Preconditioned conjugate gradient method for solution of linear operator integral equation of the first kind

K. Maleknejad and D. Rostami
School of mathematics, Iran university of science and technology, P.O.Box 64451, Narmak, Tehran, Iran 16

Abstract
In this article, the solution of linear operator equations of first kind is reconstructed with definition of an optimal method. This method is a combination of optimization theorems and conjugate gradient iteration preconditioned. Idea of preconditioning over large linear systems is transformed on equivalent problem with a small condition number, which increases the rate of convergence. Only the case of positive-definite, symmetric operators on an inner product space is considered. This method can be applied also on compact operators. Moreover, several numerical examples are given.

Key words. Integral equation, Preconditioned Conjugate Gradient, Toeplitz systems, Sparse matrix.
AMS(MOS) subject classification. 45L, 45E, 65N

1-Introduction

The main difficulty in solving a Fredholm integral equation of the first kind

\[ Kz(s) = \int_a^b k(s,t)z(t)dt = g(s), \quad c \leq s \leq d \]

arises from the instability of the (generalized) inverse operator. Tikhonov [16] proposed to damp out such oscillations and regularize the solution process by taking as an approximate solution to the function z that reduces,

\[ \int_0^1 (K(z(s) - g(s))) ds + \alpha \int_0^1 \left[ p(s)z(s) + q(s) \left( \frac{dz(s)}{ds} \right)^2 \right] ds \]

where \( p \) and \( q \) are strictly positive functions and \( \alpha \) is a positive parameter. This idea may be phrased abstractly as the problem of minimizing the functional

\[ F_\alpha(z) = \|Kz - g\|^2 + \alpha\|z\| \]

where \( K \) is a compact operator from a real Hilbert space \( H_1 \) into a real Hilbert space \( H_2 \). A family of Tikhonov-type regularization methods for solving (1) has proposed, that is

\[ F_\alpha(z) = \|Kz - g\|^2 + \alpha\|z^{(n)}\|^2, \quad n = 1, 2, \ldots \]
Where $K$ is differential operator and $n$ order of derivative. In [15], Martí has shown that Sobolev spaces may serve as a tool for a partial regularization of equations of type (1). Where a Sobolev space $H^m(0,1), m \geq 0$ is defined as the completion of the set $C^m(0,1)$ (bounded continuous and $m$-times bounded continuously differentiable real function on $(0,1)$) with respect to the Sobolev norm given by:

$$
\| u \|_{m} = \left( \sum_{i=0}^{m} \int_{0}^{1} | f^{(i)}(t) |^2 dt \right)^{\frac{1}{2}}, \quad f \in C^m(0,1).
$$

Or the norm $\| \cdot \|_{m}$ of these Hilbert spaces generated by the inner product $(\cdot, \cdot)_m$ defined by:

$$
(g, g)_m = \left( \sum_{i=0}^{m} \int_{0}^{1} f^{(i)}(t) g^{(i)}(t) dt \right), \quad f, g \in C^m(0,1).
$$

Let $W_n$ be a subset of $H^m(0,1)$ with finite dimension $n$, and $(\nu_i)_{i=1}^{n}$ be a basis of $W_n$. Now we have (1) and we defined:

$$
\| Kz_n - g \|_{m} = (Kz_n - g)^{(m)} = (Kz_n, Kz_n)_m - 2(Kz_n, g)_m + (g, g)_m
$$

Since $H^m(0,1)$ is a compact space, so $z = \sum_{i=1}^{n} \nu_i$ globally converges to $z$ [6]. Now we have:

$$
\| Kz_n - g \|_{m}^2 = \sum_{i,j=1}^{n} x_i x_j (K\nu_i, K\nu_j)_m - 2 \sum_{i=1}^{n} x_i (K\nu_i, g)_m + (g, g)_m
$$

Defining $B = [(K\nu_i, K\nu_i)_m], W = [(K\nu_i, g)_m], M = [(\nu_i, \nu_i)_m]$ and substituting (6) we obtain:

$$
\| Kz_n - g \|_{m}^2 = X^T B X - 2W^T X + (g, g)_m
$$

Similarly we have

$$
\| z \|_{m}^2 = \sum_{i,j=1}^{n} x_i x_j (\nu_i, \nu_j) = X^T M X
$$

So by (7), (8) and substitution in (3) we have

$$
P_{\alpha}(X) = X^T B X - 2W^T X + (g, g) + \alpha X^T M X \quad (\alpha > 0)
$$

