

# An Inversion Theorem for Buffered Linear Toeplitz (BLT) Matrices and Applications to Streaming Differential Privacy

H. Brendan McMahan<sup>1</sup>Krishna Pillutla<sup>2</sup><sup>1</sup>Google Research<sup>2</sup>Wadhvani School of Data Science & AI, IIT Madras

## Abstract

Buffered Linear Toeplitz (BLT) matrices are a family of parameterized lower-triangular matrices that play an important role in streaming differential privacy with correlated noise. Our main result is a BLT inversion theorem: the inverse of a BLT matrix is itself a BLT matrix with different parameters. We also present an efficient and differentiable  $O(d^3)$  algorithm to compute the parameters of the inverse BLT matrix, where  $d$  is the degree of the original BLT (typically  $d < 10$ ). Our characterization enables direct optimization of BLT parameters for privacy mechanisms through automatic differentiation.

## 1 Introduction

We consider the inverses of a family of parameterized lower-triangular matrices known as Buffered Linear Toeplitz (BLT) matrices [1]. Given a scale parameter  $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{R}^d$  and a decay parameter  $\lambda = (\lambda_1, \dots, \lambda_d) \in \mathbb{R}^d$ , the  $n \times n$  BLT matrix is defined as

$$\text{BLT}_n(\alpha, \lambda) := \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots \\ \sum_{i=1}^d \alpha_i & 1 & 0 & 0 & \cdots \\ \sum_{i=1}^d \alpha_i \lambda_i & \sum_{i=1}^d \alpha_i & 1 & 0 & \cdots \\ \sum_{i=1}^d \alpha_i \lambda_i^2 & \sum_{i=1}^d \alpha_i \lambda_i & \sum_{i=1}^d \alpha_i & 1 & \ddots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}_{n \times n}. \quad (1)$$

The matrix  $\mathbf{C} = \text{BLT}_n(\alpha, \lambda)$  is lower triangular and Toeplitz (i.e., it has equal entries along each diagonal from the top-left to the bottom-right), with ones along the principal diagonal:<sup>1</sup>

$$\mathbf{C}[j, k] = \begin{cases} 0, & \text{if } j < k, \\ 1, & \text{if } j = k, \\ \sum_{i=1}^d \alpha_i \lambda_i^{j-k-1}, & \text{if } j > k. \end{cases} \quad (2)$$

Such parameterized matrices (and their inverses) are central to streaming differential privacy with correlated noise, achieving near-optimal tradeoffs between privacy, utility, and computation cost; we describe this in detail in Section 2.

<sup>1</sup>We denote the  $(j, k)$ <sup>th</sup> entry of the matrix  $\mathbf{C} \in \mathbb{R}^{n \times n}$  by  $\mathbf{C}[j, k]$ ; Table 1 in Appendix A provides a complete notation summary.

The main result of this note is that the BLT family of matrices is closed under inversion. In particular, the inverse of a degree- $d$  BLT matrix  $\text{BLT}_n(\boldsymbol{\alpha}, \boldsymbol{\lambda})$  (which always exists) is also a BLT matrix of the same order  $d$ :

$$\text{BLT}_n(\boldsymbol{\alpha}, \boldsymbol{\lambda})^{-1} = \text{BLT}_n(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\lambda}}) \quad \text{for all } n > 0$$

for all integers  $n > 0$  for *unique* parameters  $\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\lambda}} \in \mathbb{R}^d$ . We also give an equivalence between representing a (BLT, inverse BLT) system  $\text{BLT}_n(\boldsymbol{\alpha}, \boldsymbol{\lambda})^{-1} = \text{BLT}_n(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\lambda}})$  using (a) both the parameters  $\boldsymbol{\alpha}, \boldsymbol{\lambda}$  of the first BLT, and (b) both the decay parameters  $\boldsymbol{\lambda}, \hat{\boldsymbol{\lambda}}$ . Finally, we give a differentiable algorithm to compute the BLT inverse in  $O(d^3)$  time for any size  $n$ .

Next, we provide some background on streaming differential privacy in Section 2. BLT matrices and their inverses play an important role in this setting. We give a full formal statement of our BLT inversion theorem in Section 3. We give the key ideas behind the proofs in Section 4 with full proof details in Section 5.

## 2 Background

Let  $\mathbf{G} \in \mathbb{R}^{n \times m}$  be a sequence of  $n$  vectors in  $\mathbb{R}^m$  stacked row-wise into a matrix. Each vector  $\mathbf{g}_t$  (i.e.  $t^{\text{th}}$  row of  $\mathbf{G}$ ) is assumed to satisfy  $\|\mathbf{g}_t\|_2 \leq \zeta$  for some constant  $\zeta > 0$ . We aim to estimate (in a differentially private manner) a sequence of (known) linear combinations of these vectors, represented as the rows of  $\mathbf{A}\mathbf{G} \in \mathbb{R}^{n \times m}$ ; here,  $\mathbf{A} \in \mathbb{R}^{n \times n}$  is known as the *workload matrix*.

This setup captures diverse problems such as continual counting and stochastic optimization under differential privacy. In the latter case,  $\mathbf{g}_t$  is an unbiased estimator of the loss gradient evaluated at the current model parameters  $\boldsymbol{\theta}_t$ . The workload matrix captures the optimization algorithm: stochastic gradient descent (SGD) with a constant learning rate  $\eta$  corresponds to the prefix sum workload  $\mathbf{A}_{\text{pre}}$ , which is the lower triangular matrix with all ones. This is because each iterate  $\boldsymbol{\theta}_t = \boldsymbol{\theta}_0 - \eta \sum_{\tau < t} \mathbf{g}_\tau$  of SGD relies on estimating the prefix sums  $\sum_{\tau < t} \mathbf{g}_\tau$ , which are the rows of  $\mathbf{A}_{\text{pre}}\mathbf{G}$ . Other first-order optimizers such as SGD with momentum correspond to different workloads.

The matrix mechanism for differential privacy [2, 3], known also as DP-FTRL [4, 5] in the learning setting, injects correlated noise to release private estimates of  $\mathbf{A}\mathbf{G}$ . Given a factorization  $\mathbf{A} = \mathbf{B}\mathbf{C}$  with  $\mathbf{C}$  invertible, this correlated noise mechanism is defined as

$$\mathcal{M}(\mathbf{G}) = \mathbf{B}(\mathbf{C}\mathbf{G} + \mathbf{Z}) = \mathbf{A}(\mathbf{G} + \mathbf{C}^{-1}\mathbf{Z}), \quad (3)$$

where  $\mathbf{Z} \in \mathbb{R}^{n \times m}$  is component-wise i.i.d. Gaussian noise. We scale<sup>2</sup>  $\mathbf{Z} \sim \mathcal{N}_{n \times m}(0, \text{sens}(\mathbf{C})^2 \sigma^2)$ , where  $\text{sens}(\mathbf{C})$  is the  $\ell_2$ -sensitivity of the operation  $\mathbf{G} \mapsto \mathbf{C}\mathbf{G}$ , while  $\sigma$  is a noise multiplier depending only on the desired privacy level; e.g. we take  $\sigma^2 = 1/(2\rho)$  for a  $\rho$ -zero-concentrated DP guarantee [6]. Note that multiplication by  $\mathbf{B}$  is simply a post-processing step that does not affect the privacy guarantee. The sensitivity  $\text{sens}(\mathbf{C})$  depends on how adjacent  $\mathbf{G}, \mathbf{G}'$  are allowed to differ. In the learning setting, if a data item can appear only once in training, then  $\text{sens}(\mathbf{C}) = \|\mathbf{C}\|_{\text{col}}$  is the maximum column norm of the matrix  $\mathbf{C}$ . Different expressions exist when each data item can participate more than once [see e.g. 7, Eq. (2)].

In general, we aim to find the factorization  $\mathbf{A} = \mathbf{B}\mathbf{C}$  to minimize the worst-case expected (squared)  $\ell_2$  norm across all rows of  $\mathcal{M}(\mathbf{G}) - \mathbf{A}\mathbf{G} = \mathbf{B}\mathbf{Z} = \mathbf{A}^{-1}\mathbf{C}\mathbf{Z}$ . This can be evaluated

<sup>2</sup>The notation  $\mathbf{Z} \sim \mathcal{N}_{n \times m}(0, \sigma^2)$  denotes a random matrix  $\mathbf{Z} \in \mathbb{R}^{n \times m}$  whose entries are i.i.d.  $\mathcal{N}(0, \sigma^2)$ .

(assuming the norm constant  $\zeta = 1$  w.l.o.g.) as the (square of the) *max loss*

$$L(\mathbf{C}) := \text{sens}(\mathbf{C}) \cdot \|\mathbf{A}^{-1}\mathbf{C}\|_{\text{row}}, \quad (4)$$

where  $\|\mathbf{B}\|_{\text{row}}$  denotes the maximum row norm of the matrix  $\mathbf{B}$ . Sometimes, we may choose the Frobenius norm  $\|\cdot\|_{\text{F}}$  instead of  $\|\cdot\|_{\text{row}}$  to compute the average expected (squared)  $\ell_2$  norm across rows of  $\mathbf{A}\mathbf{G}$  instead of the worst-case.

