The Problem

We consider the least squares problem: given the matrix $L \in \mathbb{C}^{m \times n}$ and right-hand side $b \in \mathbb{C}^m$, find a solution vector $x \in \mathbb{C}^n$ such that

$$\|Lx - b\|_2 = \min_{y \in \mathbb{C}^n} \|Ly - b\|_2. \tag{1}$$

It is easy to show that x is a solution of (1) if and only if

$$L^* Lx = L^* b. \tag{2}$$

The equations given in (2) are called the normal equations.

Summary of Known Methods

The following methods are known to solve a least squares problem

SIRT (Simultaneous Iterative Reconstruction Technique)
ART (Algebraic Reconstruction Technique)
CG-like methods: CGNR, CGNE, LSQR

Most of these methods are applied to the normal equations. Drawbacks of this approach are: a slow rate of convergence and sensitivity with respect to rounding errors. Both phenomena depend on $K_2(L)^2$ which is in general very large. On the other hand if b is nearly in the range of L then the sensitivity of the solution to rounding errors depends on $K_2(L)$.

The Proposed Method

We suppose that a preconditioner $T \in \mathbb{C}^{n \times m}$ is given, which is an approximation of the generalized inverse of L.

Algorithm

select x_0, eps;

$r_0 = b - Lx_0, k = 0$;

while $\|r_k\|_2 > eps$ do

  $k := k + 1, u_k^{(1)} = Tr_{k-1}, c_k^{(1)} = Lu_k^{(1)}$;

  for $i = 1, ..., k-1$ do

    $\alpha_i = c_i^* c_k^{(i)}$;

    $c_k^{(i+1)} = c_k^{(i)} - \alpha_i c_i, u_k^{(i+1)} = u_k^{(i)} - \alpha_i u_i$;

  endfor

  $c_k = c_k^{(k)} / \|c_k^{(k)}\|_2, u_k = u_k^{(k)} / \|c_k^{(k)}\|_2$;

  $x_k = x_{k-1} + u_k c_k^* r_{k-1}$;

  $r_k = r_{k-1} - c_k c_k^* r_{k-1}$;

endwhile

A Minimization Property

The following minimization property can be proved:

Theorem 1
If $\|c_i^{(i)}\|_2 \neq 0$ for every $i \leq k$, then the approximation x_k satisfies

$$\|b - Lx_k\|_2 = \min_{\xi \in x_0 + span\{u_1, \ldots, u_k\}} \|b - L\xi\|_2.$$

<u>Remark</u>
Note that this theorem is valid for every choice of $u_k^{(1)}$. If $u_k^{(1)} = Tr_{k-1}$ is used it follows that x_k satisfies

$$\|b - Lx_k\|_2 = \min_{\xi \in x_0 + K_k(TL, Tr_0)} \|b - L\xi\|_2 \tag{3}$$

Robustness of the method

The case $c_k^{(1)^*} r_{k-1} = 0$ will lead a break down at the $(k+1)^{th}$ iteration. To circumvent this we define the search direction $u_k^{(1)}$ as follows:

<u>Definition 1</u>
If $c_k^{(1)^*} r_{k-1} \neq 0$ take $u_k^{(1)} = T r_{k-1}$, else take $u_k^{(1)} = L^* r_{k-1}$.

Theorem 2
When we use $u_k^{(1)}$ as defined in Definition 1, the proposed method has no breakdown.

<u>Remark</u>
Theorem 1 and 2 imply that the algorithm converges and $\|r_k\|_2$ decreases monotonically. Since $r_k \in \mathbb{C}^m$ the algorithm finishes after m iterations.

Rate of Convergence

Inequality (3) implies $\|r_k\|_2 \leq \|p_k(LT)r_0\|_2$, for every polynomial $p_k \in \Pi_k^1$.

<u>Definition 2</u>

For the diagonalizable matrix LT there is an $S \in \mathbb{C}^{m\times m}$ such that $S^{-1}(LT)S =$diag$(\lambda_1, ..., \lambda_m)$.

Theorem 3

If $\varepsilon^{(k)}$ is defined as follows: $\varepsilon^{(k)} = \min_{p_k\in\Pi_k^1} \max_{1\leq i\leq m} |p_k(\lambda_i)|$, then the residual r_k

satisfies the following inequality: $\|r_k\|_2 \leq \varepsilon^{(k)}\|S\|_2\|S^{-1}\|_2\|r_0\|_2$.

