Schnelle Polynomtransformationen und Vorkonditionierer für Toeplitz–Matrizen

Thesen zur Dissertation von Dipl.–Math. Daniel Potts

1. Ein wesentliches Anliegen der numerischen Mathematik ist die Entwicklung effizienter, numerisch stabiler Algorithmen für häufig wiederkehrende Grundaufgaben. Zu den bekanntesten schnellen Algorithmen gehört der Radix–2–Algorithmus zur Berechnung der diskreten Fourier–Transformation (DFT). Eine schnelle Realisierung der DFT wird *schnelle Fourier–Transformation* (Fast Fourier Transform, kurz: FFT) genannt. Viele Verfahren sind erst durch die Effizienz der FFT von praktischem Interesse, so z.B. Polynommultiplikation, Matrizenmultiplikation, Matrix–Vektor–Multiplikation, Invertierung großer strukturierter Matrizen oder trigonometrische Interpolation auf feinen Gittern.
Da einerseits die Eingabewerte bei den meisten Anwendungen reell sind und da andererseits die DFT eine komplexe Arithmetik benötigt, interessiert man sich auch für reelle Versionen der DFT. Dies führt zu verschiedenen diskreten trigonometrischen Transformationen, den *diskreten Kosinus–Transformationen* (DCT), den *diskreten Sinus–Transformationen* (DST) und der diskreten Hartley–Transformation. Wir werden die schnellen reellen Algorithmen für die diskreten trigonometrischen Transformationen nutzen (siehe [7, 2, 1]).

2. Viele Anwendungen der DFT stehen in engem Zusammenhang mit zirkulanten Matrizen $\boldsymbol{C}_N \in \mathbb{C}^{N,N}$. Diese lassen sich mit Hilfe der Fourier–Matrix $\boldsymbol{F}_N$ diagonalisieren, d.h., es existiert eine eindeutige Darstellung der Form

$$\boldsymbol{C}_N \;=\; \bar{\boldsymbol{F}}_N \,(\operatorname{diag}\, \boldsymbol{c}_N)\, \boldsymbol{F}_N \quad (\boldsymbol{c}_N \in \mathbb{C}^N)\,. \tag{1}$$

Um ein großes lineares Gleichungssystem mit einer Toeplitz–Koeffizientenmatrix (kurz: Toeplitz–System) zu lösen, hatte Strang in [8] die großartige Idee, eine reguläre zirkulante Matrix, die die gegebene Toeplitz–Matrix „approximiert“, als Vorkonditionierer in einem konjugierten Gradienten–Verfahren (PCG–Verfahren) zu verwenden. Unter geeigneten Voraussetzungen erzielt man eine superlineare Konvergenz des PCG–Verfahrens. Der Erfolg dieser Idee löste seit 1986 eine wahre Flut von Arbeiten aus.

3. Aus diesen Entwicklungsrichtungen leiten sich die Zielstellungen meiner Arbeit ab:

(i) Die Eigenschaften diagonalisierbarer Matrizen bezüglich einer Kosinus– bzw. Sinus–Matrix sind herzuleiten.
(ii) Die DCT kann man auch als *diskrete Polynomtransformation* (DPT) auffassen, bei der man die Chebyshev–Polynome erster Art benutzt. Die Interpolationseigenschaft der Chebyshev–Polynome liefert einen schnellen Algorithmus zur Polynommultiplikation (siehe [2]). Auf dieser Basis ist ein schneller Algorithmus für die DPT herzuleiten, wobei die Polynome einer beliebigen Drei–Term–Rekursion genügen sollen.

(iii) Unter Verwendung diagonalisierbarer Matrizen bezüglich einer Kosinus– bzw. Sinus–Matrix soll eine einheitliche Theorie für Vorkonditionierer von reellen Toeplitz–Systemen entwickelt werden.
(iv) Die theoretischen Ergebnisse sind durch umfangreiche Testrechnungen zu belegen.