2-Relation between Integral Operator and optimization

A general quadratic function can be written as

$$
P_{\alpha}(X) = X^T G X - b^T X + c \quad (\alpha > 0)
$$
(11) \[ f(x) = \frac{1}{2}(x-x^*)^T G^a (x-x^*) + c^T (x-x^*) \]

where \( G^a x^* = b \), \( c^* = -\frac{1}{2} x^T G x^* \), \( G^a \) (or the Hessian matrix) is symmetric and maps differences in position into differences gradient (that is if \( g = \nabla f(x^*) \), \( g^* = \nabla f(x^*) \) then \( g - g^* = G^a (x^*-x^*) \)). Sufficient conditions which imply that \( x^* \) is a local minimizer of (10) are as follows.

**Theorem 1** Sufficient conditions for a strict and isolated local minimizer \( x^* \) are that \( g^* = 0 \) holds and that \((G^a)^* \) is positive definite, that is \( S^T (G^*) S > 0 \) for \( \forall S \neq 0 \).

**Proof:** In [13].

Other main difficulty is solving equation (10), to determine exact bounds on regularization parameter (that is \( \alpha > 0 \)). If \( \alpha \) is equal to zero in gradient (9) and \( x \) denote exact solution, then we have \( Bx - W \), and we have,

\[ (12) \]

\[ B(x - x^*) + \alpha M x - W = 0 \]

Assume that inverse of \( B \) exists and it is bounded. It is shown in [13] that if we are permitted an error \( \epsilon \) in the sense of \( ||x - y|| \leq \epsilon \), then we must have in

\[ J(x) = ||x - y||^2 + \alpha \phi(x) \]

and by attention to (8), \( \alpha = \frac{\epsilon^2}{x^T M x} \). Let \( \phi(x) = \frac{1}{x^T M x} \), satisfy conditions in [13]). So we have

\[ (13) \]

\[ \frac{\epsilon^2}{x^T M x} \geq \alpha > 0 \]

**Theorem 2** If \( \alpha \) exists and satisfies in (13), and \( B + \alpha M \) is positive definite then a minimizer \( x^* \) for (9) exists and \( x^* \) is a strict and isolated local minimizer.

**Proof:** Use theorem (1).

**Remark:** It seems that with identifying bounds for \( \alpha \), we can find out the solutions of an integral equation with compact linear operator, but it is difficult to find out the
\[ g(x) = (Kx, x) - 2(y, x) \]

Theorem 3) Let \( X \) and \( Y \) be subspaces of an inner product space \( \mathbb{Z} \), and let \( K: \mathbb{X} \to \mathbb{Y} \) be a positive-definite, symmetric operator. If \( x \) satisfies \( Kx = y \), then it minimizes the functional

\[ f(x) = \frac{1}{2} x^2 - \langle x, b \rangle \quad x \in \mathbb{R}^n. \]

The unique solution \( x^* \) of \( Ax = b \) is also the unique minimizer of \( f(x) \) as \( x \) varies over the finite range of \( X \). Let us choose a set of linearly independent elements \( v_1, v_2, \ldots, v_n \), \( n = 1, 2, \ldots \), in \( X \) and minimize \( q(x) \) over \( X = \text{span}(v_i) \). Then write \( z_n = \sum_{i=1}^{n} x_i v_i \) and have

\[ q(z) = f(x) = x^\top G x - 2x^\top b. \]

This is a quadratic function, but probably it is an improperly posed problem. Consider (4), if \( \nabla F(x) = 0 \), then the linear system to be solved

\[ Gx = b, \]

assumed to have a coefficient matrix \( G \) that is real and positive definite. The solution of this system is equivalent to the minimization of function (16), where \( x \in \mathbb{R}^n \).

