Numerical Solutions for the Inverse Heat Problems in $\mathbf{R}^N$.

Raymond H. Chan
University of Hong Kong
Hong Kong

September, 1990

Abstract. We investigate the numerical solutions for the inverse heat problems in $\mathbf{R}^N$. Using discrete time and spatial sampling of the domain and sinc expansion for approximating the initial data, the problems are reduced to solving linear systems with block Toeplitz coefficient matrices. The generating functions for these systems are positive and in the Wiener class. Fast Toeplitz solvers based on the preconditioned conjugate gradient methods are implemented to solve the resulting systems.

AMS Subject Classification. 65F10, 65M99

1 Introduction.

For simplicity, we begin with the inverse heat problem in $\mathbf{R}^1$. The heat equation in $\mathbf{R}$ is the following parabolic equation:

$$\partial_t u(x,t) = \partial_x^2 u(x,t), \quad \forall x \in \mathbf{R},\ t > 0 \ ,$$

with initial values given by

$$u(x,0) = f(x), \quad \forall x \in \mathbf{R}.$$

It is well-known that if $f \in L^2(\mathbf{R})$, then

$$u(x,t) = \frac{1}{\sqrt{4\pi t}} \int_{-\infty}^{\infty} \exp\left(\frac{-(x-y)^2}{4t}\right) f(y)dy, \tag{1}$$

see for instance, John [14].

The inverse heat problem in $\mathbf{R}$ is the problem of recovering the initial data $f(y)$ for all $y \in \mathbf{R}$ when for some $t > 0$, $u(x,t)$ is given for all $x \in \mathbf{R}$. Using the idea suggested in Gilliam, Martin and Lund [11], we restrict ourselves to the following class of functions.

Definition. *A function f is said to be in the class $B(S_d)$, where*

$$S_d = \{z \in \mathbf{C} : |Im(z)| < d\},$$

if it satisfies the following three conditions:

1. *f is holomorphic in S_d,*
2. *there exists $\gamma \in (0,1)$ such that for t sufficiently large,*

$$\int_{-d}^{d} |f(t+iy)|dy = O(t^\gamma),$$

3. $$N(f) \equiv \int_{-\infty}^{\infty} \{|f(t+id)| + |f(t-id)|\}dt < \infty.$$

For $f \in B(S_d)$, the following theorem by Stenger [17] shows that they can be approximate extremely good by the *sinc expansion*, where the sinc fucntion is defined as

$$\text{sinc}(x) \equiv \frac{\sin(\pi x)}{\pi x}.$$

Theorem 1. *If $f \in B(S_d)$, then the error*

$$\epsilon(y) \equiv f(y) - \sum_{k=-\infty}^{\infty} f(kh)\, \text{sinc}\left(\frac{y-kh}{h}\right)$$

satisfies

$$\|\epsilon\|_\infty \equiv \sup_{y\in\mathbf{R}} |\epsilon(y)| \leq \frac{N(f)}{2\pi d \sinh\left(\frac{\pi d}{h}\right)} = O\left(\exp\left(-\frac{\pi d}{h}\right)\right).$$

Moreover, if there exist $\kappa, \alpha > 0$ *such that*

$$|f(x)| \leq \kappa \exp(-\alpha|x|),$$

for all $x \in \mathbf{R}$*, and if we put*

$$h = \sqrt{\frac{\pi d}{\alpha n}}$$

then

$$\epsilon_n(y) \equiv f(y) - \sum_{k=-n}^{n} f(kh) \text{ sinc}\left(\frac{y-kh}{h}\right) \tag{2}$$

satisfies

$$\|\epsilon_n(y)\|_\infty = O\left(\exp\left(-\sqrt{\pi\alpha d n}\right)\right).$$

2 The Discrete Toeplitz Systems.

For $f \in B(S_d)$, we approximate $f(y)$ in (1) by

$$f(y) \approx \sum_{k=-\infty}^{\infty} f(kh) \text{ sinc}\left(\frac{y-kh}{h}\right).$$

Then (1) becomes

$$u(x,t) \approx \frac{1}{\sqrt{4\pi t}} \sum_{k=-\infty}^{\infty} f(kh) \int_{-\infty}^{\infty} \exp\left(-\frac{(x-y)^2}{4t}\right) \text{sinc}\left(\frac{y-kh}{h}\right) dy.$$