**BLT Mechanism** A major focus of prior research has been to improve the privacy-utility-compute tradeoffs of the mechanism (3) in theory and practice; see [7, 8, 9, 10] and the references therein. In particular, the computation cost of computing  $(\mathbf{C}^{-1}\mathbf{Z})[t, :]$  in each iteration  $t$  dominates the running time of the algorithm in the learning setting. The *BLT mechanism* [1] achieves state-of-the-art tradeoffs. In general, it restricts the  $\mathbf{C}$  matrix to be Toeplitz and parameterizes its first column  $c_1, c_2, \dots$  as  $c_t = \mathbf{u}^\top \mathbf{W}^{t-1} \mathbf{v}$  using a matrix  $\mathbf{W}$  and two vectors  $\mathbf{u}, \mathbf{v}$ . We focus on the *diagonal BLT* formulation described in Eq. (1), which corresponds to diagonal  $\mathbf{W}$ ; this formulation has been preferred in empirical studies for being more computationally efficient without sacrificing utility [1, 11].

A key advantage of the BLT mechanism is that the rows of the correlated noise  $\mathbf{C}^{-1}\mathbf{Z}$  in Eq. (3) can be generated in a streaming fashion with  $O(dm)$  time and space complexity [cf. 11, Alg. 2,3]; notably, this is independent of the iteration counter. Together with additive utility guarantee in the streaming setting where  $\text{sens}(\mathbf{C}) = \|\mathbf{C}\|_{\text{col}}$ , this leads to near-optimal privacy-utility-compute tradeoffs with the prefix sum workload  $\mathbf{A} = \mathbf{A}_{\text{pre}}$ . In particular, for any size  $n > 0$  and error term  $\delta > 0$ , there exist some parameters  $\boldsymbol{\alpha}, \boldsymbol{\lambda} \in \mathbb{R}^d$  for  $d = O(\log^2(n/\delta))$  that give an *additive* approximation of the optimal max error:

$$L(\text{BLT}_n(\boldsymbol{\alpha}, \boldsymbol{\lambda})) \leq \min\{L(\mathbf{C}) : \mathbf{C} \in \mathbb{R}^{n \times n} \text{ is lower-triangular \& Toeplitz}\} + \delta. \quad (5)$$

In this work, we show that the inverse of a diagonal BLT of the form of Eq. (1) is another diagonal BLT; we give a precise statement in Section 3. We also describe how to find the parameters  $\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\alpha}}$  of the inverse BLT in a differentiable manner so that max loss (4) can be optimized (as a function of the BLT parameters  $\boldsymbol{\alpha}, \boldsymbol{\lambda}$ ) using automatic differentiation.

**Parameter Restrictions** We restrict ourselves to the BLT decay parameter  $\boldsymbol{\lambda} \in (0, 1)^d$ , and our main result (Theorem 1) further focuses on the case where each  $\alpha_i > 0$  and  $\sum_{i=1}^d \alpha_i < 1$ . These restrictions are not strictly necessary, in that Eq. (1) is a well-defined (and invertible) matrix for any parameters  $\boldsymbol{\alpha}, \boldsymbol{\lambda} \in \mathbb{R}^d$ . Why these restrictions? In short, we believe they identify the most practically important subclass of BLTs where the goal is to approximate the optimal Toeplitz matrix (see the right side of Eq. (5)). This allows sharper and simpler theoretical characterizations and numerically stable mechanisms. The restriction  $\boldsymbol{\alpha} > 0$  (or  $\boldsymbol{\alpha} < 0$ ) is beneficial when optimizing BLTs, and ensure our subclass is closed under matrix inversion. For example, the previous works [1, 11] restrict the search over  $\boldsymbol{\alpha}$  to over strictly positive entries by imposing log-barrier functions in the optimization.

Do these restrictions lead to sub-optimal mechanisms? For the problem of correlated noise DP mechanisms for single-participation, strong empirical evidence from prior work [1, 11] shows that this is not the case. In particular, Dvijotham et al. [1] showed BLTs that satisfy these restrictions can for all practical purposes perfectly match the optimal Toeplitz matrix (i.e. the Toeplitz matrix  $\mathbf{C}$  minimizing  $L(\mathbf{C})$  optimizing, as in the right side of Eq. (5)), and McMahan et al. [11] showed

strong performance for a common multiple-participation setting. Nevertheless, it is possible that for some applications the additional expressive power of allowing some  $\lambda_i < 0$  could make it worth investigating this case further.

**Notation Summary** We use the shorthand  $[d] := \{1, \dots, d\}$ . Vectors are denoted by boldfaced lower-case (Greek or Latin) letters (e.g.  $\boldsymbol{\lambda}$ ,  $\boldsymbol{u}$ ) while matrices are denoted by boldfaced upper-case letters (e.g.  $\boldsymbol{C}$  or  $\boldsymbol{M}$ ); both are 1-indexed. Often, we will denote the first columns of the lower triangular and Toeplitz matrices  $\boldsymbol{C}$  and  $\boldsymbol{C}^{-1}$  with shorthand  $c_t = \boldsymbol{C}[t, 1]$  and  $\hat{c}_t = (\boldsymbol{C}^{-1})[t, 1]$  respectively. We give a detailed summary of the notation in Table 1 of Appendix A.

### 3 Main Results

Our main result is a BLT inversion theorem: the inverse of a BLT matrix is also a unique BLT. We give some properties of the inverse BLT parameters. All proofs are given in Sections 4 and 5. We say a vector of parameters  $\boldsymbol{\lambda}$  is **distinct** if it holds that

$$\lambda_i \neq \lambda_j \quad \text{for all } i, j \in [d] \text{ such that } i \neq j.$$

**Theorem 1.** *The matrix  $\text{BLT}_n(\boldsymbol{\alpha}, \boldsymbol{\lambda})$  is invertible for any integer  $n > 0$  for any parameters  $\boldsymbol{\alpha} \in \mathbb{R}^d$  and  $\boldsymbol{\lambda} \in \mathbb{R}^d$  for all integers  $n > 0$  and  $d > 0$ . In addition, if the scale parameters are positive ( $\alpha_i > 0$ ) and satisfy  $\sum_{i=1}^d \alpha_i < 1$ , and the decay parameters  $\boldsymbol{\lambda} \in (0, 1)^d$  are distinct, then there exist parameters  $\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\lambda}} \in \mathbb{R}^d$  with  $\hat{\boldsymbol{\lambda}}$  distinct such that  $\text{BLT}_n(\boldsymbol{\alpha}, \boldsymbol{\lambda})^{-1} = \text{BLT}_n(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\lambda}})$  for all integers  $n > 0$ . Further, the scale parameters of the inverse are negative (i.e.  $\hat{\alpha}_i < 0$  for all  $i$ ), and the decay parameters of the inverse satisfy have the following:*

- (a) *If  $\sum_{i=1}^d \alpha_i / \lambda_i < 1$ , then  $\hat{\lambda}_i \in (0, 1)$  for each  $i \in [d]$ .*
- (b) *If  $\sum_{i=1}^d \alpha_i / \lambda_i > 1$ , then there exists an integer  $j \in [d]$  such that  $\hat{\lambda}_j \in (-1, 0)$  and  $\hat{\lambda}_i \in (0, 1)$  for all  $i \in [d], i \neq j$ .*
- (c) *Finally, if  $\sum_{i=1}^d \alpha_i / \lambda_i = 1$ , then there exists an integer  $j \in [d]$  such that  $\hat{\lambda}_j = 0$  and  $\hat{\lambda}_i \in (0, 1)$  for all  $i \in [d], i \neq j$ .*

Furthermore, these inverse parameters  $\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\alpha}}$  are unique (up to permutations of indices).

Following Eq. (2),  $\boldsymbol{C}[2, 1] = \sum_{i=1}^d \alpha_i$ , and so the assumption  $\sum_{i=1}^d \alpha_i < 1$  is useful because we typically want the first column of the  $\boldsymbol{C}$  matrix to be decreasing in the context of streaming differential privacy [1, 11]. Furthermore, we empirically observe that BLT parameters optimized for the max loss tend to satisfy  $\hat{\lambda}_i \in (0, 1)$ , paralleling the assumption that  $\lambda_i \in (0, 1)$  for the original BLT; this is also true for the theoretical construction of Dvijotham et al. [1]. This corresponds to the regime of  $\sum_{i=1}^d \alpha_i / \lambda_i < 1$  as per Theorem 1(a).

BLTs satisfying the condition of Theorem 1(c) are degenerate in the sense that one of the decay parameters is exactly zero. We give an example where this holds.

**Example.** *Consider a BLT of degree  $d = 2$  with parameters  $\boldsymbol{\alpha} = (2/5, 1/5)$  and  $\boldsymbol{\lambda} = (4/5, 2/5)$  so that  $\sum_i \alpha_i / \lambda_i = 1/2 + 1/2 = 1$ . Then, we have that its inverse is a BLT given by the parameters<sup>3</sup>*

<sup>3</sup>This can be verified, for instance, using the upcoming Lemma 3.

$\hat{\boldsymbol{\alpha}} = (-1/15, -8/15)$  and  $\hat{\boldsymbol{\lambda}} = (3/5, 0)$ . In particular, notice that we have a decay parameter of exactly 0 in the inverse, and hence  $\hat{\alpha}_2 = -8/15$  only influences the second Toeplitz coefficient  $(\mathbf{C}^{-1})[j+1, j] = \sum_{i=1}^d \hat{\alpha}_i$ .

For practical applications, we usually optimize for the BLT parameters, and hence will not reach the measure zero set of parameters with  $\sum_i \alpha_i / \lambda_i = 1$  when the optimizer is initialized randomly. Further, a small numerical error in the BLT parameters is enough to make this degeneracy vanish.

