Supplementary Material: Structured Transforms for Small-Footprint Deep Learning

Vikas Sindhwani
Google, New York
sindhwani@google.com

1 Preliminaries: Proofs

Proposition 1.1 (Properties of f-unit-circulant matrices)**.** *: (1) Downward shift-and-scale action on a vector:* $\mathbf{Z}_f\mathbf{v} = [fv_n, v_1, v_2 \ldots v_{n-1}]^T$.
(2) Upward shift-and-scale transposed action on a vector: $\mathbf{Z}_f^T\mathbf{v} = [v_1, v_2, \ldots v_{n-1}, fv_0]^T$.
(3) f-potent: $\mathbf{Z}_f^n = f\mathbf{I}$
(4) Inverse: $\mathbf{Z}_f^{-1} = \mathbf{Z}_{f^{-1}}^T$.
(5) $\mathbf{Z}_f^T = \mathbf{J}\mathbf{Z}_f\mathbf{J}$ *and* $\mathbf{Z}_f = \mathbf{J}\mathbf{Z}_f^T\mathbf{J}$.

Proof of Proposition 1.1. The first two properties can be directly verified from the definition. The third property - which will turn out to be crucial - follows since applying $\mathbf{Z}_f$ n times cycles the vector back to its original form but with all entries scaled by f. The fourth property follows becauses $\mathbf{Z}_{f^{-1}}^T$ cancels the downward shift-and-scale action of $\mathbf{Z}_f$. The fifth property can be verified by observing the shifting/reversing actions of the left and right hand side on an arbitrary vector. □

Lemma 1.2 (Rank 2 displacement property of Toeplitz matrices)**.** *For any Toeplitz matrix* $\mathbf{T}$ *and scalars* e, f*, then* $rank(\nabla_{\mathbf{Z}_e,\mathbf{Z}_f}[\mathbf{T}]) \leq 2$

Proof of Lemma 1.2. Let $\mathbf{t} = [\mathbf{t}_-, t_0, \mathbf{t}_+]^T$ where $\mathbf{t}_- = [t_{-(n-1)}, \ldots t_{-1}]^T$ and $\mathbf{t}_+ = [t_1 \ldots t_{n-1}]$. The notation $\mathbf{T}(t)_{ij} = t_{i-j}$ denotes an $n \times n$ Toeplitz matrix. The following can be seen from the shift-and-scale properties of f-unit-circulants.

$$\mathbf{Z}_f\mathbf{T}(\mathbf{t}) = \begin{bmatrix} f\mathbf{J}\mathbf{t}_+ & ft_0 \\ \mathbf{T}(\mathbf{t}') & \mathbf{t}_- \end{bmatrix}, \quad \mathbf{T}(\mathbf{t})\mathbf{Z}_f = \begin{bmatrix} \mathbf{J}\mathbf{t}_- & ft_0 \\ \mathbf{T}(\mathbf{t}') & f\mathbf{t}_+ \end{bmatrix}$$

where $\mathbf{t}' = [t_{-(n-2)} \ldots t_{-1}, t_0, t_1 \ldots t_{n-2}]$. From this, it should be clear that for any scalars e, f, the following is true.

$$\nabla_{\mathbf{Z}_e,\mathbf{Z}_f}[\mathbf{T}(\mathbf{t})] = \mathbf{Z}_e\mathbf{T}(\mathbf{t}) - \mathbf{T}(\mathbf{t})\mathbf{Z}_f = \begin{bmatrix} \mathbf{J}(e\mathbf{t}_+ - \mathbf{t}_-) & (e-f)t_0 \\ \mathbf{0}_{(n-1)\times(n-1)} & (\mathbf{t}_- - f\mathbf{t}_+) \end{bmatrix}$$

Since any matrix of the form $\begin{bmatrix} \mathbf{u}^T & w \\ \mathbf{0}_{(n-1)\times(n-1)} & \mathbf{v} \end{bmatrix} = \mathbf{e}_0[\mathbf{u} \;\; \frac{w}{2}]^T + \begin{pmatrix} \frac{w}{2} \\ \mathbf{v} \end{pmatrix}\mathbf{e}_n^T$, it follows that $\nabla_{\mathbf{Z}_e,\mathbf{Z}_f}[\mathbf{T}(\mathbf{t})]$ has rank at most 2. □

Theorem 1.3 (Theorem 3.3, [2])**.** $\nabla_{\mathbf{A},\mathbf{B}}$ *is invertible if and only if* $\lambda_i(\mathbf{A}) \neq \lambda_j(\mathbf{B})$ *and* $\triangle_{\mathbf{A},\mathbf{B}}$ *is invertible if and only if* $\lambda_i(\mathbf{A})\lambda_j(\mathbf{B}) \neq 1$*, for any pair of eigenvalues* $\lambda_i(\mathbf{A}), \lambda_j(\mathbf{B})$ *of* $\mathbf{A}, \mathbf{B}$ *respectively.*