By letting $x = x_j = jh$ and $t_0 = \left(\frac{h}{2\pi}\right)$, and after some simplification, see Gilliam et. al. [11], we finally have

$$\begin{aligned} u(x_j, t_0) &\approx \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} f(kh) \int_{-\pi}^{\pi} \exp\left(-\frac{\tau^2}{4\pi^2}\right) e^{i(k-j)\tau} d\tau \\ &\equiv \sum_{k=-\infty}^{\infty} f(kh)\beta_{k-j}, \end{aligned}$$

where

$$\beta_\ell = \frac{1}{2\pi}\int_{-\pi}^{\pi} \exp\left(-\frac{\tau^2}{4\pi^2}\right) e^{i\ell\tau} d\tau. \tag{3}$$

We remark that β_ℓ are Fourier coefficients of the function

$$g(\tau) \equiv \exp\left(-\frac{\tau^2}{4\pi^2}\right). \tag{4}$$

For fixed $n > 0$, we then have the discrete system:

$$u(x_j, t_0) = \sum_{k=-n}^{n} f(kh)\beta_{k-j},$$

or in matrix form:

$$B_{2n+1}\vec{f} = \vec{u}. \tag{5}$$

Here B_{2n+1} is the $(2n+1)$-by-$(2n+1)$ symmetric Toeplitz matrix

$$B_{2n+1} = \begin{pmatrix} \beta_0 & \beta_1 & \cdots & \beta_{2n} \\ \beta_{-1} & \beta_0 & \cdots & \beta_{2n-1} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{-2n} & \beta_{-2n+1} & \cdots & \beta_0 \end{pmatrix}.$$

The vectors $\vec{f}$ and $\vec{u}$ are given by

$$\vec{f} = (f(-nh), f(-nh+h), \cdots, f(nh-h), f(nh))^t$$

and

$$\vec{u} = (u(x_{-n}, t_0), \cdots, u(x_0, t_0), \cdots, u(x_n, t_0))^t \ .$$

Given $\vec{u}$, we can invert (5) to determine $\vec{f}$ and use it to approximate $f(y)$ for all $y \in \mathbf{R}$. More precisely, we have

$$f(y) = \sum_{k=-n}^{n} f(kh)\,\mathrm{sinc}\left(\frac{y-kh}{h}\right) + \epsilon_n(y),$$

for all $y \in \mathbf{R}$, where $\epsilon_n(y)$ is given in (2).

The inverse heat problem has now been converted into a problem of solving the Toeplitz system (5). In the following, we will consider fast solvers for such system.

3 Fast Solvers for Toeplitz Systems.

Let A_n be a Hermitian Toeplitz matrix of order n:

$$A_n = \begin{bmatrix} a_0 & a_1 & \cdot & a_m & \cdot & a_{n-2} & a_{n-1} \\ \bar{a}_1 & a_0 & a_1 & \cdot & a_m & \cdot & a_{n-2} \\ \cdot & \bar{a}_1 & a_0 & \cdot & \cdot & \cdot & \cdot \\ \bar{a}_m & \cdot & \cdot & \cdot & \cdot & \cdot & a_m \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \bar{a}_{n-2} & \cdot & \bar{a}_m & \cdot & \cdot & a_0 & a_1 \\ \bar{a}_{n-1} & \bar{a}_{n-2} & \cdot & \bar{a}_m & \cdot & \bar{a}_1 & a_0 \end{bmatrix}, \qquad (6)$$

and we are interested in solving the system

$$A_n \vec{f} = \vec{u}.$$

The first direct Toeplitz solver for such systems was invented by Levinson [16] in 1949. His algorithm requires $O(n^2)$ operations. Variants of his algorithm, for examples, Trench [20] and Zohar [22], are still widely used today. Around 1980, faster direct solvers were developed by Brent, Gustavson and Yun [4], Bitmead and Anderson [3] and Ammar and Gragg [1]. These faster solvers require only $O(n \log^2 n)$ operations. We note that all these direct solvers are not stable unless A_n is positive definite, see Bunch [5]. However, in the following, we will show that B_n in (5) above is always positive definite. Hence all these direct solvers are applicable.

Strang [18] in 1985 first proposed using iterative method such as the conjugate gradient method for solving Toeplitz systems. He noted that the Toeplitz matrix and vector multiplication of the form $A_n y$ can be computed in $O(n \log n)$ operations by first embedding A_n into a $2n$-by-$2n$ circulant matrix and then compute the circulant matrix and vector multiplication by the Fast Fourier Transform. The cost per iteration of the conjugate gradient method is thus of $O(n \log n)$. It remains to estimate the convergence rate of the method.