A well-known iteration method for finding a minimum for a nonlinear function is the method of steepest descent [5]. For minimizing \( F(x) \) by this method, assume that an initial guess \( x \) is given. Choose a path and search for a new minimum on it, by checking conditions of theorem 2, it must be \( \nabla F(x) = 0 \). We move along the direction

\[ \delta x = -\nabla F(x_0) = g(x_0) = g_0 = b - Gx_0, \]

then solve the one-dimensional minimization problem

\[ m^2 F(x) = m g_0, \]

calling the solution \( \gamma \) (by linear search methods [18]).

Using it, define the new iteration \( x_{k+1} = x_k + \gamma g_k \quad k = 0, 1, 2, \ldots \). The method of steepest descent will converge [5,9] but the convergence is generally quite slow. The optimal
local strategy of using a direction of fastest descent is not a good strategy for finding an optimal direction for finding the global minimum. In comparison, the conjugate gradient method (CG) will be more rapid, it will take no more than \( n \) iterations, assuming there are no rounding errors.

3 -Preconditioned Conjugate Gradient Method (PCG)

A particular way of obtaining quadratic termination is to invoke the concept of the conjugacy a set of non-zero vectors \( S^{(1)}, \ldots, S^{(n)} \) to a given positive definite matrix \( G \). This is property that

\[
S^{(i)^T} GS^{(j)} = 0, \quad \forall i \neq j
\]

(18)

Theorem 4) A conjugate direction method terminates for a quadratic function in at most \( n \) exact line search, and each \( x^{(k+1)} \) is the minimizer in subspace generated by \( x^{(1)} \) and the directions \( S^{(1)}, S^{(2)}, \ldots, S^{(k)} \) (that is the set of points \( \{ x | x = x^{(1)} + \sum_{j=1}^{k} \alpha_j S^{(j)} \} \) ) Proof: In[9].

An equivalent geometric definition can be given by introducing a new inner product and norm for \( \mathbb{R}^n \): \( \langle x, y \rangle_A = x^T A y \), \( \| x \|_A = \sqrt{\langle x, x \rangle_A} = \sqrt{x^T A x} \) \( x \in \mathbb{R}^n \).

Given a set of conjugate directions \( \{ S^{(1)}, \ldots, S^{(n)} \} \) it is straightforward to solve \( Gx = b \). Let \( x^* = \sum_{k=1}^{n} x^{(k)} \) using (18),

\[
\sum_{k=1}^{n} \frac{S^{(k)^T} G x^*}{S^{(k)^T} G S^{(k)}} = 1, \quad k = 1, 2, \ldots, n
\]

(19)

We use this formula for \( x^* \) to introduce the conjugate direction method. Let \( x^{(0)} = 0 \),

\[
x^{(k)} = x^{(k-1)} + \sum_{k=1}^{n} x^{(k)} \quad 1 \leq k \leq n.
\]

(20)

Let \( r^{(k)} = Gx^{(k)} - \nabla F(x^{(k)}) \) the residual of \( x^{(k)} \) in \( Ax = b \). Obviously \( r^{(0)} = b \) and

\[
x^{(k)} = x^{(k-1)} + S^{(k)} \quad r^{(k)} = x^{(k-1)} - GS^{(k)} \quad k = 1, \ldots, n.
\]

(21)

For \( k = n \), we have \( x = x^* \), \( r^{(n)} = 0 \) and \( x \) may equal \( x^* \) with a smaller value of \( k \). To generate the directions different ways can be used, in [9, 10] simple method is using steepest descent method direction. Since \( x^{(0)} = 0 \), choose the first direction \( S^{(0)} \), as follows:

\[
S^{(0)} = -\nabla F(x^{(0)}) = x^0 + b.
\]

An inductive construction is given for the remaining directions.

Assume \( x^{(0)}, \ldots, x^{(k)} \) have been generated, along with the conjugate directions \( S^{(0)}, S^{(1)}, \ldots, S^{(k)} \). A new direction \( S^{(k+1)} \) must be chosen, assume \( x^{(k)} \neq x^* \) thus \( r^{(k)} \neq 0 \),

\[
S^{(k+1)} = -\nabla F(x^{(k)}) + S^{(k)}.
\]
otherwise, we would have the solution $x^*$ and there would be no point to continue. By theorem (3), $r^{(k)}$ is orthogonal to $M - (S^{(1)}, ..., S^{(k)})$, $r^{(k)}$ does not belong to $M'$. Putting $S_{k+1} = r^{(k)} + S_{k}$, then the condition $S^{(k)} G S^{(k)} = 0$ implies

$$
\beta_{k+1} = \frac{S^{(k)} G r^{(k)}}{S^{(k)} G S^{(k)}}
$$

The error analysis of the CG method is based on the following optimization result.