Corollary 1.4. $\nabla_{\mathbf{Z}_1,\mathbf{Z}_{-1}}$ *is invertible.*

Proof. Let λ, μ be an eigenvalue of $\mathbf{Z}_1$ and $\mathbf{Z}_{-1}$ respectively. Then, for the associated respective eigenvectors $\mathbf{v}, \mathbf{u}$:

$$\begin{aligned} \mathbf{Z}_1\mathbf{v} &= \lambda\mathbf{v} \\ \mathbf{Z}_{-1}\mathbf{u} &= \mu\mathbf{u} \end{aligned}$$

The first equation above implies $[v_n, v_1 \dots v_{n-1}]^T = \lambda[v_1 \dots v_n]$ which in turn implies that $v_n = \lambda v_1, v_1 = \lambda v_2, \dots v_{n-1} = \lambda v_n$. It is easy to see that since $\mathbf{v} \neq 0$, it must be true that $\lambda^n = 1$. A similar argument for $\mathbf{Z}_1$ shows that $\mu^n = -1$. Hence, $\lambda \neq \mu$ and therefore $\mathbf{Z}_1, \mathbf{Z}_{-1}$ satisfy the invertibility conditions of Theorem 1.3. □

Several invertibility formulae in [2] rely on the following simple but far reaching result:

Theorem 1.5 (Theorem 3.3, [2])**.** *For any $m \times m$ matrix $\mathbf{A}$, $n \times n$ matrix $\mathbf{B}$, $m \times n$ matrix $\mathbf{M}$ and for all natural numbers k, we have,*

$$\mathbf{M} = \mathbf{A}^k\mathbf{M}\mathbf{B}^k + \sum_{i=0}^{k-1} \mathbf{A}^i \triangle_{\mathbf{A},\mathbf{B}}[\mathbf{M}]\mathbf{B}^i \tag{1}$$

Proof of Theorem 1.5. For $k = 0$, the identity is trivial. Let us show that it holds for $k+1$ under the assumption that it is true for k. Multiplying the identity on the left by $\mathbf{A}$ and right by $\mathbf{B}$, we have,

$$\begin{aligned} \mathbf{AMB} &= \mathbf{A}^{k+1}\mathbf{M}\mathbf{B}^{k+1} + \sum_{i=0}^{k-1} \mathbf{A}^{i+1}\left(\mathbf{M} - \mathbf{AMB}\right)\mathbf{B}^{i+1} \\ &= \mathbf{A}^{k+1}\mathbf{M}\mathbf{B}^{k+1} + \sum_{i=0}^{k} \mathbf{A}^{i}\left(\mathbf{M} - \mathbf{AMB}\right)\mathbf{B}^{i} - \left(\mathbf{M} - \mathbf{AMB}\right) \end{aligned}$$

Canceling $\mathbf{AMB}$ from both sides yields the identity for $k+1$. □

Theorem 1.6 (Properties of Displacement Operators, [1])**.**

$$\nabla_{\mathbf{A},\mathbf{B}}[\mathbf{M}^{-1}] = -\mathbf{M}^{-1}\nabla_{\mathbf{A},\mathbf{B}}[\mathbf{M}]\mathbf{M}^{-1} \tag{2}$$

$$\nabla_{\mathbf{A},\mathbf{C}}[\mathbf{MN}] = \nabla_{\mathbf{A},\mathbf{B}}[\mathbf{M}]\mathbf{N} + \mathbf{M}\nabla_{\mathbf{B},\mathbf{C}}[\mathbf{N}] \tag{3}$$

Lemma 1.7 ([2], Theorem 3.1)**.** *If $\mathbf{A}$ is non-singular, $\nabla_{\mathbf{A},\mathbf{B}} = \mathbf{A}\triangle_{\mathbf{A}^{-1},\mathbf{B}}$. If $\mathbf{B}$ is non-singular,* $\nabla_{\mathbf{A},\mathbf{B}} = -\triangle_{\mathbf{A},\mathbf{B}^{-1}}\mathbf{B}$

Proof. The statement of the theorem follows from the following simple observations: If $\mathbf{A}$ is invertible, we have $\mathbf{AM} - \mathbf{MB} = \mathbf{A}(\mathbf{M} - \mathbf{A}^{-1}\mathbf{MB})$, and if $\mathbf{B}$ is invertible, we have $\mathbf{AM} - \mathbf{MB} = -(\mathbf{M} - \mathbf{AMB}^{-1})\mathbf{B}$. □

2 Krylov Decomposition and Circulant-SkewCirculant decomposition for Toeplitz-like Matrices

Proof of Theorem 2.2

Proof. The statement of the proof follows from Theorem 1.5 setting $k = n$, using $\mathbf{A}^n = a\mathbf{I}, \mathbf{B}^n = b\mathbf{I}$; inserting $\triangle_{\mathbf{A},\mathbf{B}}[\mathbf{M}] = \mathbf{G}\mathbf{H}^T$ in the sum in the second term of Eqn. 1.

$$\begin{aligned} \mathbf{M} &= \mathbf{A}^n\mathbf{M}\mathbf{B}^n + \sum_{i=0}^{n-1} \mathbf{A}^i\triangle_{\mathbf{A},\mathbf{B}}[\mathbf{M}]\mathbf{B}^i \qquad (4) \\ &= ab\,\mathbf{M} + \sum_{i=0}^{n-1} \mathbf{A}^i\mathbf{G}\mathbf{H}^T\mathbf{B}^i \\ &= ab\,\mathbf{M} + \sum_{i=1}^{r} [\mathbf{g}_i\ \mathbf{A}\mathbf{g}_i\ \mathbf{A}^2\mathbf{g}_i \dots \mathbf{A}^{n-1}\mathbf{g}_i][\mathbf{h}_i\ \mathbf{B}^T\mathbf{h}_i\ (\mathbf{B}^T)^2\mathbf{h}_i \dots (\mathbf{B}^T)^{n-1}\mathbf{h}_i]^T \end{aligned}$$

and observing that the resulting expressions can be rewritten in terms of Krylov matrices generated by $\mathbf{A}, \mathbf{B}^T$ applied to columns of $\mathbf{G}, \mathbf{H}$. □

We need the following simple identity in preparation for the Proof of Theorem 2.4.