It is well-known that the convergence rate of the conjugate gradient method depends on the whole spectrum of A_n, see for instance Golub and van Loan [12], Axelsson and Barker [2]. In general, the more clustered the eigenvalues are, the faster will be the convergence. In order to study the spectrum of A_n, we first introduce the following definition.

Definition. *Let*

$$g(\tau) = \sum_{k=-\infty}^{\infty} a_k e^{-ik\tau}, \quad \forall \tau \in [-\pi, \pi).$$

It is said to be a function in the Wiener class if its Fourier coefficients a_k *are absolutely summable, i.e.*

$$\sum_{k=-\infty}^{\infty} |a_k| \leq \infty.$$

It is called the generating function of a Toeplitz matrix A_n *if its Fourier coefficients* a_k *are the diagonals of* A_n*.*

For the discrete inverse heat problem (5), its generating function is given by $g(\tau)$ in (4). Clearly $g(\tau) \in C^\infty[-\pi, \pi]$ and is a function in the Wiener class. Moreover,

$$e^{-0.25} \leq g(\tau) \leq 1$$

for all $\tau \in [\pi, \pi)$. It is then easy to show that the spectrum of $\sigma(B_n)$ of B_n satisfies

$$\sigma(B_n) \subseteq [e^{-0.25}, 1],$$

see for instance, Grenander and Szegö [13]. In particular, B_n are all positive definite for all n. Using standard error analysis of the conjugate gradient method, see Golub and van Loan [12], we can get

Theorem 2. *The convergence rate of the conjugate gradient method is linear, i.e. there exists* $0 \leq \gamma < 1$ *such that*

$$\lim_{q \to \infty} \frac{||e_{q+1}||}{||e_q||} = \gamma.$$

Here e_q *is the error vector at the* q *iteration.*

It follows easily that for a given tolerance, the number of iterations required for convergence is a fixed constant independent of the size n of the matrix B_n. Recall that the number of operations per iteration is of $O(n \log n)$, therefore, the total number of operations in solving the Toeplitz system is also of $O(n \log n)$. Thus the inverse heat problem (5) can be solved in $O(n \log n)$ operations.

4 Superlinear Convergence Rate.

One can even reduce the number of iterations required for convergence by using the preconditioned conjugate gradient method with some suitably chosen preconditioners. Strang in [18] proposed using circulant matrix C_n as preconditioners for solving Toeplitz systems. The main idea behind is that $C_n^{-1}y$ can be computed in $O(n \log n)$ operations by Fast Fourier Transform. Therefore the cost per iteration remains at $O(n \log n)$. However, the convergence rate of the preconditioned conjugate gradient method will then depend on the specturm of $C_n^{-1}A_n$.

For Toeplitz matrix A_n given in (6), Strang defined his preconditioner as:

$$S_n = \begin{bmatrix} a_0 & a_1 & \cdot & a_m & \cdot & \bar{a}_2 & \bar{a}_1 \\ \bar{a}_1 & a_0 & a_1 & \cdot & a_m & \cdot & \bar{a}_2 \\ \cdot & \bar{a}_1 & a_0 & \cdot & \cdot & \cdot & \cdot \\ \bar{a}_m & \cdot & \cdot & \cdot & \cdot & \cdot & a_m \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ a_2 & \cdot & \bar{a}_m & \cdot & \cdot & a_0 & a_1 \\ a_1 & a_2 & \cdot & \bar{a}_m & \cdot & \bar{a}_1 & a_0 \end{bmatrix},$$

i.e. S_n copies only the central diagonals of A_n. Here for simplicity we assume that $n = 2m$. The convergence rate of the preconditioned conjugate gradient method with preconditioner S_n was analysed by R. Chan and Strang [6].

Theorem 3. *Let g be the generating function of the matrices A_n. If g is a positive function in the Wiener class, then the spectrum of $S_n^{-1}A_n$ clustered around one. More precisely, we have for all $\epsilon > 0$, there exist $N, M > 0$, such that for all $n > N$, at most M eigenvalues of*

$$S_n^{-1}A_n - I_n = S_n^{-1}(A_n - S_n)$$

have absolute value larger than ϵ.

Using standard error analysis of the conjugate gradient method, it is also proved in R. Chan and Strang [6] that

Corollary. *Let the generating function g be a positive function in the Wiener class. Then the convergence rate of the preconditioned conjugate gradient* *method with the Strang's preconditioner is superlinear, i.e.*

$$\lim_{q\to\infty} \frac{||e_{q+1}||}{||e_q||} = 0, \tag{7}$$

where e_q *is the error vector at the* q *iteration.*

Comparing this Corollary with Theorem 2, we see that the preconditioned method converges faster than the non-preconditioned one.