**Theorem 5** The sequence $(x^{(k)})$ of the CG method satisfies in

$$
\|x^* - x^{(k)}\|_G = \min_{\text{des}(\lambda) < k} \|x^* - q(\lambda) b\|_G
$$

**Proof:** [14], (where $q(\lambda)$ is a polynomial, for example $q(\lambda) = a + a\lambda + a\lambda^2$).

Now, if the eigenvalues of $G$ be denoted by $\sigma_1 < \sigma_2 < ... < \sigma_n$ and let $u_1, ..., u_n$ denote an orthonormal basis of eigenvectors. Using this basis, we can write

$$
x^* = \sum_{j=1}^{n} a_j u_j, \quad b = Gx^* = \sum_{j=1}^{n} a_j \sigma_j u_j
$$

with attention to (23), (24) we have:

$$
\|x^* - q(\lambda) b\|_G = \left[ \sum_{j=1}^{n} a_j^2 \left( 1 - \sigma_j^2 \right) \right]^{\frac{1}{2}}
$$

In [9] we can find better known bounds:

$$
\frac{\|x^* - x^{(k)}\|_G}{\|x^*\|_G} \leq 2 \left( \frac{1 - \sqrt{c}}{1 + \sqrt{c}} \right)^k
$$

where $c = \sigma_1 \sigma_n^{-1}$, $c$ is the condition number of $G$. The bound (26) implies that the CG methods can converge quite slowly. To increase the rate of convergence, or at least to grant a rapid rate of convergence, the problem (1) is transformed to an equivalent problem with smaller condition number. The bounds in (26), will be smaller and one expects that the iteration will converge more rapidly.

If we have (17), then matrix $Q$ is chosen such that $Q$ is nonsingular, so that we transform (17) by $(Q^{-1} Q^T x) = Q^{-1} b$, converted to $\tilde{G}x = \tilde{b}$ where $\tilde{G} = Q^{-1} G Q^T, \tilde{x} = Q^T x, \tilde{b} = Q^{-1} b$, then $\text{cond}(G)$ is smaller than $\text{cond}(\tilde{G})$ (see appendix).

Finding $Q$ requires a careful analysis of the original problem (1) and understanding the structure of $G$ in order to choose $Q$. If $G = \varphi G \varphi^T$ with $\varphi$ to be chosen with eigenvalues near 1 in magnitude, for example, if $\varphi$ is about the identity $I$, then $\varphi = Q \varphi^T$.

Approximate Cholesky factors are used in preconditioners in some cases [11]. The
Approximate Cholesky factors are used in preconditioners in some cases [11]. The success of the pre-conditioned conjugate gradient method depends upon the choice of the preconditioner. Assume matrix $M$ has given, following (17) $M^{-1}Gx=M^{-1}b$. So we have

\[ \frac{\|x^* - x^1\|^2}{\|x^0 - x^*\|^2} \leq 4 \left( \frac{\sqrt{c} - 1}{\sqrt{c} + 1} \right)^{2\sigma} \]  

(27)

It has been shown that if $\sigma = n - 1$, $\sigma$ has $p$ distinct eigenvalues then CG method converges in at most $p$ iterations. If we consider the system of equations (17) where $G$ is a symmetric positive definite matrix, then we re-write $G = M - N$, where $M$ is symmetric and positive definite, we assume that a given vector, say, $d$ can solve the system $Mz = d$ easily. So the preconditioned conjugate gradient (PCG) method proceeds as follows:

**Algorithm**

1. **Generate** $G, M, b$
   
   \[ k = 0; \quad x^0 = 0; \quad r^0 = b \]

2. **while** ($r^k \neq 0$)
   