Lemma 2.1. $\mathbf{Z}_1(\mathbf{J}\mathbf{Z}_1^T\mathbf{g})^T = \mathbf{Z}_1(\mathbf{g})$

Proof. We show that $\mathbf{Z}_1(\mathbf{J}\mathbf{Z}_1^T\mathbf{g}) = \mathbf{Z}_1(\mathbf{g})^T$. Explicitly, by taking downshifts of $\mathbf{g}$ and stacking them as rows, we have,

$$Z_1(\mathbf{g})^T = \begin{bmatrix} g_0 & g_1 & \cdots & g_{n-1} \\ g_{n-1} & g_0 & \cdots & g_{n-2} \\ \vdots & \vdots & \vdots & g_1 \\ g_1 & \cdots & g_{n-1} & g_0 \end{bmatrix}$$

At the same time, observe that,

$$\mathbf{J}\mathbf{Z}_1^T\mathbf{g} = \begin{pmatrix} g_0 \\ g_{n-1} \\ \vdots \\ g_1 \end{pmatrix}$$

which is the first column of $Z_1(\mathbf{g})^T$. Since the rest of the columns follow by taking downward shifts, the identity follows. $\square$

Proof of Theorem 2.4

Proof. By Lemma 1.7, it follows that if $\nabla_{\mathbf{Z}_1,\mathbf{Z}_{-1}}[\mathbf{M}] = \mathbf{G}\mathbf{H}^T$, then $\triangle_{\mathbf{Z}_1^T,\mathbf{Z}_{-1}} = \left(\mathbf{Z}_1^T\mathbf{G}\right)\mathbf{H}^T$. Plugging $\mathbf{A} = \mathbf{Z}_1^T, \mathbf{B} = \mathbf{Z}_{-1}, a = 1, b = -1$ in Theorem 2.2 in the main paper, we get,

$$\begin{aligned}
\mathbf{M} &= \frac{1}{2}\sum_{i=0}^{r-1} krylov(\mathbf{Z}_1^T, \mathbf{Z}_1^T\mathbf{g_i})krylov(\mathbf{Z}_{-1}^T, \mathbf{h_i})^T \\
&= \frac{1}{2}\sum_{i=0}^{r-1} \mathbf{J}\mathbf{Z}_1(\mathbf{J}\mathbf{Z}_1^T\mathbf{g}_i)\left[\mathbf{J}\mathbf{Z}_1(\mathbf{J}\mathbf{h}_i)\right]^T && (5)\\
&= \frac{1}{2}\sum_{i=0}^{r-1} \mathbf{J}\mathbf{Z}_1(\mathbf{J}\mathbf{Z}_1^T\mathbf{g}_i)\mathbf{Z}_1(\mathbf{J}\mathbf{h}_i)^T\mathbf{J} && (6)\\
&= \frac{1}{2}\sum_{i=0}^{r-1} \left(\mathbf{J}\mathbf{Z}_1(\mathbf{J}\mathbf{Z}_1^T\mathbf{g}_i)\mathbf{J}\right)\left(\mathbf{J}\mathbf{Z}_1(\mathbf{J}\mathbf{h}_i)^T\mathbf{J}\right) && (7)\\
&= \frac{1}{2}\sum_{i=0}^{r-1} \mathbf{Z}_1(\mathbf{J}\mathbf{Z}_1^T\mathbf{g}_i)^T\mathbf{Z}_{-1}(\mathbf{J}\mathbf{h}_i) && (8)\\
&= \frac{1}{2}\sum_{i=0}^{r-1} \mathbf{Z}_1(\mathbf{g}_i)\mathbf{Z}_{-1}(\mathbf{J}\mathbf{h}_i) && (9)
\end{aligned}$$

$\square$

Above, we use the following facts (1) $\mathbf{J}^2 = \mathbf{I}$, (2) Property 5 in Proposition 1.1 to deduce that $krylov(\mathbf{Z}_1^T, \mathbf{v}) = krylov(\mathbf{J}\mathbf{Z}_1\mathbf{J}, \mathbf{v}) = \mathbf{J}\mathbf{Z}_1(\mathbf{J}\mathbf{v})$ and (3) Lemma 2.1.