<u>Remarks</u>

- if all eigenvalues λ_i are on the positive real axis then $\varepsilon^{(k)} \leq 2\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^k$,
- if all eigenvalues are enclosed in a circle centered at $C \in I\!R$ with $C > 0$ and having radius R with $C > R$ then $\varepsilon^{(k)} \leq \left(\frac{R}{C}\right)^k$

Implementation Aspects

- The vectors u_k are called search directions. The companion vectors $c_1, \ldots, c_k$ form an orthonormal basis for the Krylov subspace $K_k(LT, LTr_0)$. Both vector sequences should be stored in memory. If many iterations are needed, then the memory requirements can be bounded by restarting or truncating the algorithm.

- The residual r_k is obtained by updating r_{k-1}. Due to rounding errors it is possible that r_k and $b - Lx_k$ become different. If the difference is too large one can use the following strategies:
  - restart the algorithm, or
  - calculate $r_k = b - Lx_k$, and change x_k and r_k as follows:

  for $i = 1, \ldots, k$ do

  $$x_k = x_k + u_i c_i^* r_k; \; r_k = r_k - c_i c_i^* r_k;$$

  endfor

  and continue the algorithm.

Example: seismic inverse scattering

We now consider a least-squares inverse problem from exploration seismology. The aim of exploration seismology is to determine structures in the earth's subsurface from seismic measurements at the surface. This type of inverse problem is underdetermined.

We obtain the following problem:

$$L\Delta\kappa \; = \; p^{sc} \, , \qquad L \in \mathbb{C}^{m\times n}$$

with

$$L\Delta\kappa(x_j, \omega_k) \; = \; \int\limits_V \Delta\kappa(x', z') \, \mathbf{L}(x', z', x_j, \omega_k) \; dx' dz' \; ,$$

where $\mathbf{L}$ is given by

$$\mathbf{L}(x', z', x_j, \omega_k) \; = \; -\frac{i \; s(\omega_k) \; \omega_k}{8\pi c^{(0)}} \cdot \frac{\exp\{-i\omega_k \tau(x', z', x_j)\}}{R(x', z', x_j)} \; .$$

Choice of the preconditioning operator T

Our preconditioner is chosen as close as possible to L^{-1}, the generalized inverse of L. The operator T is based on a Born inversion method, where this inverse is determined approximately (with the aid of high-frequency asymptotics).

Other choices for preconditioners are:

1. T is an approximation of $(L^T L)^{-1} L^T$,
2. Use one or more iterations of the SIRT or ART method,
3. Use as preconditioner T an incomplete LQ (or QR) decomposition of L.

Numerical experiments (convergence)

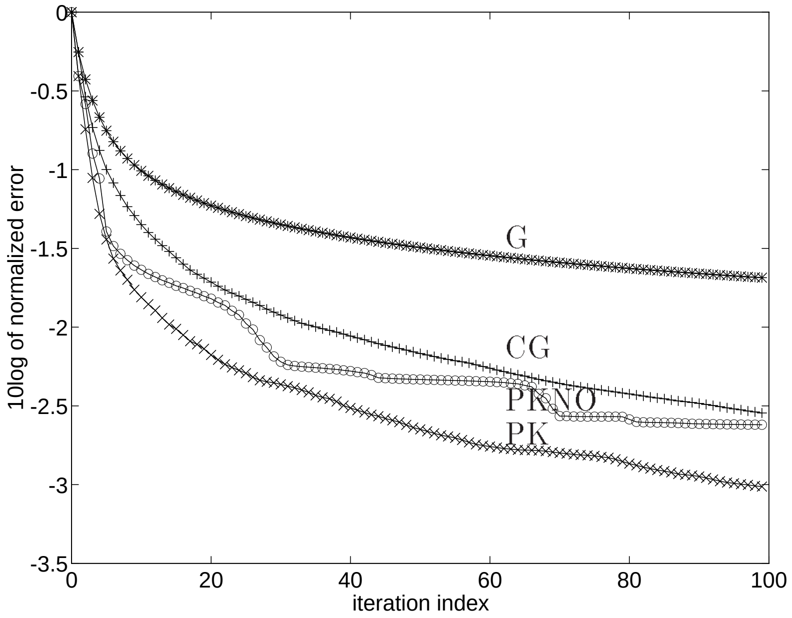

Figure 1: Rate of convergence for the following schemes: gradient scheme (G), conjugate gradient scheme (CG), preconditioned Krylov without orthogonalization (PKNO), preconditioned Krylov scheme (PK).