4. Um reelle Algorithmen zu entwickeln, werden wir zunächst verschiedene orthogonale trigonometrische Matrizen $\boldsymbol{O}_N$ betrachten. Unter einer *trigonometrischen Transformation der Länge N* verstehen wir eine lineare Abbildung von $\mathbb{R}^N$ in sich, die jedem Vektor $\boldsymbol{x}_N := (x_j)_{j=0}^{N-1} \in \mathbb{R}^N$ den Vektor $\hat{\boldsymbol{x}}_N := (\hat{x}_j)_{j=0}^{N-1} \in \mathbb{R}^N$ mit

$$\hat{\boldsymbol{x}}_N \;:=\; \boldsymbol{O}_N\, \boldsymbol{x}_N$$

zuordnet. Zur Berechnung dieser Matrix–Vektor–Multiplikation sind schnelle, reelle und numerisch stabile Algorithmen vorhanden (siehe [7, 2, 1]). Motiviert durch die Darstellung (1) untersuchen wir reelle Matrizen der Form

$$\boldsymbol{O}'_N(\mathrm{diag}\ \boldsymbol{d}_N)\, \boldsymbol{O}_N \quad (\boldsymbol{d}_N \in \mathbb{R}^N)\ . \tag{2}$$

Dies führt auf spezielle Toeplitz–plus–Hankel–Matrizen. Beispielsweise gilt für die DCT–II mit $\boldsymbol{C}_N^{II} := \left(\frac{2}{N}\right)^{1/2} \left(\varepsilon_j^N\ \cos\frac{j(2k+1)\pi}{2N}\right)_{j,k=0}^{N-1} \in \mathbb{R}^{N,N}$, $\varepsilon_j^N = 1/\sqrt{2}$ $(j = 0, N)$ und $\varepsilon_j^N = 1$ $(j = 1, \ldots, N-1)$ die Gleichung

$$(\boldsymbol{C}_N^{II})'\ \mathrm{diag}(d_0, \ldots, d_{N-1})'\ \boldsymbol{C}_N^{II} =$$

$$\frac{1}{2}\begin{pmatrix} a_0 & a_1 & \ldots & a_{N-2} & a_{N-1} \\ a_1 & a_0 & \ldots & a_{N-3} & a_{N-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{N-2} & a_{N-3} & \ldots & a_0 & a_1 \\ a_{N-1} & a_{N-2} & \ldots & a_1 & a_0 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} a_1 & a_2 & \ldots & a_{N-1} & 0 \\ a_2 & a_3 & \ldots & 0 & a_{N-1} \\ \vdots & \vdots & \iddots & \vdots & \vdots \\ a_{N-1} & 0 & \ldots & a_3 & a_2 \\ 0 & a_{N-1} & \ldots & a_2 & a_1 \end{pmatrix}$$

mit

$$(d_0, \ldots, d_N)' \;:=\; \tilde{\boldsymbol{C}}_{N+1}^{I}\ (a_0, \ldots, a_{N-1}, 0)'\,,$$

$$\tilde{\boldsymbol{C}}_{N+1}^{I} := \left(\frac{2}{N}\right)^{1/2} \left((\varepsilon_k^N)^2\ \cos\frac{jk\pi}{N}\right)_{j,k=0}^{N} \in \mathbb{R}^{N+1,N+1}\ .$$

Diese Ergebnisse nutzen wir, um eine neue Variante zur schnellen Multiplikation einer reellen Toeplitz–Matrix mit einem beliebigen Vektor anzugeben.