   1. **solve** $MZ^k - b - Gx^k - r^k$
   2. $k = k + 1$
   3. If $k = 1$
      
      \[ S^0 = Z^0 \]
   4. Else
      
      \[ \beta = \frac{(r^{k-1}Z^k)}{(r^{k-2}Z^k)} \]
      
      \[ S^k = Z^k \] + \[ \beta S^{k-2} \]
   5. **end**

   \[ y = \frac{(r^{k-1}Z^k)}{(S^{k-2}GS^k)} \]

   \[ x^k = x^{k-1} + y S^k \]

   \[ r^k = r^{k-1} - y GS^k \]

   **end**

\[ x = x^k \]

Matrix $M$ in the algorithm is the preconditioning matrix. In many integral equations, matrix $G$ of the following structure arises (or sparse matrices representation of discretization of Integral operators [11]).
In this case matrix G is symmetric and positive definite so the associated eigenvalues are also positive definite and also Q is a $p \times p$ matrix. Construction of a preconditioner should be easy, but in the case of non-symmetric we can use other heuristics [1,2,8,12,17]. Suppose we wish to solve the system (17) where
\[
\begin{pmatrix}
\frac{B_1}{b_1},&\frac{B_2}{b_2},&\cdots,
\end{pmatrix}
\]
\[
\begin{pmatrix}
\frac{G_1}{0},&\cdots,0,
\end{pmatrix}
\]
\[
\begin{pmatrix}
B_1,
\cdots,0
\end{pmatrix}
\]
\[
\begin{pmatrix}
B_2
\end{pmatrix}
\]
\[
\cdots
\]
\[
\begin{pmatrix}
0,
\cdots,G_r,
\end{pmatrix}
\]
\[
\begin{pmatrix}
0,
\cdots,B_r
\end{pmatrix}
\]
\[
\begin{pmatrix}
0,
\cdots,Q
\end{pmatrix}
\]
One possibility is to apply the CG method to (17) with
\[
\begin{pmatrix}
G_1,0,\cdots,0
\end{pmatrix}
\]
\[
\begin{pmatrix}
0,G_2,\cdots,0
\end{pmatrix}
\]
\[
\begin{pmatrix}
\cdots,\cdots,\cdots,0
\end{pmatrix}
\]
\[
\begin{pmatrix}
0,0,\cdots,G_r,B_r
\end{pmatrix}
\]
\[
\begin{pmatrix}
0,0,\cdots,0,Q
\end{pmatrix}
\]
Note that the matrix $M^{-1}(G - M)$ is 2-cyclic. We obtain
\[
\sum_{i=1}^{r} B_i^T G_i \nu = \lambda^2 Q \nu
\]
instead of solving (17) directly, we could eliminate $x_{1,2,\ldots,r}$ and solve for $\xi$. This leads the system
\[(Q - B_1 G_1^{-1} B_1) \xi = c - \sum_{i=1}^{r} B_i^T G_i^{-1} B_i \]
if Q is a $p \times p$ matrix then we can apply the CG method to (29). We will obtain, the solution in at most $p$ iterations. The matrix $A = Q - \sum_{i=1}^{p} G_i B_i^T$ is the Schur complement of in G and hence A is symmetric and positive definite. Associated with (29) we can choose a preconditioner $\tilde{M}$. For example $\tilde{M} = I$ or $\tilde{M} = Q$. Note that if $\tilde{M} \xi = \tilde{N} \xi = 0$, then the convergence properties of the algorithm are determined by the eigenvalues of $M N$.
Let $M \tilde{N} w = \gamma w$. Then if $M = Q$, we have

$$
\sum_{i=1}^{n} \left( \begin{array}{c}
B \\
G B
\end{array} \right)^{i-1} w = \gamma \left( \begin{array}{c}
Q \\
v
\end{array} \right)
$$

Hence the eigenvalues of (28) are the squares of eigenvalues of (30). The rate of convergence of the two procedures are essentially the same, because we are able to eliminate half of the numerical operations. On the other hand Toeplitz matrix arises in many applications of integral equations. A Toeplitz matrix is constant along its diagonals and thus, in the symmetric case, is determined by the $n$ elements of the first row, $a_{0}, a_{1}, ..., a_{n-1}$.

$$
G = 
\begin{pmatrix}
0 & a & a & a & a \\
a & 0 & 1 & 2 & 3 \\
a & a & 0 & 1 & 2 \\
a & a & a & 0 & 1 \\
a & a & a & a & 0
\end{pmatrix}
$$