**Comparison to Previous Literature** We make two remarks where Theorem 1 significantly simplifies and extends prior work. First, Dvijotham et al. [1, Lemma 5.2] showed (using our notation) that given an appropriate pair of decay parameters  $(\boldsymbol{\lambda}, \hat{\boldsymbol{\lambda}})$ , there exist scale parameters  $(\boldsymbol{\alpha}, \hat{\boldsymbol{\alpha}})$  such that  $\text{BLT}(\boldsymbol{\alpha}, \boldsymbol{\lambda})^{-1} = \text{BLT}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\lambda}})$ . We strengthen this result with the help of Theorem 1 to show that a (BLT, inverse BLT) system can be parameterized in two equivalent ways. We also give significantly simplified expressions compared to [1, Lemma 5.2].

**Theorem 2.** Let  $\boldsymbol{\lambda}, \hat{\boldsymbol{\lambda}} \in \mathbb{R}^d$  be distinct non-zero vectors (i.e.,  $\lambda_i \neq \lambda_j$  and  $\hat{\lambda}_i \neq \hat{\lambda}_j$  for all  $i \neq j$ ) that also satisfy  $\lambda_i \neq \hat{\lambda}_j$  for all  $i, j \in [d]$ . Then, there exist unique scale parameters  $\boldsymbol{\alpha}, \hat{\boldsymbol{\alpha}} \in \mathbb{R}^d$  that achieve  $\text{BLT}_n(\boldsymbol{\alpha}, \boldsymbol{\lambda}) = \text{BLT}_n(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\lambda}})^{-1}$  for all  $n > 0$ , given by:

$$\alpha_i = \frac{\prod_{j=1}^d \lambda_i - \hat{\lambda}_j}{\prod_{j \neq i} \lambda_i - \lambda_j}, \quad \text{and} \quad \hat{\alpha}_i = \frac{\prod_{j=1}^d \hat{\lambda}_i - \lambda_j}{\prod_{j \neq i} \hat{\lambda}_i - \lambda_j}. \quad (6)$$

Furthermore, the following two parameterizations describe the same class of (BLT, inverse BLT) systems satisfying  $\text{BLT}(\boldsymbol{\alpha}, \boldsymbol{\lambda})^{-1} = \text{BLT}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\lambda}})$ :

- (a) positive scale parameters  $\boldsymbol{\alpha} \in \mathbb{R}_{++}^d$  and distinct decay parameters  $\boldsymbol{\lambda} \in (0, 1)^d$  of the BLT that satisfy  $\sum_{i=1}^d \alpha_i / \lambda_i < 1$ ;
- (b) a pair of decay parameters  $\boldsymbol{\lambda}, \hat{\boldsymbol{\lambda}}$  that satisfy the strict interlacing condition

$$1 > \lambda_1 > \hat{\lambda}_1 > \lambda_2 > \hat{\lambda}_2 > \dots > \hat{\lambda}_{d-1} > \lambda_d > \hat{\lambda}_d > 0.$$

Given either parameterization, the system  $\text{BLT}(\boldsymbol{\alpha}, \boldsymbol{\lambda})^{-1} = \text{BLT}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\lambda}})$  is uniquely determined.

Next, we turn to BLT inversion theorems. Dvijotham et al. [1, Proposition 5.6] show that the inverse of  $\text{BLT}_n(\boldsymbol{\alpha}, \boldsymbol{\lambda})$  is a lower-triangular and Toeplitz matrix whose first column  $\hat{c}_1, \hat{c}_2, \dots$  is given by

$$\hat{c}_t = \begin{cases} \mathbf{u}^\top \mathbf{v} + \kappa, & \text{if } t = 1, \\ \mathbf{u}^\top \mathbf{W}^{t-1} \mathbf{v} & \text{if } t > 1. \end{cases} \quad (7)$$

for vectors  $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$ , a matrix  $\mathbf{W} \in \mathbb{R}^{d \times d}$  and a scalar  $\kappa \in \mathbb{R}$ .<sup>4</sup> Theorem 1 implies that we can instead take

$$\mathbf{W} = \begin{pmatrix} \hat{\lambda}_1 & & \\ & \ddots & \\ & & \hat{\lambda}_d \end{pmatrix} \in \mathbb{R}^{d \times d}, \quad \mathbf{u} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^d, \quad \mathbf{v} = \begin{pmatrix} \hat{\alpha}_1 / \hat{\lambda}_1 \\ \vdots \\ \hat{\alpha}_d / \hat{\lambda}_d \end{pmatrix} \in \mathbb{R}^d, \quad \kappa = 1 - \sum_{i=1}^d \frac{\alpha_i}{\lambda_i}.$$

<sup>4</sup>This representation is not unique, as  $\tilde{\mathbf{u}} = \mathbf{M}\mathbf{u}$ ,  $\tilde{\mathbf{W}} = \mathbf{M}\mathbf{W}\mathbf{M}^\top$ ,  $\tilde{\mathbf{v}} = \mathbf{M}\mathbf{v}$  satisfies  $\mathbf{u}^\top \mathbf{W}^\tau \mathbf{v} = \tilde{\mathbf{u}}^\top \tilde{\mathbf{W}}^\tau \tilde{\mathbf{v}}$  for all  $\tau \geq 0$  for any orthonormal matrix  $\mathbf{M}$ .

We can verify by direct computation that this produces  $\hat{c}_1 = 1$  and  $\hat{c}_t = \sum_{i=1}^d \hat{\alpha}_i \hat{\lambda}_i^{t-2}$  for  $t \geq 2$ , as desired. In other words, Theorem 1 shows the existence of a  $\mathbf{W}$  of rank  $d$  and all real eigenvalues satisfying Eq. (7).

The representation in Theorem 1 is more convenient from a computational perspective, as operations on diagonal  $\mathbf{W}$  can be implemented more efficiently. Of course, given any  $\mathbf{u}, \mathbf{v}, \mathbf{W}$  that satisfy  $\hat{c}_t = \mathbf{u}^\top \mathbf{W}^{t-1} \mathbf{v}$  (e.g. by the approach of [1, Proposition 5.6]), we can then find the inverse BLT decay parameter  $\hat{\lambda}$  by diagonalizing  $\mathbf{W} = \mathbf{M} \text{diag}(\hat{\lambda}) \mathbf{M}^{-1}$ , *assuming* it is possible, and scale parameter  $\hat{\alpha} = (\mathbf{M}^\top \mathbf{u}) \odot (\mathbf{M}^{-1} \mathbf{v})$ , where  $\odot$  denotes component-wise multiplication of vectors. By showing that  $\mathbf{W}$  has all real and unique eigenvalues, Theorem 1 establishes that this matrix is diagonalizable.

**Algorithms for BLT Inversion** We give an algorithm to directly find the parameters  $\hat{\alpha}$  and  $\hat{\lambda}$  of the inverse BLT. Specifically, the decay parameter  $\hat{\lambda}$  is obtained from the finding the roots of a degree- $d$  polynomial  $q$  (whose roots are guaranteed to all be real). While these roots can be analytically obtained for degrees  $d = 4$  or lower, analytical expressions for the roots of general polynomials of degree  $d \geq 5$  are impossible. (This is also known as the Abel–Ruffini theorem, see e.g. [12, 13].)

Instead, we use numerical polynomial root-finding procedures to find all the roots of the polynomial  $q$ . The typical approach proceeds by constructing a non-symmetric matrix, known as the companion matrix, and then finding its eigenvalues in  $O(d^3)$  time (see e.g. [14] or the upcoming Section 6)—these eigenvalues are exactly the roots of the polynomial  $q$ . Importantly, all these operations are supported by typical automatic differentiation frameworks, including JAX and PyTorch. This allows us to parameterize the optimization of a BLT and its inverse as  $(\alpha, \lambda)$ , rather than as  $(\lambda, \hat{\lambda})$  as in [1]. See Section 6 for details.

## 4 Technical Tools and Proof Outline

The proofs of Theorems 1 and 2 rely on a deep connection between Toeplitz matrices and ordinary generating functions.

The **ordinary generating function** of a sequence  $(c_t)_{t=1}^\infty$  is the formal power series:<sup>5</sup>

$$f_c(x) := \sum_{t=0}^{\infty} c_{t+1} x^t.$$

This is closely related to the  $Z$ -transform, which can be obtained by the symbolic substitution  $z = 1/x$ . The sequence  $(c_t)_{t=1}^\infty$  can be obtained from the Maclaurin expansion of the generating function:

$$f_c(x) = \sum_{t=0}^{\infty} \frac{f_c^{(t)}(0)}{t!} x^t = \sum_{t=0}^{\infty} c_{t+1} x^t \iff c_{t+1} = \frac{f_c^{(t)}(0)}{t!},$$

where  $f_c^{(t)}$  denotes the  $t^{\text{th}}$  derivative of the function  $f_c$  (assuming it exists).

The key relationship between lower-triangular Toeplitz matrices and generating functions is that the product of two Toeplitz matrices (i.e., the convolution of their first columns) is equivalent to

---

<sup>5</sup>The variable  $x$  in a formal power series should be interpreted as a formal symbol rather than a numerical value. Specifically, we neglect any concerns related to convergence.

the product of the generating functions of their respective coefficients. (This is analogous to the fact that the convolution of two sequences can be obtained from the product of their  $Z$ -transforms.)