Lemma 2.2. *For any scalars $e \neq 0, f \neq 0$, $rank(\nabla_{\mathbf{Z}_e,\mathbf{Z}_f}[\mathbf{M}]) \leq r$ if and only if $rank(\nabla_{\mathbf{Z}_1,\mathbf{Z}_{-1}}[\mathbf{M}]) \leq r$.*

Proof. Observe that $\mathbf{Z}_e = diag([e, 1_{n-1}]^T)\mathbf{Z}_1 = \mathbf{Z}_1 diag([e, 1_{n-1}]^T)$ i.e. the scaling action is delegating to a diagonal matrix, via pre- or post-multiplication. Likewise, $\mathbf{Z}_f = diag([-f, 1_{n-1},])\mathbf{Z}_{-1} = \mathbf{Z}_{-1}diag([-f, 1_{n-1},])$. Hence,

$$\begin{aligned}
\nabla_{\mathbf{Z}_e,\mathbf{Z}_f}[\mathbf{M}] &= \mathbf{Z}_e\mathbf{M} - \mathbf{M}\mathbf{Z}_f && (10)\\
&= diag([e, 1_{n-1}])\mathbf{Z}_1\mathbf{M} - \mathbf{M}\mathbf{Z}_{-1}diag([-f, 1_{n-1}]) \\
&= \mathbf{G}\mathbf{H}^T
\end{aligned}$$

It follows that $\mathbf{Z}_1\mathbf{M} - \mathbf{M}\mathbf{Z}_{-1} = \bar{\mathbf{G}}\bar{\mathbf{H}}^T$, where $\bar{\mathbf{G}} = diag([e^{-1}, 1_{n-1}])\mathbf{G}, \bar{\mathbf{H}} = \mathbf{H}diag([-f^{-1}, 1_{n-1}])$. The converse can be shown similarly. □

3 Learning Toeplitz-like Matrices: Proofs

Proof of Theorem 3.1

Proof. Toeplitz-like structured matrices have the form:

$$\mathbf{M}(\mathbf{G}, \mathbf{H}) = \sum_{i=1}^{r} \mathbf{Z}_1(\mathbf{g}_i)\mathbf{Z}_{-1}(\mathbf{h}_i) \tag{11}$$

1. For $r = 1$, when $\mathbf{H} = [\mathbf{e}_0]$, we have $\mathbf{Z}_{-1}(\mathbf{h}_0) = \mathbf{I}$. Hence, the sum in Eqn. 11 reduces to a general Circulant term. Likewise, when $\mathbf{G} = [\mathbf{e}_0]$, we have $\mathbf{Z}_1(\mathbf{g_0}) = \mathbf{I}$ Eqn. 11 reduces to a general skew-circulant.

2. The result follows from Theorem 2.4 in the main paper for Toeplitz matrices by re-defining $\mathbf{h} \equiv \mathbf{J}\mathbf{h}$.

3. For any Toeplitz matrix $\mathbf{T}$, using Theorem 1.6, Eqn. 2, we have

$$\begin{aligned} \nabla_{\mathbf{Z}_1,\mathbf{Z}_{-1}}[\mathbf{T}^{-1}] &= -\mathbf{T}^{-1}\nabla_{\mathbf{Z}_1,\mathbf{Z}_{-1}}[\mathbf{T}]\mathbf{T}^{-1} \\ &= -\left(\mathbf{T}^{-1}\mathbf{G}\right)\left(\mathbf{H}^T\mathbf{T}^{-1}\right) \end{aligned} \tag{12}$$

   where $\mathbf{G}, \mathbf{H}$ are factors with rank upto 2. The last expression shows that $\nabla_{\mathbf{Z}_1,\mathbf{Z}_{-1}}[\mathbf{T}^{-1}]$ also has rank upto 2. We can now use Theorem 2.4.

4. The proof follows by induction. For $t = 1$, i.e. the result is true by Theorem 2.4 for a single Toeplitz matrix and the previous assertion concerning inverses of Toeplitz matrices. Assume it is true for t.

   Now in Theorem 1.6 Eqn 3, let $\mathbf{M} = \mathbf{A}_1 \ldots \mathbf{A}_t$ and let $\mathbf{N} = \mathbf{A}_{t+1}$ and set the operator matrices to be $\mathbf{A} = \mathbf{Z}_1, \mathbf{C} = \mathbf{Z}_{-1}$ and $\mathbf{B} = \mathbf{Z}_e$ for some scalar $e \neq 1$ or -1. Then we have,

$$\nabla_{\mathbf{Z_1},\mathbf{Z}_{-1}}[\mathbf{A}_1 \ldots \mathbf{A}_{t+1}] \quad = \quad \nabla_{\mathbf{Z}_1,\mathbf{Z}_e}[\mathbf{A}_1 \ldots \mathbf{A}_t]\mathbf{A_{t+1}} + \mathbf{A}_1 \ldots \mathbf{A}_t \nabla_{\mathbf{Z}_e,\mathbf{Z}_{-1}}[\mathbf{A}_{t+1}]$$

   In the first term above, $\nabla_{\mathbf{Z}_1,\mathbf{Z}_e}[\mathbf{A}_1 \ldots \mathbf{A}_t]$ has rank at most $2t$ if and only if $\nabla_{\mathbf{Z}_1,\mathbf{Z}_{-1}}[\mathbf{A}_1 \ldots \mathbf{A}_t]$ has rank at most $2t$ by Lemma 2.2; and the latter is true by the by the inductive assumption. In the second term $\nabla_{\mathbf{Z}_1,\mathbf{Z}_{-1}}[\mathbf{A}_{t+1}]$ has rank at most 2 by Theorem 2.4. Hence, the new displacement rank is no more than $2t + 2$. The completes the inductive argument.