A circulant matrix, which is special case of a Toeplitz matrix, is determined by only the first $\frac{n}{2}+1$ elements of first row if $n$ is even. Each successive row contains the elements of row above shifted one to the right, with the last element wrapped around to become the first, i.e.;

$$
\begin{pmatrix}
0 & a & a & a & c \\
a & c & c & c & 1 \\
c & c & c & c & 1 \\
1 & c & c & c & 2 \\
c & 2 & 1 & 0 & c
\end{pmatrix}
$$

Since the eigenvectors of a circulant matrix are given by successive powers of the $n$th roots of unity, systems with circulant coefficient matrix are amenable to solution by the FFT in $O(n \log n)$ operations. Usually a preconditioner (31) can be obtained as follows:
The matrix $S$ is circulant preconditioner for $G$ \[4,7\]. Now, we are in position to construct circulant matrix in linear operator integral equation by using periodic splines \[6,7\].

4-Numerical Experiments

Consider \[\int_a^b k(x,y)f(x)\varphi(x)\,dx, a \leq \infty\], let $f$ be approximated by

\[f(x) = \sum_{j=0}^{M-1} \alpha_j B_h(x;x_j) = \sum_{j=0}^{M-1} \left( \frac{1}{6} \right)^k \binom{4}{k} \left( \frac{x}{k}j - k \right) \quad \text{where } x = \max(0, x) \text{ and } B_h(x;x_j) \text{ are periodic cubic cardinal B-splines with period } T = M \cdot h \text{ and knot spacing } h. \]

$M$ is the number of B-splines. The vector $\alpha = (\alpha_0, \ldots, \alpha_{M-1})^T$ of unknown coefficients is to be determined. Since $B_h(x;x_j)$ is periodic on $(0,T)$, it has the Fourier series

\[B_h(x;x_j) = \frac{1}{T} \sum_{q=-\infty}^{\infty} B_q e^{i2\pi q x/T}, \quad \text{where } w_q = \frac{2\pi q}{T} \text{ and } B_h = \int_0^T B_h(x;x_i)e^{-iwx} \, dx = h \left( \frac{\sin \left( \frac{h w_q}{2} \right)}{h w_q} \right)^{M-1} \text{ so}
\]

\[B = B_h e^{-iwx}, \quad q = \ldots, 1, 0, 1, \ldots, j = 0, \ldots, M-1 \text{ the spline in equation } f(x) = \sum_{j=0}^{M-1} \alpha_j B_h(x;x_j) \text{ has the}
\]

Fourier series $f(x) = \frac{1}{T} \sum_{q=-\infty}^{\infty} f_q \exp(iwx)$ with Fourier coefficients $f_q = \sum_{j=0}^{M-1} \alpha_j B_q$. Moreover functions $k$ and $g$ are approximated on $(0,T)$ by:

\[k(x) = \frac{1}{N} \sum_{q=0}^{N-1} k \exp(iwx), \quad g(x) = \frac{1}{N} \sum_{q=0}^{N-1} g \exp(iwx)\]
where \( k = \sum_{n=0}^{N-1} \exp(i w x) \), \( g = \sum_{q=0}^{N-1} \exp(i w x) q = 0,1, \ldots, N-1 \) with \( g(x) = g(x) \) and \( k(x) = k(x) \).

and \( w = \frac{2\pi}{Nh} \). So we have \( k \ast f = g \). Using least square method and Plancherel’s theorem we have \( \| k \ast f - g \| = \frac{T}{N} \sum_{q=0}^{N-1} | k \ast f - g |^2 \), with minimizing this functional we have the system \( Aa = b \) where \( A = W^H P W \), \( b = W^H b \), and \( W = K \cdot B \).

