Richard Huber
richard.huber@uni-graz.at

6. Exercise Sheet

Basic information: Conjugate gradients and Tikhonov regularization
The conjugate gradients method (with which you might be familiar from classical numerical analysis) can be extended to linear inverse problems, or more precisely, to the normal equation (CG requires positive semi-definite and selfadjoint operators).

Conjugate gradients on the normal equation (CGNE):

INPUT: Operator T between Hilbert spaces X and Y, $x^* \in X$ initial guess, data $y^\delta \in Y$.

```
Initialize x_0^δ = x* ∈ X, d_0 = y^δ − T x_0^δ ∈ Y and p_1 = s_0 = T* d_0 ∈ X.
for k = 1, 2, ... unless s_{k−1} = 0 do
    q_k = T p_k
    α_k = ( ||s_{k−1}|| / ||q_k|| )^2
    x_k^δ = x_{k−1}^δ + α_k p_k
    d_k = d_{k−1} − α_k q_k
    s_k = T* d_k
    β_k = ( ||s_k|| / ||s_{k−1}|| )^2
    p_{k+1} = s_k + β_k p_k
```

Initialize $x_0^\delta = x^* \in X$, $d_0 = y^\delta - Tx_0^\delta \in Y$ and $p_1 = s_0 = T^* d_0 \in X$.
for $k = 1, 2, \dots$ **unless** $s_{k-1} = 0$ **do**
- $q_k = Tp_k$
- $\alpha_k = \left(\frac{\|s_{k-1}\|}{\|q_k\|}\right)^2$
- $x_k^\delta = x_{k-1}^\delta + \alpha_k p_k$
- $d_k = d_{k-1} - \alpha_k q_k$
- $s_k = T^* d_k$
- $\beta_k = \left(\frac{\|s_k\|}{\|s_{k-1}\|}\right)^2$
- $p_{k+1} = s_k + \beta_k p_k$

Output: Sequence of x_k^δ which converges to a solution of $T^*Tx = T^*y^\delta$, terminates if a solution is found.

The CGNE algorithm minimizes over the Krylov spaces, i.e., the iterates satisfy

$$\|Tx_k^\delta - y^\delta\| = \min\left\{\|Tx - y^\delta\| \;\middle|\; (x - x^*) \in \mathcal{K}^k(T^*(y^\delta - Tx^*), T^*T)\right\}, \qquad (1)$$

where $\mathcal{K}^k(z, A) := \operatorname{span}(z, Az, A^2 z, \dots, A^{k-1} z)$. In particular, if the algorithm terminates, the last iterate x_k^δ is a solution to the normal equation, if not, the iterates converge to a solution. Let $T\colon Z \to Y$ be linear and continuous between Banach spaces with Z reflexive and X a dense subspace of Z. For $y^\delta \in Y$ and $\alpha > 0$ one can consider the Tikhonov approach

$$x^* \in \operatorname*{argmin}_{x \in Z} \frac{\|Tx - y^\delta\|_Y^q}{q} + \frac{\alpha}{p}\|x\|_X^p, \qquad \text{where } \|x\|_X = \infty \text{ if } x \in Z \setminus X. \qquad (2)$$

The main factors in proving solvability are norms being weakly lower semi-continuous functions, coercivity, and bounded sets in Banach spaces being sequentially weakly compact (Eberlein-Shmulyan + Banach-Alaoglu, in fact for Banach spaces, this is equivalent to being reflexive). Then, one considers a minimizing sequence, which due to coercivity is bounded, and by compactness possesses a convergent subsequence, whose limit by semi-continuity is a minimizer.

Example 6.1) [Analysis of conjugate gradients]
We consider the CGNE algorithm with operator T, some data y^δ, and initial guess x^*.

a) Show via induction that $d_k = y^\delta - Tx_k^\delta$ and $s_k = T^*(y^\delta - Tx_k^\delta)$ for all k and conclude that $\|d_k\|$ is monotonously decreasing. Further, show that $p_k \in \operatorname{Rg}(T^*)$ and conclude that if $x^* \notin \ker(T)^\perp$, neither is $x_k^\delta \notin \ker(T)^\perp$. (Hence, the algorithm only converges to $T^\dagger y^\delta$ if the initial guess was in $\ker(T)^\perp$.)

b) We denote by R_K the solution operator $y^\delta \mapsto x_K^\delta$ by executing CGNE with $x^* = 0$ and stopping after at most K iterations (understanding x_K^δ as the final iteration in case of termination before K). Assume that T is compact with singular system (σ_n, u_n, v_n), and $y^\delta = \sum_{i\in\mathcal{I}} a_i v_i$ for an index set $I \subset \mathbb{N}$ and $a_i \in \mathcal{K}$ with $N = |\{\sigma_i \mid i \in I\}|$ the number of different singular values in I.

Show that R_K terminates after at most N iterations when $N < K$, while $s_K \neq 0$ if $N = \infty$.

c) Conclude for compact T with infinitely many singular values that the solution operator $R_K \colon y^\delta \mapsto x_K^\delta$ is neither linear nor continuous.

Remark. *Note that the conjugate gradients method is not linear and therefore cannot be analyzed with the spectral filtering techniques (unlike Landweber iterations). On the other hand, only considering linear methods is a restriction and non-linear methods have the potential of improving performance.*

Example 6.2) [Implementation of CG algorithm]
Implement the CGNE algorithm with starting point $x^* = 0$ and with discrepancy principle as a stopping criterion. Test it analogously to Example 5.3). Redo the test with random x^* (e.g., via 'randn' suitably scaled), what do you observe and how can this be explained?

Example 6.3) [Tikhonov regularization with semi-norms]
Let Y, Z, U be Banach spaces, Z reflexive, U separable, $T \colon Z \to Y$ linear and continuous. Further, let $A\colon \operatorname{Dom}(A) \subset Z \to U^*$ densely defined, linear and (sequentially) weak-weak* closed with closed range and finite-dimensional kernel.

We consider for $y^\delta \in Y$, $p, q \in [1, \infty[$ and $\alpha > 0$ the minimization problem

$$\min_{x\in Z} \frac{\|Tx - y^\delta\|_Y^q}{q} + \frac{\alpha}{p}\|Ax\|_{U^*}^p, \tag{3}$$

understanding $\|Ax\|_{U^*} = \infty$ if $x \notin \operatorname{Dom}(A)$.

a) Show that $\|Ax\|_{U^*}$ is (sequentially) weakly lower semi-continuous on Z. Assuming that A is injective, conclude the existence of a constant c such that $\|x\| \leq c\|Ax\|$ for all $x \in \operatorname{Dom}(A)$ (**Hint:** closed graph theorem).

b) Show the existence of a solution to (3).

Hint. *A few general observations, the norm of U^* is weak* lower semi-continuous, being a weak or weak* closed set implies being (norm) closed. By Banach-Alaoglu, bounded sets in the dual of a separable Banach space are sequentially weak* precompact. For b) note that there are continuous projections onto finite-dimensional spaces, and consequently $Z = M + N + K$ (as a direct sum of subspaces M, N, Z which intersect pairwise only at the origin) such that $M := \ker(A) \cap \ker(T)$ and $M + N = \ker(A)$.*

Remark. *The classical Tikhonov approach uses the so-called 'penalty term' $\|x\|_X$. This can be generalized in many ways, the most basic (nonetheless important) one is discussed above. Theoretically, the penalty only needs to be weak lower semi-continuous, and create coercivity in a suitable sense, but in practical applications A (and U) are chosen with further considerations in mind.*