50x50	CG	PK	G	PKNO
k	10	5	26	5
CPU	68.45	40.5	177.37	40.4

Table 1: Number of iterations and CPU time (in seconds) for the small problem

100x100	CG	PK	G	PKNO
k	12	4	42	4
CPU	331	132	1146	127

Table 2: Number of iterations and CPU time (in seconds) for the large problem

Numerical experiments (results)

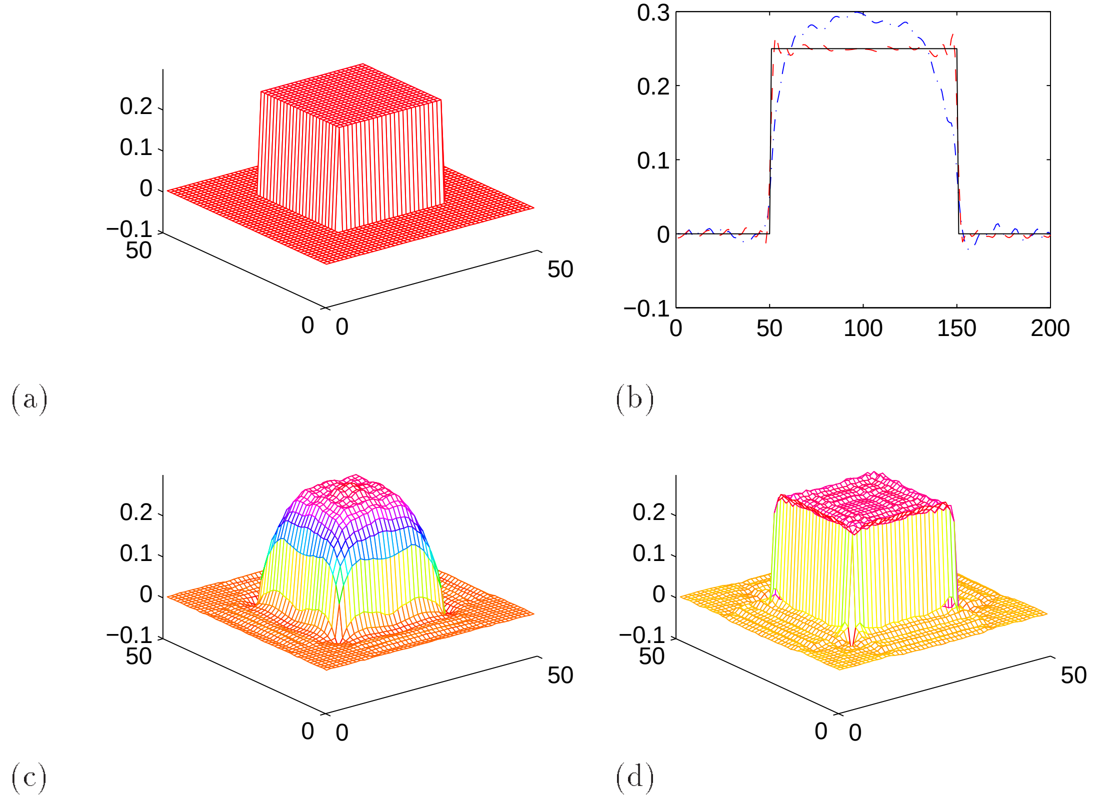

Figure 2: Comparison of imaged contrasts: (a) Actual model, (b) Cross-section of models (a), (c), and (d), (c) Estimated model after 1 iteration, (d) Estimated model after 4 iterations

Conclusions

- A Krylov subspace method has been presented to solve Least Squares problems. This method has no break down and the norm of the residuals decreases monotonically.
- The CG method and the preconditioned Krylov method have been applied to a seismic inverse scattering problem. For this problem a suitable preconditioner is obtained by using asymptotic theory.
- From the experiments it follows that the preconditioned Krylov methods are three times as fast as the CG method for large problems.

For further details see:

Kees Vuik, Agur G.J. Sevink, Gerard C. Herman
A preconditioned Krylov subspace method for the solution of least squares problems in inverse scattering
J. Comp. Phys., 123, pp. 330-340, 1996.