5. We use the fact that for any linear displacement operator L, $L[\sum_{i=1}^{p} \alpha_i[\mathbf{M}_i]] = \sum_{i=1}^{p} \alpha_i L[\mathbf{M}_i]$. If each term in the sum has rank at most $2t$, then the sum has rank atmost $2tp$.

6. Follows from Corollary 1.4 and the fact that for any $n \times n$ matrix $\mathbf{M}$, $rank(\nabla_{\mathbf{Z_1},\mathbf{Z}_{-1}}[\mathbf{M}])$ is atmost n.

□

Proof of Proposition 3.4

Proof. The Jacobian of a vector valued function $f : \mathbb{R}^m \mapsto \mathbb{R}^n$ is the $n \times m$ matrix

$$[Jf]_{ij} = \frac{\partial f_i}{\partial x_j},$$

where $f(\mathbf{x}) = [f_1(\mathbf{x}), \ldots f_n(\mathbf{x})]^T$. So consider the vector-valued function $f(\mathbf{v}) = \mathbf{Z}_f(\mathbf{v})\mathbf{x}$ for a fixed $\mathbf{x}$. By the diagonalization of f-Circulant matrices, Theorem 3.2 in the main paper,

$$f(\mathbf{v}) = D_{\mathbf{f}}^{-1}\Omega_n^{-1}\,\mathrm{diag}(\Omega_n(\mathbf{f}\circ\mathbf{v}))\Omega_n(\mathbf{f}\circ\mathbf{x})$$

Define $\mathbf{U}_f = \Omega_n D_{\mathbf{f}}$ and $\mathbf{y} = \mathbf{U}_f\mathbf{v}$. Then $f(\mathbf{v}) = h(g(\mathbf{v}))$ where $h(\mathbf{v}) = \mathbf{U}_f\mathbf{v}$ and $g(\mathbf{v}) = diag(\mathbf{U}_f\mathbf{v})\mathbf{y}$. Note that $\frac{\partial g_i}{\partial v_j} = y_i[\mathbf{U}_f]_{ij}$. The Jacobian of h is simply $\mathbf{U}_f^{-1}$ while Jacobian of g with respect to $\mathbf{v}$ is simply $diag(\mathbf{y})\mathbf{U}_f$. From the chain rule it follows that the Jacobian of f is $\mathbf{U}_f^{-1}diag(\mathbf{U}_f\mathbf{x})\mathbf{U}_f = \mathbf{Z}_f(\mathbf{x})$. □

Proof of Proposition 3.5

Proof. The transform under consideration is,

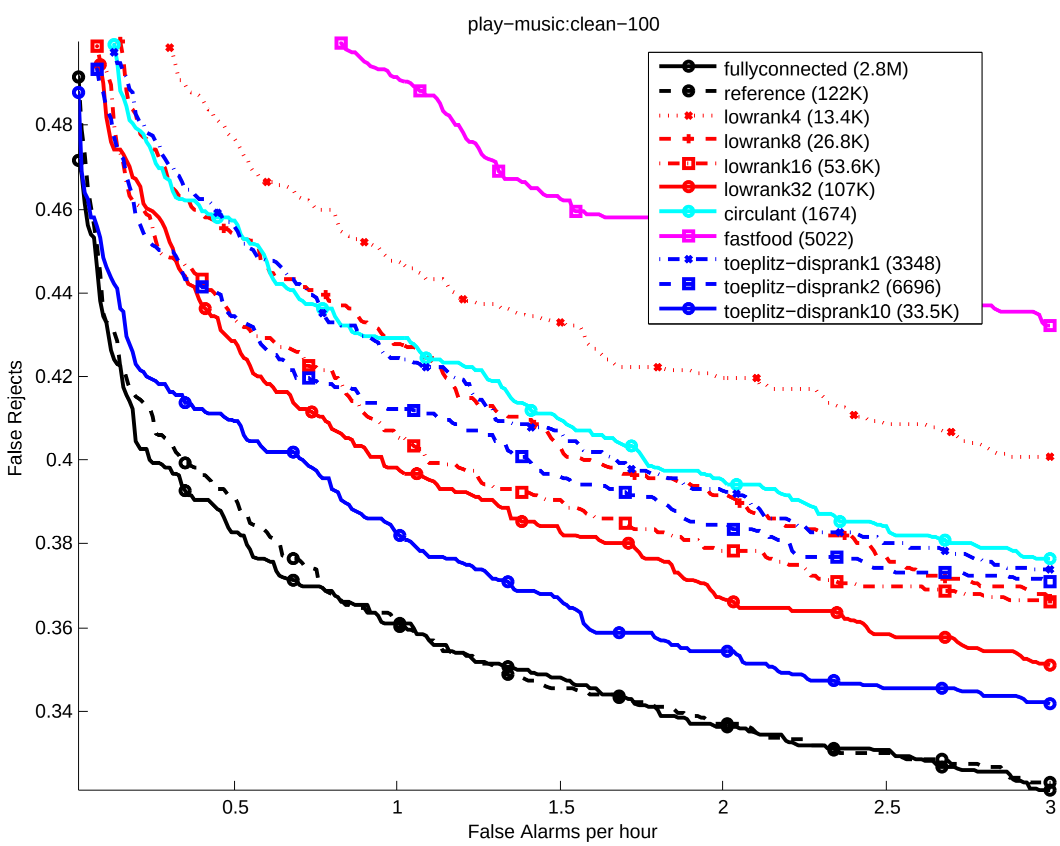
$$f(\mathbf{x}, \mathbf{G}, \mathbf{H}) = \sum_{i=1}^{r} \mathbf{Z}_1(\mathbf{g}_i)\mathbf{Z}_{-1}(\mathbf{h}_i)\mathbf{x} \tag{13}$$