It is interesting to note that if extra smoothness conditions are imposed on g, we can get a more precise bound on the convergence rate given in (7). Theorems 4 and 5 below are given in Trefethen [19] and R. Chan [8] respectively.

Theorem 4. *Suppose* $g(z) = \sum a_j z^j$ *is analytic in a neighborhood of* $|z| = 1$, *then there exist* c *and* $0 \le r < 1$ *such that*

$$\frac{||e_{q+1}||}{||e_q||} \le cr^{q^2}.$$

Theorem 5. *Let* g *be a* $(l+1)$*-times differentiable function and*

$$g^{(l+1)} \in L^1[0, 2\pi),$$

where $l > 0$. *Then there exists* c *such that for large* n,

$$\frac{||e_{2q+2}||}{||e_{2q}||} \le \frac{c}{q^{2l}}.$$

We note that the Strang's preconditioner is not the only circulant matrix enjoying these properties. Other circulant preconditioners have been proposed, see T. Chan [10], R. Chan [8], Trytyshinkov [21] and Ku and Kuo [15]. The superlinear convergence rate of these preconditioned systems are proven in R. Chan [7], R. Chan [8], R. Chan, Jin and Yeung [9] and Ku and Kuo [15] respectively. We summarize these results in the following theoerm.

Theorem 6. *Let* A_n *be a Hermitian Toeplitz matrix with generating function being a positive function in the Wiener Class. Let* C_n *be the circulant* *preconditioner proposed by either T. Chan [10], R. Chan [8], Trytyshinkov [21] or Ku and Kuo [15]. Then for n sufficiently large, the eigenvalues of the preconditioned matrix $C_n^{-1}A_n$ are clustered around one. Hence the preconditioned conjugate gradient method converges superlinearly. Moreover, these preconditioned systems converge at the same rate as the Strang's preconitioned system.*

In §6, we will apply the preconditioned conjugate gradient method to the inverse heat problem with different choices of preconditioners.

5 Inverse Heat Problems in $\mathbf{R}^N$.

For simplicity, we consider the inverse heat problem in $\mathbf{R}^2$. The case in $\mathbf{R}^N$ for general N can be treated similarly. In $\mathbf{R}^2$, one can easily check that the discrete Toeplitz system to solve is of the following form:

$$(B_n \otimes B_n)\vec{f} = \vec{u}, \tag{8}$$

where $\otimes$ is the tensor product and B_n is the discrete Toeplitz matrix in $\mathbf{R}^1$. In this case we can precondition it with $C_n \otimes C_n$, where C_n is one of the circulant preconditioners discussed above. We will see in the next section that this will be a good preconditioner.

6 Numerical Results.

In this section, we apply the Toeplitz solvers mentioned in previous sections to the inverse heat problems (5) and (8). We begin with the one-dimensional problems. We solve the Toeplitz system $B_n\vec{f} = \vec{u}$ for different n. The right hand side vector $\vec{u}$ is chosen to be the vector of all ones and the initial guess for the solution $\vec{f}$ is the zero vector. Convergence is said to occur when

$$\frac{\|r^q\|_2}{\|r^0\|_2} \le 10^{-7},$$

where r^q is the residual at the q-th iteration. The experiments were carried out on the VAX 6420 in University of Hong Kong.

The following tables show the number of iterations for different choices of preconditioners. Here R_n, S_n and T_n denote the circulant preconditioners proposed by R. Chan [8], T. Chan [10] and Strang [18] respectively and B_n denotes no preconditioning.

n	B_n	$R_n^{-1}B_n$	$S_n^{-1}B_n$	$T_n^{-1}B_n$
256	5	3	3	3
512	5	3	3	3
1024	5	3	3	3
2048	5	3	3	3
4096	5	3	3	3

Table 1. One-Dimensional Inverse Heat Problem.

The next table shows that time required to solve the inverse heat problem of a given size n. The direct solver we used here was proposed by Trench [20] and is the one currently available in the IMSL package.

n	B_n	$T_n^{-1}B_n$	Trench
512	0.260	0.250	1.119
1024	0.490	0.470	3.561
2048	1.170	1.150	14.289
4096	2.361	2.109	58.320

Table 2. Time in Seconds for Different Solvers.