5. In Verallgemeinerung von diskreten trigonometrischen Transformationen betrachten wir DPT. Sei $\{P_n\}_{n\in\mathbb{N}_0}$ eine Folge von Polynomen, die durch die Drei–Term–Rekursion

$$\begin{aligned} P_n(x) \;&:=\; (\alpha_n x + \beta_n)\, P_{n-1}(x) \;+\; \gamma_n P_{n-2}(x) \quad (n = 1, 2, \ldots)\,, \\ P_{-1}(x) \;&:=\; 0\,, \quad P_0(x) \;:=\; 1 \end{aligned}$$

mit $\alpha_n, \beta_n, \gamma_n \in \mathbb{R}$ und $\alpha_n > 0\,, \gamma_n \neq 0$ $(n \in \mathbb{N})$ gegeben ist. Ein reeller schneller Algorithmus für die DPT wird entwickelt (siehe [5]), d.h., für gegebene Koeffizienten

$a_k \in \mathbb{R}$ $(k = 0, \ldots, N)$ berechnen wir die DPT$(N+1, M+1)$: $\mathbb{R}^{N+1} \to \mathbb{R}^{M+1}$ mit $(M \geq N)$, die durch

$$\hat{a}_j := \sum_{k=0}^{N} a_k \, P_k(\cos \frac{j\pi}{M}) \quad (j = 0, \ldots, M)$$

definiert ist. Die bekannten Konzepte zur Herleitung schneller Algorithmen, wie die Teile–und–Herrsche–Strategie, werden dabei benutzt. Zusammen mit einer schnellen Polynommultiplikation erhalten wir einen Algorithmus der Komplexität $\mathcal{O}(N \log^2 N) + \mathcal{O}(M \log M)$. Die Algorithmen zur DPT werden zunächst für Zweierpotenzlängen angegeben. Ohne Schwierigkeiten lassen sich diese auf beliebige Transformationslängen N ausdehnen. Ursprünglich wurde ein Algorithmus unter Verwendung der DFT von Driscoll und Healy (siehe [4]) entwickelt. Der neue Algorithmus benutzt die DCT und ist einfacher.

6. Wir entwickeln *Vorkonditionierer* zur iterativen Lösung von *reellen, positiv definiten, symmetrischen Toeplitz–Systemen* $\boldsymbol{T}_N \boldsymbol{x} = \boldsymbol{b}$ mittels PCG–Verfahren. Dazu sind Vorkonditionierer $\boldsymbol{M}_N$ mit den folgenden vier Eigenschaften gesucht:

(E1) Die Matrizen $\boldsymbol{M}_N$ sind symmetrisch und positiv definit.

(E2) Die arithmetische Komplexität zur Berechnung von $\boldsymbol{M}_N$ ist nicht größer als die arithmetische Komplexität bei einer Matrix–Vektor–Multiplikation mit der Koeffizientenmatrix $\boldsymbol{T}_N$.

(E3) Die arithmetische Komplexität zur Lösung von $\boldsymbol{M}_N \boldsymbol{z} = \boldsymbol{r}$ ist nicht größer als die arithmetische Komplexität zur Berechnung von $\boldsymbol{T}_N \boldsymbol{p}$.

(E4) Die Matrizen $\boldsymbol{M}_N^{-1} \boldsymbol{T}_N$ $(N \in \mathbb{N})$ haben „geclusterte“ Eigenwerte um 1.

7. Ausgehend von den Matrizen (2) haben wir reelle Vorkonditionierer aus der Algebra

$$\mathcal{A}_{\boldsymbol{O}_N} := \{\boldsymbol{O}_N'(\text{diag } \boldsymbol{d}_N)\, \boldsymbol{O}_N : \quad \boldsymbol{d}_N \in \mathbb{R}^N\}$$

untersucht und trigonometrische Strang–Typ–Vorkonditionierer erhalten. Ähnlich wie T. Chan optimale zirkulante Vorkonditionierer (siehe [3]) entwickelt hat, werden wir optimale trigonometrische Vorkonditionierer berechnen. Dazu werden wir ein *neues allgemeines Konzept* benutzen. Damit ist es dann möglich, auch superoptimale trigonometrische Vorkonditionierer schnell zu berechnen. Wir zeigen, daß die trigonometrischen Vorkonditionierer die Eigenschaften (E1) – (E4) erfüllen.
Aus vielen numerischen Testbeispielen ergeben sich folgende Beobachtungen:

(i) Unter zusätzlichen Voraussetzungen konvergiert das PCG–Verfahren mit dem Strang–Typ–Vorkonditionierer, mit dem optimalen trigonometrischen Vorkonditionierer und mit dem superoptimalen trigonometrischen Vorkonditionierer superlinear.
(ii) In allen Fällen liefert ein trigonometrischer Vorkonditionierer in weniger Schritten ein gleich gutes Ergebnis wie ein zirkulanter Vorkonditionierer. Da die arithmetische Komplexität mit einer diskreten trigonometrischen Transformation geringer als bei einer DFT mit reellen Eingabedaten ist, sind diese Vorkonditionierer für reelle Toeplitz–Matrizen vorzuziehen.

(iii) In allen getesteten Beispielen liefert der superoptimale trigonometrischen Vorkonditionierer keine besseren Ergebnisse als der Strang–Typ–Vorkonditionierer oder der optimale trigonometrische Vorkonditionierer.

Diese Ergebnisse wurden auf symmetrische Block–Toeplitz–Matrizen mit symmetrischen Toeplitz–Blöcken übertragen (siehe [6]).

8. In dieser Arbeit werden wir erstmals optimale trigonometrische Vorkonditionierer aus der Algebra $\mathcal{A}_{\boldsymbol{O}_N}$ für die Normalgleichung einer unsymmetrischen und rechteckigen Toeplitz–Matrix schnell berechnen. Für schlecht konditionierte Toeplitz–Matrizen geben wir einfache Vorkonditionierer aus der Algebra $\mathcal{A}_{\boldsymbol{O}_N}$ an. Falls die symmetrische Toeplitz–Matrix von einer nichtnegativen Funktion $f \in C([-1,1])$ erzeugt wird, so definieren wir vereinfachte Strang–Typ–Vorkonditionierer $\boldsymbol{M}_N(f, \boldsymbol{O}_N)$. Für die DCT–II gilt beispielsweise

$$\boldsymbol{M}_N(f, \boldsymbol{C}_N^{II}) \;:=\; (\boldsymbol{C}_N^{II})' \operatorname{diag}\left(f(\cos(\tfrac{k\pi}{N}))_{k=0}^{N-1}\right) \boldsymbol{C}_N^{II} \,.$$

Im Fall $f(\cos(\frac{k\pi}{N})) > 0$ für $k = 0, \ldots, N-1$ weisen wir die Eigenschaften (E1) – (E4) für $\boldsymbol{M}_N(f, \boldsymbol{C}_N^{II})$ nach.
Umfangreiche numerische Testrechnungen belegen den Wert der neuen theoretischen Ergebnisse.

Literatur

[1] G. Baszenski. Programmpaket zur Berechnung diskreter trigonometrischer Transformationen. 1995. http://www.iuk.fh-dortmund.de/~baszenski/.

[2] G. Baszenski and M. Tasche. Fast polynomial multiplication and convolution related to the discrete cosine transform. *Linear Algebra Appl.*, 252:1 – 25, 1997.

[3] T. F. Chan. An optimal circulant preconditioner for Toeplitz systems. SIAM *J. Sci. Statist. Comput.*, 9:766 – 771, 1988.

[4] J. Driscoll and D. M. Healy. Computing Fourier transforms and convolutions on the 2–sphere. *Adv. Appl. Math.*, 15:202 – 240, 1994.

[5] D. Potts, G. Steidl, and M. Tasche. Fast algorithms for discrete polynomial transforms. *Math. Comp.* in print.

[6] D. Potts, G. Steidl, and M. Tasche. Trigonometric preconditioners for block Toeplitz systems. In G. Nürnberger, J. W. Schmidt, and G. Walz, editors, *Multivariate Approximation and Splines*, pages 219 – 234, Basel, 1997. Birkhäuser.

[7] G. Steidl. Fast radix–p discrete cosine transform. *Appl. Algebra in Engrg. Comm. Comput.*, 3:39 – 46, 1992.

[8] G. Strang. A proposal for Toeplitz matrix calculations. *Studies in Appl. Math.*, 74:171 – 176, 1986.