**Lemma 3.** *For any sequences  $(a_t)_{t=1}^\infty, (b_t)_{t=1}^\infty, (c_t)_{t=1}^\infty$  that take values in a field  $K$  (e.g. the field  $\mathbb{R}$  of reals or  $\mathbb{C}$  of complex numbers), the following are equivalent:*

- (i) *The respective ordinary generating functions  $f_a, f_b$ , and  $f_c$  of sequences  $(a_t)_{t=1}^\infty, (b_t)_{t=1}^\infty$ , and  $(c_t)_{t=1}^\infty$  satisfy  $f_a(x) = f_b(x)f_c(x)$ .*
- (ii) *For any integer  $n > 0$ , the  $n \times n$  lower triangular Toeplitz matrices  $\mathbf{M}_a, \mathbf{M}_b, \mathbf{M}_c$  with respective first columns given by sequences  $(a_t)_{t=1}^n, (b_t)_{t=1}^n, (c_t)_{t=1}^n$  satisfy  $\mathbf{M}_a = \mathbf{M}_b \mathbf{M}_c$ .*

We are most interested in real sequences. In particular, Lemma 3 tells us that we can calculate the inverse  $\hat{\mathbf{C}} = \mathbf{C}^{-1}$  of any lower-triangular Toeplitz matrix  $\mathbf{C}$  whose first column is obtained as a prefix of the sequence  $(c_t)_{t=0}^\infty$  by the following steps:

- Compute its generating function  $f_c(x)$ ;
- Calculate the reciprocal  $\hat{f}_c(x) = 1/f_c(x)$ ;
- Calculate its Maclaurin series  $\hat{c}_{t+1} = \hat{f}_c^{(t)}(0)/(t!)$  for  $t \geq 0$ ;
- Construct the lower-triangular Toeplitz matrix  $\hat{\mathbf{C}}$  with first column  $\hat{c}_1, \hat{c}_2, \dots$

Moreover, this holds for any leading principal sub-matrix: that is,  $\mathbf{C}[1 : n, 1 : n]^{-1} = \hat{\mathbf{C}}[1 : n, 1 : n]$  for any integer  $n > 0$ . Thus, it suffices to consider infinite Toeplitz matrices; we denote infinite BLT matrices as  $\text{BLT}(\boldsymbol{\alpha}, \boldsymbol{\lambda})$  by dropping the subscript  $n$ .

**BLT to Generating Function** We start by computing the generating function of the BLT and inverse BLT:

**Lemma 4.** *The generating function  $f(x)$  of  $\text{BLT}(\boldsymbol{\alpha}, \boldsymbol{\lambda})$  is given by*

$$f(x) = 1 + x \frac{r(x)}{p(x)} = \frac{q(x)}{p(x)}, \quad (8)$$

where we define the polynomials

$$p(x) = \prod_{i=1}^d (1 - \lambda_i x), \quad r(x) = \sum_{i=1}^d \frac{\alpha_i p(x)}{1 - \lambda_i x}, \quad \text{and} \quad q(x) = p(x) + x r(x). \quad (9)$$

Moreover, the generating function of  $\hat{f}$  of the inverse matrix  $\text{BLT}(\boldsymbol{\alpha}, \boldsymbol{\lambda})^{-1}$  is given by

$$\hat{f}(x) = \frac{p(x)}{q(x)} = 1 - x \frac{r(x)}{q(x)}. \quad (10)$$

*Proof.* By summing the geometric series, we get

$$f(x) = 1 + \sum_{t=1}^{\infty} \left( \sum_{i=1}^d \alpha_i \lambda_i^{t-1} \right) x^t = 1 + \sum_{i=1}^d \frac{\alpha_i x}{1 - \lambda_i x}.$$

Simplifying this gives Eq. (8). For the inverse, we have from Lemma 3 that

$$\hat{f}(x) = \frac{1}{f(x)} = \frac{p(x)}{p(x) + x r(x)} = 1 - \frac{x r(x)}{p(x) + x r(x)} = 1 + x \frac{-r(x)}{q(x)}.$$

□

Note that the polynomial  $p$  is of degree  $d$ , while the polynomial  $r$  is of degree  $d - 1$ . The polynomial  $q(x) = p(x) + x r(x)$  is thus a sum of two degree- $d$  polynomials, and its degree  $D = \deg(q)$  can be at most  $d$ . The generating functions  $f$  and  $\hat{f}$  are rational functions of degree at most  $d$  in both the numerator and denominator.

**Generating Function to Inverse BLT** To reconstruct the inverse BLT, we need to find the Maclaurin series of its generating function  $\hat{f}$ . When  $\hat{f}$  is a rational function, as is the case in Lemma 4, this is most easily obtained by a partial fraction decomposition of  $\hat{f}(x)$ . This approach is also commonly used in solving recursions in combinatorics and discrete mathematics, and filter design in digital signal processing.

The first step to construct a partial fraction decomposition is to reason about the roots of the denominator  $q(x) = p(x) + x r(x)$ .

**Proposition 5.** *Consider the setting of Theorem 1 with parameters  $\alpha \in \mathbb{R}_{++}^d$  and  $\lambda \in (0, 1)^d$ , and let  $r(x)$  and  $p(x)$  be as defined in Eq. (9) (see Lemma 4). Then, the polynomial  $q(x) = p(x) + x r(x)$  has degree  $D = d - 1$  if  $\sum_{i=1}^d \alpha_i / \lambda_i = 1$  and  $D = d$  otherwise. Moreover, all its roots  $\nu_1, \dots, \nu_D$  are unique and real.*

Since the polynomial  $q(x) = p(x) + x r(x)$  has all unique and real roots, the partial fraction decomposition of  $r(x)/q(x)$  takes a very simple form:

**Proposition 6.** *Consider the setting of Proposition 5. The polynomials  $r$  and  $q$  are co-prime and we have a unique partial fraction decomposition*

$$\frac{r(x)}{q(x)} = - \sum_{i=1}^d \frac{\hat{\alpha}_i}{1 - \hat{\lambda}_i x} \tag{11}$$

for some  $\hat{\alpha}, \hat{\lambda} \in \mathbb{R}^d$  with  $\hat{\alpha}$  non-zero component-wise. Further, we have the following based on  $\deg(q) = D$ :

- (a) If  $D = d$ , then  $\hat{\lambda}_i \neq 0$  for all  $i \in [d]$ .
- (b) If  $D = d - 1$ , then  $\hat{\lambda}_i \neq 0$  for  $i \in [d - 1]$  and  $\hat{\lambda}_d = 0$ .
- (c) Finally, we have  $\hat{\lambda}_i = 1/\nu_i$  for  $i \in [D]$  where  $\nu_1, \dots, \nu_D$  are the roots of the degree- $D$  polynomial  $q$ .

Together with Lemmas 3 and 4, this immediately implies a partial fraction decomposition of the generating function  $\hat{f}$  of  $\text{BLT}(\alpha, \lambda)^{-1}$ :

**Corollary 7.** *In the setting of Proposition 6, we have that the generating function of  $\text{BLT}(\boldsymbol{\alpha}, \boldsymbol{\lambda})^{-1}$  is given by*

$$\hat{f}(x) = 1 + x \frac{-r(x)}{q(x)} = 1 + \sum_{i=1}^d \frac{\hat{\alpha}_i x}{1 - \hat{\lambda}_i x}, \quad (12)$$

where  $\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\lambda}}$  are as given in Proposition 6. In particular, we have that  $\text{BLT}(\boldsymbol{\alpha}, \boldsymbol{\lambda})^{-1} = \text{BLT}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\lambda}})$ .

Lemma 18 of Appendix A conveniently summarizes all the generating function results that we have established so far.

**Inverse BLT Parameter Properties** In order to complete the proof, we must argue about the signs and magnitudes of  $\hat{\alpha}_i$ 's and  $\hat{\lambda}_i$ 's depending on the value of  $\sum_i \alpha_i / \lambda_i$ . The main technical result we show is:

**Proposition 8.** *Consider the setting of Theorem 1 with parameters  $\boldsymbol{\alpha} \in \mathbb{R}_{++}^d$  and  $\boldsymbol{\lambda} \in (0, 1)^d$  distinct, and let  $r(x)$  and  $p(x)$  be as defined in Eq. (9) (see Lemma 4). Then, we have the following:*

- (a) *If  $\sum_{i=1}^d \alpha_i / \lambda_i < 1$ , then all roots  $\nu_1, \dots, \nu_d$  of  $q$  lie in  $(1, \infty)$ . Thus, the parameter  $\hat{\lambda}_i = 1/\nu_i$  in the denominator of the partial fraction decomposition (11) lies in  $(0, 1)$  for each  $i = 1, \dots, d$ .*
- (b) *If  $\sum_{i=1}^d \alpha_i / \lambda_i > 1$ , then one root of  $q$  lies in  $(-\infty, -1)$  while all other roots lie in  $(1, \infty)$ . Thus, we have  $\hat{\lambda}_i \in (0, 1)$  for  $i = 1, \dots, d-1$  and  $\hat{\lambda}_d \in (-1, 0)$ .*
- (c) *Irrespective of the value of  $\sum_{i=1}^d \alpha_i / \lambda_i$ , we have that the numerator of the the partial fraction decomposition (11) satisfies  $\hat{\alpha}_i < 0$  for  $i = 1, \dots, d$ .*

Given Propositions 6 and 8, the proof of Theorem 1 is immediate.

*Proof of Theorem 1.* The matrix  $\text{BLT}_n(\boldsymbol{\alpha}, \boldsymbol{\lambda})$  is a lower-triangular matrix with all ones on the diagonal; thus it is invertible for all  $n > 0$ . By Corollary 7, we have that  $\text{BLT}(\boldsymbol{\alpha}, \boldsymbol{\lambda})^{-1} = \text{BLT}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\lambda}})$ , where  $\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\lambda}}$  are as in Proposition 6. Then, the signs and magnitudes of  $\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\lambda}}$  imply the various parts of Theorem 1. In particular, part (a) of Theorem 1 follows directly from from part (a) of Proposition 8. Similarly, part (b) of Theorem 1 follows from Proposition 8(b). Next, Theorem 1(c) follows from Proposition 6(b), while the negative scale parameters follows from Proposition 8(c). Finally, the uniqueness of the inverse BLT parameters follows from two observations:

- the decay parameters  $\hat{\boldsymbol{\lambda}}$  are obtained as the roots of the polynomial  $q(x) = p(x) + x r(x)$  and are unique.
- the scale parameters  $\hat{\boldsymbol{\alpha}}$  are the coefficients of a partial fraction and are unique as per Proposition 6.

