularized inversion methods. In M. Bertero and E. Pike (Eds.), *Inverse Problems in Scattering and Imaging, Proceedings of the NATO Advanced Research Workshop held at Cape Cod, USA, 14-19 April 1991*, pp. 345–370. IOP Publishing.

Groetsch, C. (1977). *Generalized Inverses of Linear Operators*. Monographs and Textbooks in Pure and Applied Mathematics. Marcel Dekker.

Groetsch, C. (1984). *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind*, Volume 105 of *Research Notes in Mathematics Series*. Pitman.

Nashed, M. (1970). Steepest descent for singular linear operator equations. *SIAM Journal on Numerical Analysis 7*(3), 358–362.

Petryshyn, W. (1967). On generalized inverses and on the uniform convergence of $(I - \beta K)^n$ with applications to iterative methods. *Journal of Mathematical Analysis and Applications 18*, 417–439.