\( B = (B)_{1} \), \( K = (k \delta)_{1} \), \( P = (\sqrt{T} \delta)_{1} \), and \( T = \frac{1}{\| q \|} \). A is a Toeplitz matrix because \( a = \sum_{q=0}^{N-1} \exp(\frac{2\pi}{N} iq(r-s)) \), \( a = \frac{T}{N^2} k \cdot B \). If kernel is not smooth then the system \( Aa = b \) is ill-posed, so for example let \( A \) be a Hilbert-Toeplitz matrix, as follow:

i) \( A = (i-j)^{M-1} \)

ii) \( A = e^{i \cdot H} \)

Now \( C \) and \( S \) represent, respectively Tony chan’s[4], circulant preconditioner, and \( S \) Strang[3] preconditioner, then matrices \( C, S \) whose entries are given by

\[
C\left(\frac{i}{\sqrt{M}} + \frac{(M-i)a}{\sqrt{M}}\right)_{i=0, \ldots, (M-1)} \quad \text{and} \quad S\left(\sum_{j=0}^{M-1} a_{j} Q_j\right)_{j=0, \ldots, M-1}
\]

where

\[
Q = \begin{bmatrix}
0 & \cdots & 1 \\
1 & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & 0
\end{bmatrix},
\quad C = \text{circ}(c, c, \ldots, c),
\quad S = \text{circ}(s, s, \ldots, s),
\quad A = (a = a).
\]

Tables 1 and 2 showing that \( \text{cond}(A) > \text{cond}(S^{-1}A) \geq \text{cond}(C^{-1}A) \) and \( \text{cond}(A) > \text{cond}(S^{-1}A) \geq \text{cond}(C^{-1}A) \).
Table 1

<table>
<thead>
<tr>
<th>M</th>
<th>Cond(A)₁</th>
<th>Cond(S⁻¹A)₁</th>
<th>Cond(C⁻¹A)₁</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0.9E01</td>
<td>0.2E01</td>
<td>0.1E01</td>
</tr>
<tr>
<td>10</td>
<td>1.1E08</td>
<td>0.6E01</td>
<td>0.2E01</td>
</tr>
<tr>
<td>20</td>
<td>8.2E18</td>
<td>0.8E01</td>
<td>0.3E01</td>
</tr>
<tr>
<td>100</td>
<td>4.7E19</td>
<td>1.1E01</td>
<td>0.4E01</td>
</tr>
</tbody>
</table>

Table 2

<table>
<thead>
<tr>
<th>M</th>
<th>Cond(A)₂</th>
<th>Cond(S⁻¹A)₂</th>
<th>Cond(C⁻¹A)₂</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>7.6E01</td>
<td>0.1E01</td>
<td>0.1E01</td>
</tr>
<tr>
<td>10</td>
<td>1.1E35</td>
<td>2.1E00</td>
<td>1.3E00</td>
</tr>
<tr>
<td>20</td>
<td>×&gt;1.0E10</td>
<td>4.3E00</td>
<td>2.7E00</td>
</tr>
</tbody>
</table>

5- Conclusion

We have shown in this paper the difficulty of finding optimal value of regularization parameter α, in spite of having bound for it. For finding α we have to solve large system of equations with large condition numbers. By using PCG method we don’t need to find optimal value of α. Constructing special preconditioner we obtain a system of equations with a small condition number. We can summarize some advantages of our algorithm as follows:

(i) Reduction of the order of operations.
(ii) Increasing speed and accuracy of algorithm and saving memories.
(iii) Transform to parallel algorithms.

This paper is only a preliminary study with many open problems.

Appendix

If function F is the generating function of the matrices A(F) (i.e., F(α)=f₁αf₂α²⋯), then theorem 6 is satisfied.

Theorem 6 ) Let function F belong to Banach space of all 2π-periodic continues real-valued functions, we have:

(i) Then σ(A(F)) ⊆ [F, F] or |A(F)| ≤ ∥F∥ (where σ is spectrum of Toeplitz matrix).
(ii) Let $C(F)$ be the $n$-by-$n$ circulant preconditioner of $A(F)$ then $C(F)$ is the minimizer of $\| B - A(F) \|_F$ over all $n$-by-$n$ circulant matrices $B$.

(iii) Then spectrum of $A(F) - C(F)$ is clustered around zero (if $F$ is in the Wiener class).

(iv) Then spectrum of $C(F)A(F)$ is clustered around one (if $F$ is in the Wiener class).

(Definition: function $F$ is in the Wiener class if its Fourier coefficients are absolutely summable, i.e. $\sum_{k=0}^{\infty} |a_k(F)| < \infty$.) Proof [3,4].

Acknowledgements
The authors are grateful to referees who suggested some improvements of the original version.

References