□

We summarize the key ideas behind the proofs of Propositions 5, 6 and 8, with full proofs appearing in Section 5 (see Figures 1 and 2 for a key idea):

- Let  $\mu_1 < \dots < \mu_d$  denote the roots of the polynomial  $p(x) = \prod_{i=1}^d (1 - \lambda_i x)$ . Since  $\lambda_i \in (0, 1)$ , we have  $\mu_i = \lambda_i^{-1} \in (1, \infty)$ . Notably, all these roots are real.

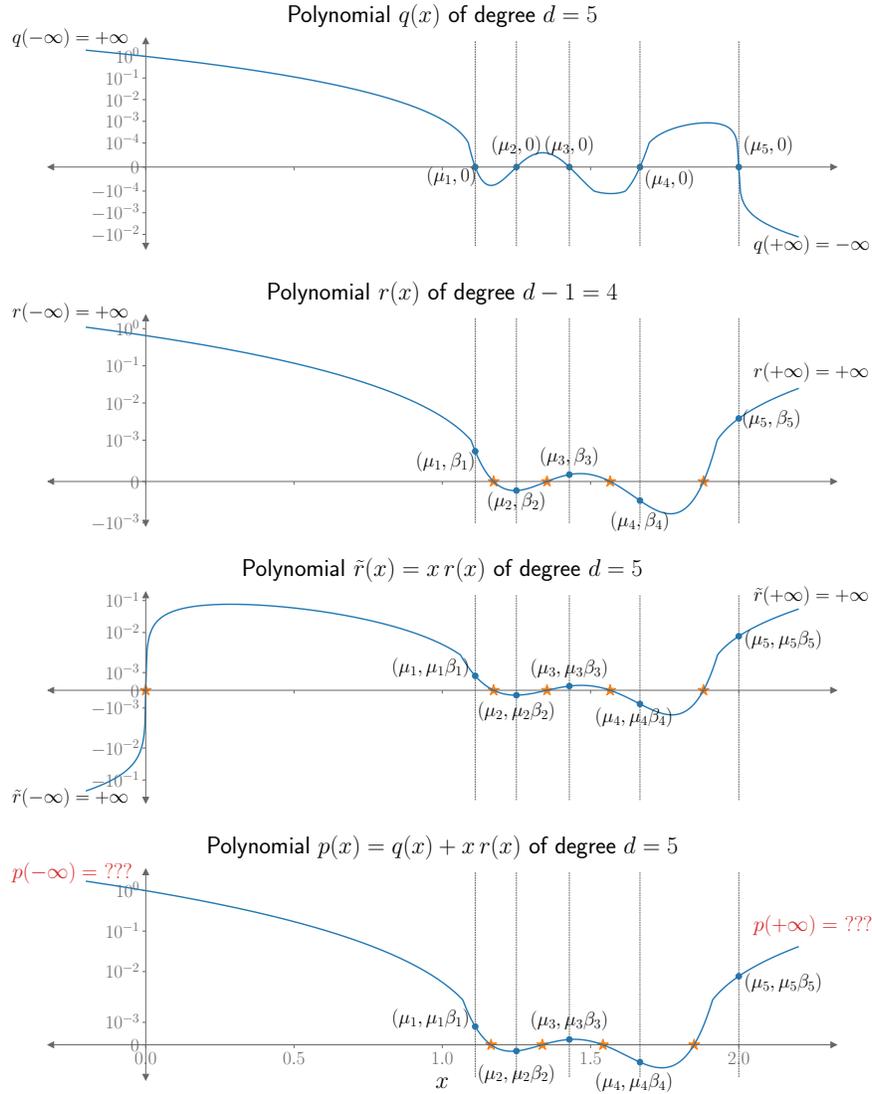


Figure 1: Illustrations of the polynomials  $r, p, q$  for  $d = 5$  in symmetrical log scale. **First row:** Let  $\mu_i := 1/\lambda_i > 1$  for  $i = 1, \dots, d$  denote the roots of  $p(x)$  in ascending order. **Second row:** We show that  $\beta_i := r(\mu_i)$  is positive for  $i$  odd and negative for  $i$  even. Due to sign changes, each of the  $d - 1$  roots of  $r$  lies between  $(\mu_i, \mu_{i+1})$  for some  $i$  (denoted by the orange star). **Third row:** By the same argument, each  $d - 1$  non-zero roots of  $x r(x)$  lie in  $(\mu_i, \mu_{i+1})$  for some  $i$ . **Last row:** The same argument accounts for  $d - 1$  roots of the degree- $d$  polynomial  $q(x) = p(x) + x r(x)$ . Since  $d - 1$  roots of the degree- $d$  real polynomial  $q$  are real, the final root must be real as well, establishing Proposition 5. This is continued in Figure 2; a similar argument also works  $d$  even—see Figure 3.

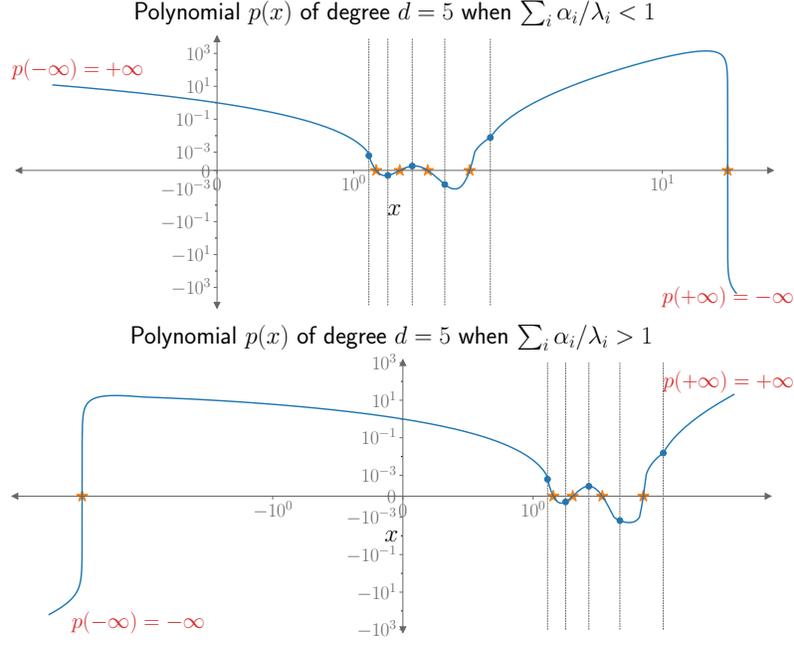


Figure 2: Continued from Figure 1, which shows  $d - 1$  roots of  $q(x) = p(x) + xr(x)$  for an example with  $d = 5$ . This figure illustrates how the final  $d^{\text{th}}$  root of  $q(x)$  depends on the BLT parameters  $\alpha, \lambda$ . As previously, the dotted lines denote the roots  $\mu_1, \dots, \mu_d$  of  $p(x)$  (where  $\mu_i = 1/\lambda_i$ ) and the orange stars denote the roots of  $q(x)$ . **Top:** When  $\sum_{i=1}^d \alpha_i/\lambda_i < 1$ , the last root of  $q$  is positive as well, as in Proposition 8(a). **Bottom:** When  $\sum_{i=1}^d \alpha_i/\lambda_i > 1$ , then  $q$  has a negative root, as in Proposition 8(b).

- The key step in the proof is to argue about the sign of  $\beta_i := r(\mu_i)$ . We show in the upcoming Property 9 that  $\beta_i > 0$  for  $i$  odd and  $\beta_i < 0$  for  $i$  even.
- Thus, we have that  $q(\mu_i) = p(\mu_i) + \mu_i r(\mu_i) = \mu_i \beta_i$  is positive for  $i$  odd and negative for  $i$  even. Thus,  $q$  admits a root in the interval  $(\mu_i, \mu_{i+1})$  for each  $i = 1, \dots, d$ . This accounts for  $d - 1$  roots of  $q$ , which are real and positive. Thus,  $q$  is of degree  $d - 1 \leq D \leq d$ . If  $q$  is of degree  $D = d$ , then the  $d^{\text{th}}$  root of  $q$  is real as well, since complex roots of a real polynomial can only occur in conjugate pairs—this gives Proposition 5.
- Next, reasoning about the partial fraction decomposition of the ordinary generating function  $\hat{f}$  of the inverse BLT (defined in Lemma 4) gives us Proposition 6. In view of Corollary 7, all that is now left for Proposition 8 is to reason about the final root of  $q$  (it can either be smaller than 0 or larger than 1) as well as the  $\hat{\alpha}_i$  coefficients.
- Next, we argue about the last root of  $q$ , in the case that its degree is  $D = d$ . The leading coefficients of  $p(x)$  and  $xr(x)$  have opposite signs; see Figure 1. We calculate the coefficient  $q_d$  of  $x^d$  in the polynomial  $q$  as

$$q_d = (-1)^d \left( \prod_{i=1}^d \lambda_i \right) \left( 1 - \sum_{i=1}^d \frac{\alpha_i}{\lambda_i} \right).$$

Thus, we get the following cases (see Figure 2):

- (Case I) If  $\sum_{i=1}^d \alpha_i/\lambda_i < 1$ , then  $q(-\infty) = +\infty$ . We show that  $q(\mu_d)$  and  $q(\infty)$  have opposite signs. Thus, the  $d^{\text{th}}$  root of the polynomial  $q$  is also larger than 1.
- (Case II) If  $\sum_{i=1}^d \alpha_i/\lambda_i > 1$ , then  $q(-\infty) = -\infty$ . We show that  $q(-1) > 0$ , leading to the conclusion that the  $d^{\text{th}}$  root of the polynomial  $q$  lies in  $(-\infty, -1)$ .
- (Case III) If  $\sum_{i=1}^d \alpha_i/\lambda_i = 1$ , the leading order terms of  $p(x)$  and  $xr(x)$  cancel out and  $q(x)$  is a polynomial of degree  $d - 1$ . Thus, it only has  $d - 1$  roots, all of which have previously been accounted for.