Next we consider the two-dimensional discrete inverse heat problem (8). The tolerance, the right hand side vector and the initial guess are the same as in one-dimensional case.

n	$N = n^2$	B_N	$R_N^{-1}B_N$	$S_N^{-1}B_N$	$T_N^{-1}B_N$
32	1024	6	4	4	4
64	4096	6	4	4	4
128	16384	6	4	4	4
256	65536	6	4	4	4

Table 3. Two-Dimensional Inverse Heat Problem.

n	$N = n^2$	B_N	$T_N^{-1} B_N$	Trench
32	1024	1.031	0.699	0.648
64	4096	3.010	2.529	2.320
128	16384	13.52	11.95	14.20
256	65536	67.56	47.92	107.15

Table 4. Time in Seconds for Different Solvers.

References:

1. Ammar, G. and Gragg, W., *Superfast Solution of Real Positive Definite Toeplitz Systems*, SIAM J. Matrix Anal. Appl., V9 (1988), pp. 61-76.

2. Axelsson, O. and Barker, V., *Finite Element Solution of Boundary Value Problems*, Academic Press, London, 1983.

3. Bitmead, R. and Anderson, B., *Asymptotically Fast Solution of Toeplitz and Related Systems of Equations*, Linear Algebra Appl., V34 (1980), pp. 103-116.

4. Brent, R., Gustavson, F. and Yun, D., *Fast Solution of Toeplitz Systems of Equations and Computations of Pade Approximations*, J. Algorithms, V1 (1980), pp. 259-295.

5. Bunch, J., *Stability of Methods for Solving Toeplitz Systems of Equations*, SIAM J. Sci. Statist. Comput., V6 (1985), pp. 349-364.

6. Chan, R. and Strang, G., *Toeplitz Equations by Conjugate Gradients with Circulant Preconditioner*, SIAM J. Sci. Statist. Comput., V10 (1989), pp. 104-119.

7. Chan, R., *The Spectrum of a Family of Circulant Preconditioned Toeplitz Systems*, SIAM J. Numer. Anal., V26 (1989), pp. 503-506.

8. Chan, R., *Circulant Preconditioners for Hermitian Toeplitz Systems*, SIAM J. Matrix Anal. Appl., V10 (1989), pp.542-550.

9. Chan, R., Jin, X. and Yeung M., *The Spectra of Super-optimal Circulant Preconditioned Teoplitz Systems*, SIAM J. Numer. Anal., to appear.

10. Chan, T., *An Optimal Circulant Preconditioner for Toeplitz Systems*, SIAM J. Sci. Statist. Comput., V9 (1988), pp. 766-771.

11. Gilliam, D., Martin, C. and Lund J., *Analytic and Numerical Aspects of the Observation of the Heat Equation*, Proceeding of the 26th Conference on Decision and Control, 1987.

12. Golub, H. and van Loan, C., *Matrix Computations*, The Johns Hopkins University Press, Maryland, 1983.

13. Grenander, U. and Szegö, G., *Toeplitz Forms and Their Applications*, 2nd Ed., Chelsea Pub. Co., New York, 1984.

14. John, F., *Partial Differential Equations*, 4th Ed., Springer-Verlag, New York, 1982.

15. Ku, T. and Kuo, J., *Design and Analysis of Toeplitz Preconditioners*, SIPI Report # 155, USC, May 1990.

16. Levinson, N., *The Wiener rms (root-mean-square) Error Criterion in Filter Design and Prediction*, J. Math. Phys., V25 (1947), pp. 261-278.

17. Stenger, F., *Numerical Methods Based on the Whittaker Cardinal, or Sinc Functions*, SIAM Review, V23 (1981), pp. 165-224.

18. Strang, G., *A Proposal for Toeplitz Matrix Calculations*, Studies in App. Math., V74 (1986), pp. 171-176.

19. Trefethen, L., *Approximation Theory and Numerical Linear Algebra*, Numerical Analysis Report 88-7, MIT, December 1988.

20. Trench, W., *An Algorithm for the Inversion of Finite Toeplitz Matrices*, SIAM J. Appl. Math., V12 (1964), pp. 515-522.

21. Tyrtyshnikov, E., *Optimal and Super-optimal Circulant Preconditioners*, Technical Report, Dept. of Numerical Math., USSR Academy of Science, 1990.

22. Zohar, S., *Toeplitz Matrix Inversion: the Algorithm of W. F. Trench*, J. Assoc. Comput. Mach., V16 (1969) pp. 592-601.