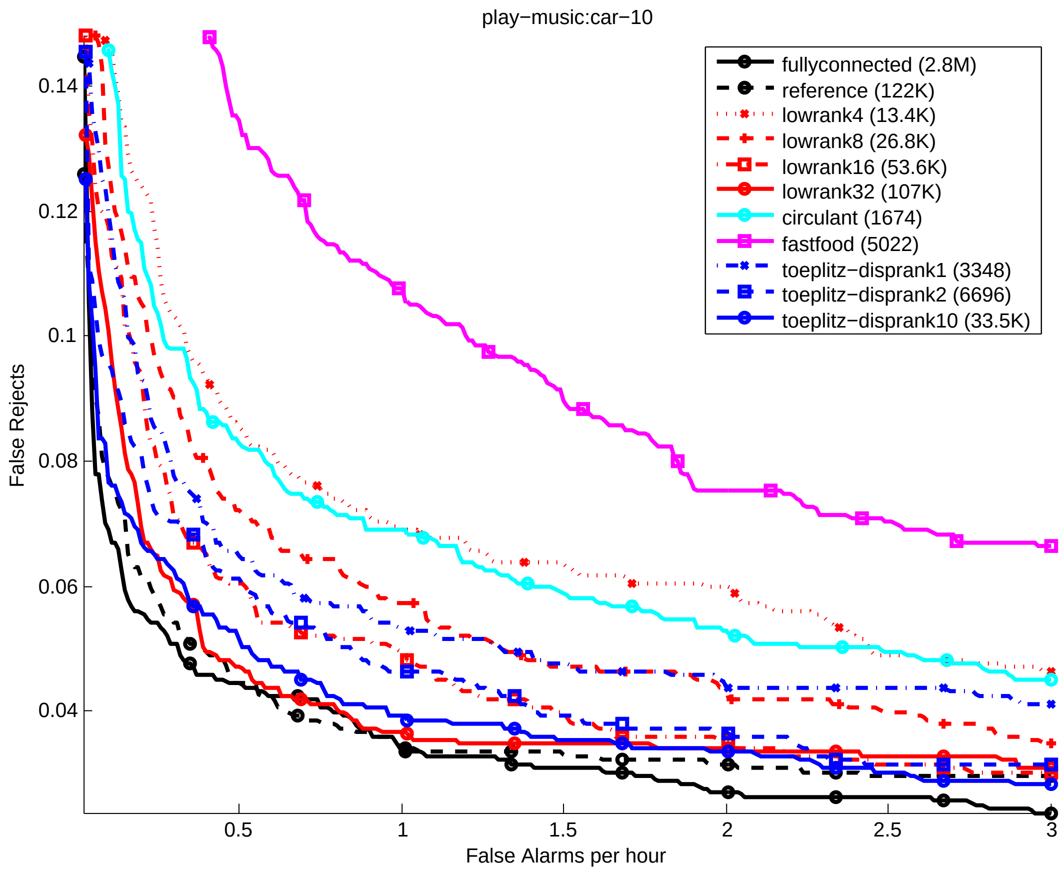
The Jacobian with respect to $\mathbf{g}_i$ is simply the Jacobian of the transform $\mathbf{Z}_1(\mathbf{g}_i)\mathbf{y}$ where $\mathbf{y} = \mathbf{Z}_{-1}(\mathbf{h}_i)\mathbf{x}$. Hence, we can apply Proposition 3.4 for $f = -1$ to get that

$$J_{\mathbf{g}_j} f|_{\mathbf{x}} = \mathbf{Z}_1\left(\mathbf{Z}_{-1}(\mathbf{h}_j)\mathbf{x}\right)$$

Similarly, the Jacobian with respect to $\mathbf{h}_j$ follows immediately from the chain rule and Proposition 3.5 for $f = -1$. □