This yields parts (a) and (b) of Proposition 8.

- Finally, we need to argue that the scale parameters  $\hat{\alpha}_i$  are negative. We do so by arguing about their signs from their closed-form expressions.

See Section 5 for full proofs of each of these steps.

**Proof of Theorem 2** We conclude this section with a proof of Theorem 2.

*Proof of Theorem 2.* The expressions of Eq. (6) can be obtained by simplifying Eq. (5.2) of Dvijotham et al. [1, Lemma 5.2]. Alternatively, Lemma 14 of Section 5 gives a short and elementary proof of the expression for  $\hat{\alpha}_i$ . Then, the expression for  $\alpha_i$  can be obtained by symmetry.

Next, we turn to the equivalence of various representations.

- (a)  $\implies$  (b): Assuming w.l.o.g. the ordering  $\lambda_1 > \dots > \lambda_d$ , Theorem 1(a) gives us unique  $\hat{\alpha} < 0$  and  $\hat{\lambda} \in (0, 1)^d$ . The proof that  $\lambda, \hat{\lambda}$  satisfy the claimed ordering is given in Corollary 13 in Section 5.
- (b)  $\implies$  (a): We can evaluate the signs of the scale parameters  $\alpha_i$  in Equation (6). For  $\alpha_1$ , we have that all the terms in Eq. (6) are positive. For  $i = 2$ , we have only one negative term  $(\lambda_2 - \hat{\lambda}_1)$  in the numerator and only one negative term  $(\lambda_2 - \lambda_1)$  in the denominator, and is thus positive. Similarly, each  $\alpha_i$  has  $i - 1$  negative terms each in the numerator and denominator, so that  $\alpha_i > 0$ . To show the bound on  $\sum_i \alpha_i/\lambda_i$ , rearranging Eq. (19) of Lemma 14 (with  $(\alpha, \lambda)$  and  $(\hat{\alpha}, \hat{\lambda})$  swapped) gives:

$$\sum_{i=1}^d \frac{\alpha_i}{\lambda_i} = 1 - \frac{\prod_{i=1}^d \hat{\lambda}_i}{\prod_{i=1}^d \lambda_i} < 1,$$

since  $0 < \hat{\lambda}_i/\lambda_i < 1$  by assumption.

□

## 5 Full Proof Details

We give the full proofs of all remaining statements from Section 4, i.e., Propositions 5, 6 and 8. In particular, Proposition 5 from Section 4 is a special case of Properties 11 and 12(a), while Proposition 6 is established in Step 4 below and Proposition 8 follows directly from Property 12

and Proposition 15 below. The assumptions of Theorem 1 are assumed to hold throughout. We also fix the BLT parameters  $\alpha \in \mathbb{R}_{++}^d$  and  $\lambda \in (0, 1)^d$  throughout, except when explicitly mentioned otherwise (e.g. Lemma 14 is a notable exception).

**Step 1: Notation and Properties of  $p(x)$**  As introduced previously, define  $\mu_i = \lambda_i^{-1}$  for  $i = 1, \dots, d$ . These are the roots of the polynomial  $p(x)$  from Eq. (9). We assume that  $\mu_1 < \mu_2 < \dots < \mu_d$ ; this is without loss of generality as we assumed in Theorem 1 that  $\lambda_i$ 's are distinct. Finally, define the constant

$$M = \prod_{i=1}^d \mu_i. \quad (13)$$

Using this, we can rewrite  $p(x)$  from Eq. (9) as

$$p(x) = \frac{(-1)^d}{M} \prod_{i=1}^d (x - \mu_i). \quad (14)$$

**Step 2: Behavior of  $r(x)$  at the roots of  $p(x)$**  Let  $\beta_i := r(\mu_i)$  be the values of the polynomial  $r$  at the roots  $\mu_1, \dots, \mu_d$  of  $p$ . (See Eq. (9) for the definition of  $r$ .) We now argue that  $\beta_i$  is positive if  $i$  is odd and negative otherwise.

**Property 9.** *We have*

$$\beta_i := r(\mu_i) = \frac{\alpha_i \mu_i}{M} \prod_{j \neq i} (\mu_j - \mu_i).$$

*In particular,  $\beta_i > 0$  if  $i$  is odd and  $\beta_i < 0$  if  $i$  is even.*

*Proof.* Starting from the definition of the degree- $(d-1)$  polynomial  $r$  from Eq. (9), we have,

$$r(x) = \sum_{i=1}^d \alpha_i \prod_{j \neq i} (1 - \lambda_j x) = \sum_{i=1}^d \frac{\alpha_i \mu_i}{M} \prod_{j \neq i} (\mu_j - x), \quad (15)$$

where we substituted  $\mu_i = \lambda_i^{-1}$  and  $M = \prod_{i=1}^d \mu_i$ . When computing  $r(\mu_i)$ , we note that all but the  $i^{\text{th}}$  term will be zero, yielding the claimed expression for  $\beta_i$ . Next, we turn to the signs:  $\beta_i$  has  $(i-1)$  negative terms in the product, so its sign is the same as  $(-1)^{i-1}$ .  $\square$

**Remark 10.** *The polynomial  $r(x)$  can also be interpreted through the lens of Lagrange interpolation. We can rewrite Eq. (15) as*

$$r(x) = \sum_{i=1}^d \beta_i \prod_{j \neq i} \frac{x - \mu_j}{\mu_i - \mu_j},$$

*which is the Lagrange interpolant through  $(\mu_1, \beta_1), \dots, (\mu_d, \beta_d)$ . In fact,  $r(x)$  is the unique degree  $(d-1)$  polynomial interpolating these  $d$  points. This fact was used in the proof of [1, Lemma 5.2], but we prove Property 9 (and the upcoming Lemma 14) by more direct and elementary means.*

**Step 3: Roots of  $q(x)$**  The polynomial  $q(x) = p(x) + x r(x)$  is of degree  $D \leq d$ . Let  $\nu_1, \dots, \nu_D$  be the roots of the polynomial (if they exist). We establish Proposition 5 with some additional properties which will be useful later:

**Property 11.** *The polynomial  $q(x) = p(x) + x r(x)$  is of degree  $D \in \{d-1, d\}$  are all its roots are unique and real. Further, there exists a root  $\nu_i$  of  $q$  in the interval  $(\mu_i, \mu_{i+1})$  for  $i = 1, \dots, d-1$ .*

*Proof.* Recall that the  $\mu_i > 1$  are the roots of  $p(x)$ . From Property 9, we deduce that  $q(\mu_i) = p(\mu_i) + \mu_i r(\mu_i) = \mu_i \beta_i$  is positive for  $i$  odd and negative for  $i$  even. Next, we invoke the intermediate value theorem: since  $q(\mu_i)$  and  $q(\mu_{i+1})$  have opposite signs and  $q$  is a polynomial (and thus continuous), there exists  $\nu_i \in (\mu_i, \mu_{i+1})$  such that  $q(\nu_i) = 0$  for each  $i = 1, \dots, d-1$ .

Finally, since  $d-1$  roots exist, the polynomial  $q$  can only be of degree  $d$  or  $d-1$ . If  $q$  is of degree  $D = d-1$ , there is nothing left to prove, so suppose that  $q$  is of degree  $D = d$ . Since  $d-1$  roots of a degree- $d$  real polynomial are real, then the final  $d^{\text{th}}$  root should be real as well—this is because complex roots can only occur in conjugate pairs.  $\square$

It remains to argue about the potential final root of the polynomial  $q$ :

**Property 12.** *In the setting of Property 11, we have the following:*

- (a) *If  $\sum_{i=1}^d \alpha_i / \lambda_i = 1$ , then  $q$  is a polynomial of degree  $D = d-1$ . Otherwise, it is of degree  $D = d$ .*
- (b) *If  $\sum_{i=1}^d \alpha_i / \lambda_i < 1$ , then, the  $d^{\text{th}}$  root  $\nu_d$  of  $q$  satisfies  $\nu_d > \mu_d > 1$ .*
- (c) *If  $\sum_{i=1}^d \alpha_i / \lambda_i > 1$  but  $\sum_{i=1}^d \alpha_i < 1$ , then, the  $d^{\text{th}}$  root  $\nu_d$  of  $q$  satisfies  $\nu_d < -1$ .*

*Proof.* We reason about the leading coefficient  $q_d$  of  $x^d$  in the polynomial  $q(x) = p(x) + x r(x)$  such that  $q(x) - q_d x^d$  is a polynomial of degree  $(d-1)$ . From Eqs. (14) and (15), we deduce that

$$q_d = \frac{(-1)^d}{M} \left( 1 - \sum_{i=1}^d \alpha_i \mu_i \right) = \frac{(-1)^d}{M} \left( 1 - \sum_{i=1}^d \frac{\alpha_i}{\lambda_i} \right).$$

Part (a) follows because  $q_d = 0$  if and only if  $\sum_{i=1}^d \alpha_i / \lambda_i = 1$ . Next, suppose that  $\sum_{i=1}^d \alpha_i / \lambda_i < 1$ . We have two cases:

- $d$  is odd (see Figure 2 for reference): We have that  $q(\mu_d) > 0$  and  $\lim_{x \rightarrow \infty} q(x) = -\infty$  (because  $q_d < 0$ ).
- $d$  is even (see Figure 4 for reference): We have that  $q(\mu_d) < 0$  and  $\lim_{x \rightarrow \infty} q(x) > 0$  (because  $q_d > 0$ ).

In both cases, by the intermediate value theorem, the final root  $\nu_d$  of  $q$  must lie in  $(\mu_d, \infty)$ , yielding part (b).

The proof of part (c) also proceeds similarly. We have  $\lim_{x \rightarrow -\infty} q(x) < 0$  both for even and odd degree  $d$ . On the other hand, we have

$$q(-1) = p(-1) - r(-1) = \prod_{i=1}^d (1 + \lambda_i) \left( 1 - \sum_{i=1}^d \frac{\alpha_i}{1 + \lambda_i} \right) > 0,$$

where the last inequality followed from  $\alpha_i, \lambda_i > 0$  and

$$\sum_{i=1}^d \frac{\alpha_i}{1 + \lambda_i} < \sum_{i=1}^d \alpha_i < 1$$

by assumption. Another invocation of the intermediate value theorem implies that  $\nu_d \in (-\infty, -1)$ , completing the proof.  $\square$

We summarize the ordering of the decay parameters  $\boldsymbol{\lambda}, \hat{\boldsymbol{\lambda}}$ :

**Corollary 13.** *In the setting of Proposition 8, we have*

$$\lambda_1 > \hat{\lambda}_1 > \lambda_2 > \hat{\lambda}_2 > \cdots > \hat{\lambda}_{d-1} > \lambda_d > \hat{\lambda}_d.$$

*Proof.* Recall that we assumed the roots  $\mu_i = \lambda_i^{-1}$  of  $p(x)$  are ordered as  $\mu_1 < \cdots < \mu_d$ . Proposition 5 and Property 12 tell us that the roots  $\nu_1, \dots, \nu_d$  of  $q(x)$  satisfy  $\mu_i < \nu_i < \mu_{i+1}$  for  $i = 1, \dots, d-1$ . Thus, we have that  $\lambda_i = \mu_i^{-1}$  for  $i \in [d]$  and  $\hat{\lambda}_i = \nu_i^{-1}$  for  $i \in [d]$  are ordered as claimed. If  $\sum_i \alpha_i / \lambda_i < 1$  (as in Proposition 8(a)), we have that  $\nu_d > \mu_d$ , leading to  $\lambda_d > \hat{\lambda}_d$ . If not, we have  $\hat{\lambda}_d \leq 0$  by Property 12 again and  $\lambda_d > 0$  by assumption, leading to the claimed order.  $\square$

**Step 4: Partial Fraction Decomposition** We now give the proof of Proposition 6 from Section 4.

*Proof of Proposition 6.* We first prove that the polynomials  $r$  and  $q$  are co-prime, meaning that they do not share any roots.<sup>6</sup> Note that  $r$  and  $p$  are co-prime because  $r(\mu_i)$  is non-zero as per Property 9, where  $\mu_i$ 's are the roots of  $p$ . Then, we get that the greatest common divisor (denoted “gcd”) of  $r$  and  $q$  is

$$\gcd(r(x), q(x)) = \gcd(r(x), p(x) + x r(x)) = \gcd(r(x), p(x)) = 1,$$

where the second equality used the property that  $\gcd(r, p) = \gcd(r, p + \varphi r)$  for any polynomial  $\varphi$  (we take  $\varphi(x) = x$ ). Thus, the  $r(x)/q(x)$  is the ratio of degree  $d-1$  polynomial to a degree  $D \in \{d-1, d\}$  polynomial. Using the fact that the roots  $\nu_1, \dots, \nu_D$  of the polynomial  $q$  are real and unique by Proposition 5, we have the general form of the partial fraction decomposition

$$\frac{r(x)}{q(x)} = \kappa_0 + \sum_{i=1}^D \frac{\kappa_i}{x - \nu_i}, \quad (16)$$

for some reals  $\kappa_0, \kappa_1, \dots, \kappa_D$ . Note that  $\kappa_1, \dots, \kappa_D$  are non-zero, while  $\kappa_0$  is allowed to take zero values. Indeed, if any of  $\kappa_1, \dots, \kappa_D$  were zero, we would not obtain the correct degree in the denominator on the left side. We consider two separate cases depending on the degree  $D$  of the polynomial  $q$ .

- Case  $D = d$ : If  $\kappa_0 \neq 0$ , then numerator on the right side would have degree  $d$ . This is a contradiction because the degree of the numerator  $r$  on the left side is  $d-1$ . Thus, we must have  $\kappa_0 = 0$ . In this case, we set  $\hat{\lambda}_i = 1/\nu_i$  and  $\hat{\alpha}_i = \kappa_i/\nu_i$  for  $i \in [d]$  to obtain (11)—note that all of these are non-zero.

---

<sup>6</sup>Note that if  $r$  and  $q$  are co-prime, then the rational function  $r/q$  is irreducible.

- Case  $D = d-1$ : If  $\kappa_0 = 0$ , then the numerator on the right side would have degree  $D-1 = d-2$ , which is again a contradiction (since  $\deg(r) = d-1$  on the left side). Thus, we must have  $\kappa_0 \neq 0$ . In this case, we set  $\hat{\lambda}_i = 1/\nu_i$  and  $\hat{\alpha}_i = \kappa_i/\nu_i$  for  $i \in [d-1]$ —note that all of these are non-zero. We set  $\hat{\lambda}_d = 0$  and set the constant term  $\hat{\alpha}_d = -\kappa_0$  to obtain Eq. (11), as desired.

Further, since the factors of the denominator are linear and unique (since  $q(x)$  has no repeated roots), there exist unique coefficients  $\kappa_0, \dots, \kappa_D$  that satisfy the partial fraction decomposition.  $\square$

**Step 5: Scale Parameters of the Inverse BLT** It remains to show Proposition 8(c) regarding the sign of the scale parameters. We start with a self-contained proof of the expression for the scale parameters from Theorem 2, and then argue about its sign.

**Lemma 14.** *Let  $\boldsymbol{\lambda}, \hat{\boldsymbol{\lambda}} \in (\mathbb{R} \setminus \{0\})^d$  be distinct non-zero vectors (i.e.,  $\lambda_i \neq \lambda_j$  and  $\hat{\lambda}_i \neq \hat{\lambda}_j$  for all  $i \neq j$ ).<sup>7</sup> Suppose also that  $\lambda_i \neq \hat{\lambda}_j$  for all  $i, j \in [d]$ . Then, the constants  $\hat{\alpha}_1, \dots, \hat{\alpha}_d$  defined by*

$$\hat{\alpha}_i = \frac{\prod_{j=1}^d \hat{\lambda}_i - \lambda_j}{\prod_{j \neq i} \hat{\lambda}_i - \hat{\lambda}_j} \quad (17)$$

satisfy the following partial fraction decomposition for the rational function

$$\hat{f}(x) = \frac{p(x)}{q(x)} = \frac{\prod_{i=1}^d (1 - \lambda_i x)}{\prod_{i=1}^d (1 - \hat{\lambda}_i x)} = 1 + \sum_{i=1}^d \frac{\hat{\alpha}_i x}{1 - \hat{\lambda}_i x}. \quad (18)$$

Further,  $\hat{\alpha}_1, \dots, \hat{\alpha}_d$  from Eq. (17) are the unique values that satisfy the decomposition of Eq. (18). They also satisfy the identity

$$\sum_{i=1}^d \frac{\hat{\alpha}_i}{\hat{\lambda}_i} + \frac{\prod_{i=1}^d \lambda_i}{\prod_{i=1}^d \hat{\lambda}_i} = 1. \quad (19)$$

*Proof.* Eq. (17) can be derived by simplifying Eq. (5.3) of [1, Lemma 5.2], but we give a short elementary proof here. We start by multiplying Eq. (18) through by  $(1 - \hat{\lambda}_i x)$  and taking the limit  $x \rightarrow \hat{\lambda}_i^{-1}$  to get:

$$\lim_{x \rightarrow \hat{\lambda}_i^{-1}} (1 - \hat{\lambda}_i x) \hat{f}(x) = \lim_{x \rightarrow \hat{\lambda}_i^{-1}} \left[ (1 - \hat{\lambda}_i x) + \sum_{j=1}^d \frac{\hat{\alpha}_j x (1 - \hat{\lambda}_i x)}{1 - \hat{\lambda}_j x} \right] = \frac{\hat{\alpha}_i}{\hat{\lambda}_i}.$$

Thus, we can evaluate the scale parameter as

$$\hat{\alpha}_i = \lim_{x \rightarrow \hat{\lambda}_i^{-1}} \hat{\lambda}_i (1 - \hat{\lambda}_i x) \hat{f}(x) = \frac{\hat{\lambda}_i \prod_{j=1}^d (1 - \lambda_j / \hat{\lambda}_i)}{\prod_{j \neq i} (1 - \hat{\lambda}_j / \hat{\lambda}_i)},$$

and rearranging yields Eq. (17).