4 Additional Empirical Results

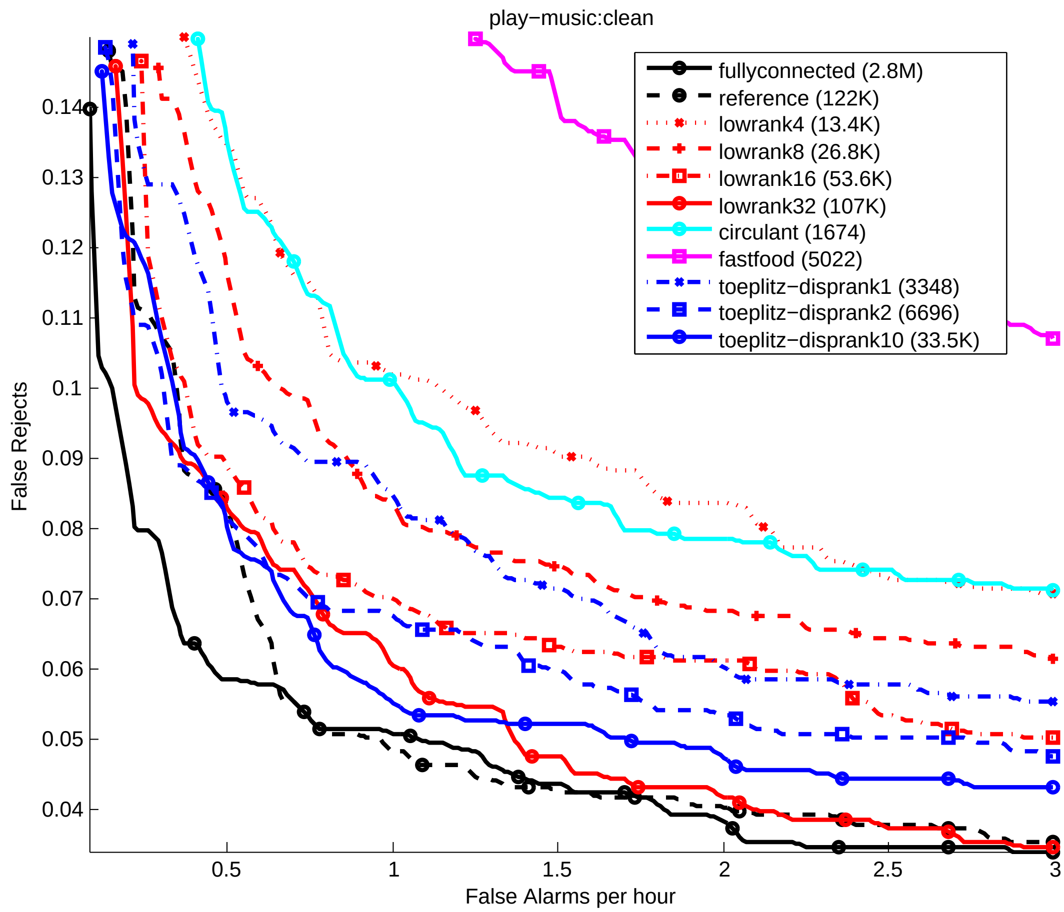




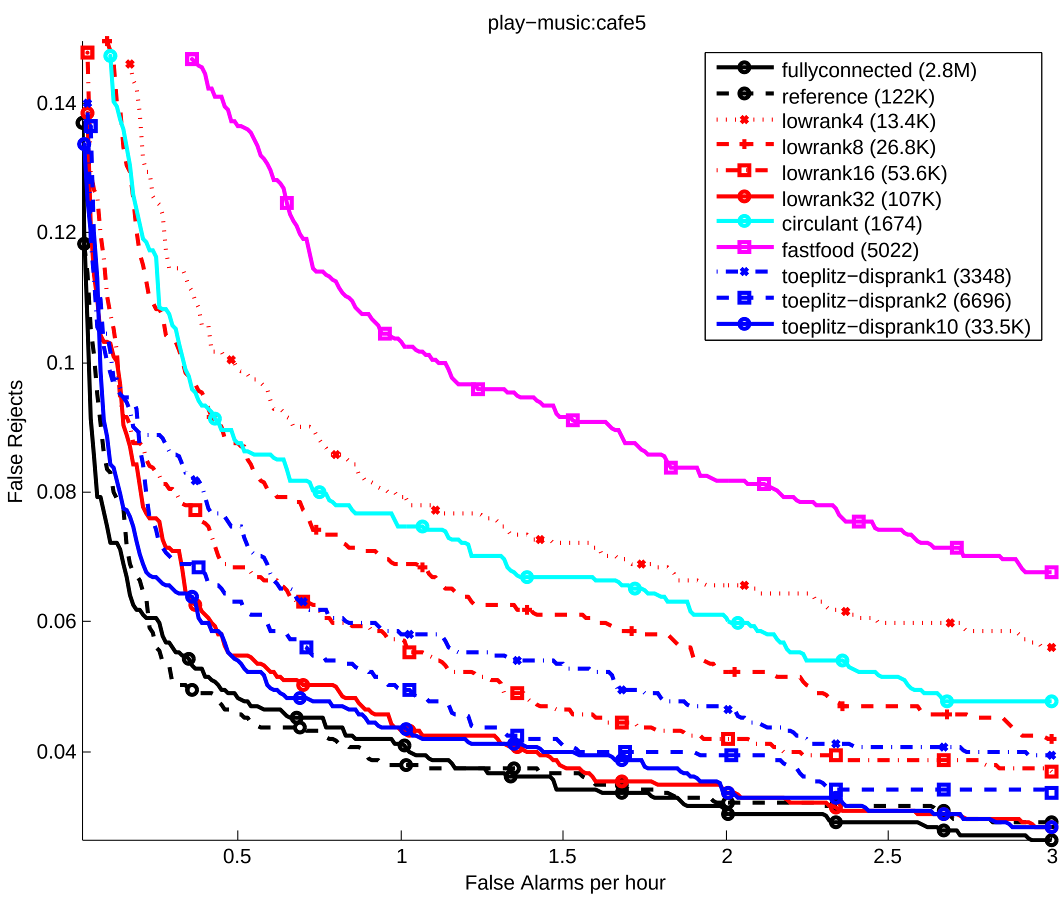
References

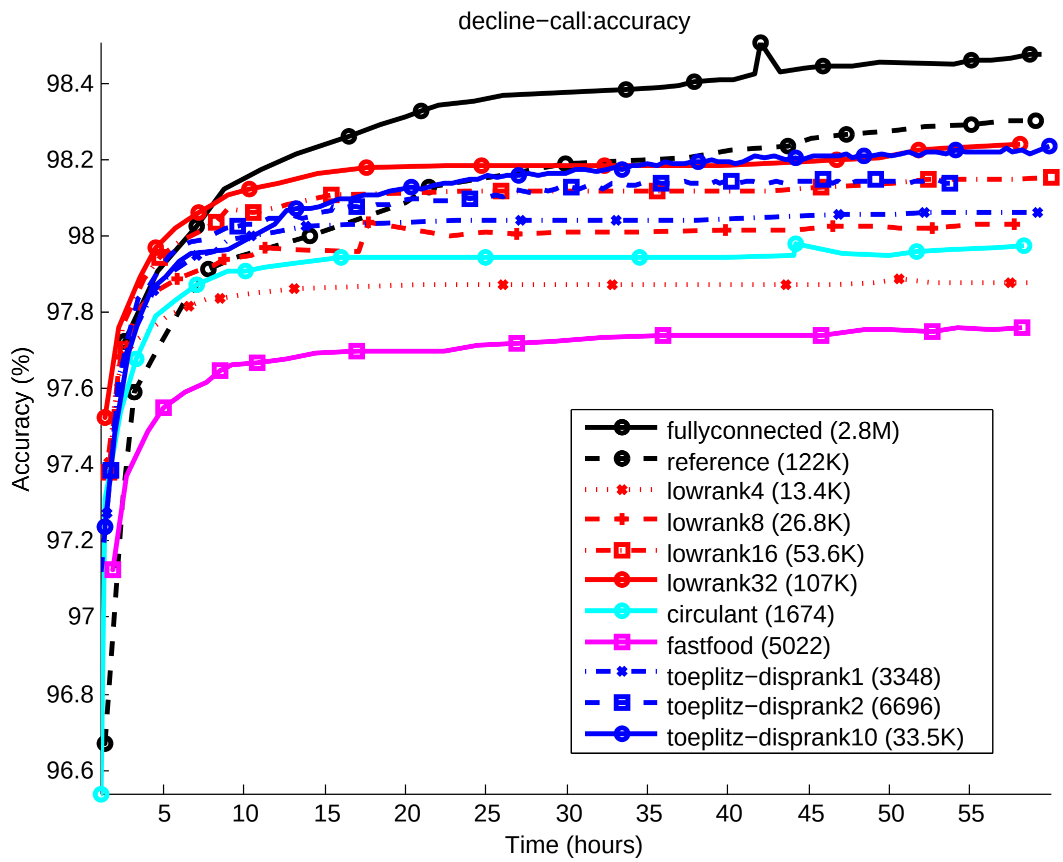
[1] V. Pan. *Structured Matrices and Polynomials: Unified Superfast Algorithms*. Springer, 2001.

[2] V. Pan. Inversion of displacement operators. *SIAM Journal of Matrix Analysis and Applications*, pages 660–677, 2003.



play−music:clean
fullyconnected (2.8M)
reference (122K)
lowrank4 (13.4K)
lowrank8 (26.8K)
lowrank16 (53.6K)
lowrank32 (107K)
circulant (1674)
fastfood (5022)
toeplitz−disprank1 (3348)
toeplitz−disprank2 (6696)
toeplitz−disprank10 (33.5K)
False Rejects
False Alarms per hour
0.14
0.13
0.12
0.11
0.1
0.09
0.08
0.07
0.06
0.05
0.04
0.5
1
1.5
2
2.5
3


play−music:cafe5
fullyconnected (2.8M)
reference (122K)
lowrank4 (13.4K)
lowrank8 (26.8K)
lowrank16 (53.6K)
lowrank32 (107K)
circulant (1674)
fastfood (5022)
toeplitz−disprank1 (3348)
toeplitz−disprank2 (6696)
toeplitz−disprank10 (33.5K)
False Rejects
0.04
0.06
0.08
0.1
0.12
0.14
False Alarms per hour
0.5
1
1.5
2
2.5
3


decline−call:accuracy
Accuracy (%)
Time (hours)
fullyconnected (2.8M)
reference (122K)
lowrank4 (13.4K)
lowrank8 (26.8K)
lowrank16 (53.6K)
lowrank32 (107K)
circulant (1674)
fastfood (5022)
toeplitz−disprank1 (3348)
toeplitz−disprank2 (6696)
toeplitz−disprank10 (33.5K)