---

<sup>7</sup>Note that (a) we do not restrict any of the decay parameters to lie in  $(0, 1)$ , and (b) the assumptions allow us to swap the roles of  $(\boldsymbol{\lambda}, \boldsymbol{\alpha})$  and  $(\hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\alpha}})$ .

Next, we can show Eq. (19) by taking the limit of  $x \rightarrow \infty$  in Eq. (18):

$$\frac{(-1)^d \prod_{i=1}^d \lambda_i}{(-1)^d \prod_{i=1}^d \hat{\lambda}_i} = \lim_{x \rightarrow \infty} \frac{p(x)}{q(x)} = \lim_{x \rightarrow \infty} \left( 1 + \sum_{i=1}^d \frac{\hat{\alpha}_i x}{1 - \hat{\lambda}_i x} \right) = 1 - \sum_{i=1}^d \frac{\hat{\alpha}_i}{\hat{\lambda}_i}.$$

Finally, the uniqueness of the  $\alpha_i$  parameters follows from the uniqueness properties of the coefficients of a partial fraction decomposition of a rational function  $\hat{f}$  whose numerator and denominator are co-prime degree- $d$  polynomials. Specifically, this proof is identical to that of Proposition 6 with the observation that the degree- $(d-1)$  polynomial  $r(x) = (q(x) - p(x))/x$  is co-prime with  $q(x)$ , and is omitted for brevity.  $\square$

Finally, we argue about the signs of the scale parameters.

**Proposition 15.** *Consider the setting of Proposition 8. Then, we have that  $\hat{\alpha}_i < 0$  for all  $i \in [d]$ .*

*Proof.* Corollary 13 tells us that  $\lambda_1 > \hat{\lambda}_1 > \lambda_2 > \hat{\lambda}_2 > \dots > \hat{\lambda}_{d-1} > \lambda_d > \hat{\lambda}_d$  (irrespective of the value of  $\sum_i \alpha_i / \lambda_i - 1$ ). We use these relations to argue about the signs of each term in Eq. (17). For example, for  $i = 1$ , we have the term  $\hat{\lambda}_1 - \lambda_1$  in the numerator is negative but all other terms are strictly positive, leading to  $\hat{\alpha}_1 < 0$ . In general the expression for  $\hat{\alpha}_i$  has  $i$  negative terms in the numerator and  $i - 1$  negative term in the denominator, lead to  $\hat{\alpha}_i < 0$ .  $\square$

## 6 Differentiable Algorithms for Inverse BLTs

We now give an algorithm to find the parameters corresponding to the inverse BLT.

From the preceding sections proving Theorem 1, we see that the decay parameters  $\hat{\lambda}$  of the inverse BLT are obtained from the reciprocal of the roots of the polynomial  $q(x) = p(x) + x r(x)$ , with scale parameters  $\hat{\alpha}_i$  as derived in Lemma 14. The entire procedure is summarized in Algorithm 1.

In particular, the standard procedure to find roots of the polynomial  $q$  is finding the eigenvalues of the companion matrix; numerical procedures to compute the eigenvalues of a matrix are widely available in common software packages. This is based on the following result:

**Lemma 16** (Thm. 3.3.14 of [15]; [16]). *For any  $q_0, \dots, q_{d-1} \in \mathbb{C}$ , the characteristic polynomial  $q(x) = \det(\mathbf{M} - x\mathbf{I}_d)$  of the matrix*

$$\mathbf{M} = \left( \begin{array}{c|c} \mathbf{0}_{d-1}^\top & -q_0 \\ \hline \mathbf{I}_{d-1} & \begin{array}{c} -q_1 \\ \vdots \\ -q_{d-1} \end{array} \end{array} \right) \in \mathbb{C}^{d \times d}.$$

*is given by  $q(x) = q_0 + q_1 x + \dots + q_{d-1} x^{d-1} + x^d$ . In particular, the eigenvalues  $\nu_1, \dots, \nu_d$  (real or complex) of  $\mathbf{M}$  are the roots of the degree- $d$  polynomial  $q$ . Furthermore, the map  $(q_0, \dots, q_{d-1}) \mapsto (\nu_1, \dots, \nu_d)$  is continuously differentiable over the set*

$$R = \left\{ (q_0, \dots, q_{d-1}) \in \mathbb{C}^d : \text{the roots } \nu_1, \dots, \nu_d \text{ of } q(x) \text{ are distinct} \right\}.$$

Thus, we get the following correctness guarantee:

---

**Algorithm 1** Inverse BLT Parameterization
 

---

**Input:** BLT parameters  $\boldsymbol{\alpha} \in \mathbb{R}_{++}^d$  and  $\boldsymbol{\lambda} \in (0, 1)^d$  with  $\sum_{i=1}^d \alpha_i/\lambda_i \neq 1$  and  $\alpha_i$ 's distinct.

**Return** Parameters  $\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\lambda}} \in \mathbb{R}^d$  such that  $\text{BLT}_n(\boldsymbol{\alpha}, \boldsymbol{\lambda})^{-1} = \text{BLT}_n(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\lambda}})$  for all integers  $n > 0$ .

1: Define polynomials  $r, p, q$  as

$$p(x) = \prod_{i=1}^d (1 - \lambda_i x), \quad r(x) = \sum_{i=1}^d \alpha_i \prod_{j \neq i} (1 - \lambda_j x), \quad q(x) = p(x) + x r(x).$$

2: Find the (real and distinct) roots  $\nu_1, \dots, \nu_d$  of  $q$ . One way is to return the eigenvalues of the companion matrix

$$\mathbf{M} = \left( \begin{array}{c|c} \mathbf{0}_{d-1}^\top & -q_0/q_d \\ \hline \mathbf{I}_{d-1} & \begin{array}{c} -q_1/q_d \\ \vdots \\ -q_{d-1}/q_d \end{array} \end{array} \right) \in \mathbb{R}^{d \times d},$$

where  $q_0, q_1, \dots, q_d$  are the coefficients of  $q$  so that  $q(x) = \sum_{k=0}^d q_k x^k$ , and  $\mathbf{0}_m \in \mathbb{R}^m$  denotes the vector of zeros, while  $\mathbf{I}_m \in \mathbb{R}^{m \times m}$  denotes the identity matrix.

3: For  $i = 1, \dots, d$ , set the decay parameters  $\hat{\lambda}_i = 1/\nu_i$  and

$$\hat{\alpha}_i = \frac{\prod_{j=1}^d \hat{\lambda}_i - \lambda_j}{\prod_{j \neq i} \hat{\lambda}_i - \hat{\lambda}_j}.$$

4: **Return**  $\hat{\boldsymbol{\alpha}} = (\alpha_1, \dots, \alpha_d)$  and  $\hat{\boldsymbol{\lambda}} = (\lambda_1, \dots, \lambda_d)$ .

---

**Corollary 17.** Given parameters  $\boldsymbol{\alpha} \in \mathbb{R}_+^d$  and  $\boldsymbol{\lambda} \in (0, 1)^d$  such that  $\lambda_i$ 's are distinct and  $\sum_{i=1}^d \alpha_i/\lambda_i \neq 1$ .

1. Then, Algorithm 1 returns parameters  $\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\lambda}}$  such that  $\text{BLT}_n(\boldsymbol{\alpha}, \boldsymbol{\lambda})^{-1} = \text{BLT}_n(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\lambda}})$  for all  $n > 0$ . Furthermore, the map  $(\boldsymbol{\alpha}, \boldsymbol{\lambda}) \mapsto (\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\lambda}})$  is continuously differentiable.

*Proof.* The correctness of Algorithm 1 follows from Theorem 1, Propositions 8 and 15, and Lemma 3. We get the following composition of continuously differentiable functions to get from  $\boldsymbol{\alpha}, \boldsymbol{\lambda}$  to  $\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\lambda}}$ :

- $\boldsymbol{\alpha}, \boldsymbol{\lambda}$  to the coefficients of the polynomials  $r(x), p(x)$ , and  $q(x) = p(x) + x r(x)$ ;
- the coefficients of the degree- $d$  polynomial  $q(x)$  to its roots  $\nu_1, \dots, \nu_d$  (which are unique due to Property 11, and thus this map is continuously differentiable due to Lemma 16);
- $\nu_1, \dots, \nu_d$  and the coefficients of the polynomial  $p$  to  $\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\lambda}}$  as per Line 3 of Algorithm 1.

Thus, As a composition of continuously differentiable functions, the map  $(\boldsymbol{\alpha}, \boldsymbol{\lambda}) \mapsto (\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\lambda}})$  is also continuously differentiable.  $\square$

**Compatibility with Automatic Differentiation** The max loss from Eq. (4) is also a differentiable function of the BLT parameters  $\boldsymbol{\alpha}, \boldsymbol{\lambda}$ , thanks to Corollary 17. We can thus optimize  $\boldsymbol{\alpha}, \boldsymbol{\lambda}$  to minimize the max loss (4) using first-order optimization, provided we can find the gradients of the loss w.r.t.  $\boldsymbol{\alpha}$  and  $\boldsymbol{\lambda}$ . This can be achieved with automatic differentiation packages, including

JAX and PyTorch which support hardware accelerators like GPUs. Indeed, the only non-trivial operations (excluding addition, subtraction, multiplication, division) used to obtain the inverse parameters is the eigenvalue computation, and this function is differentiable when the eigenvalues are unique [16].

## 7 Discussion and Open Problems

We give an inversion theorem for a family of Buffered Linear Toeplitz (BLT) matrices, a family of parameterized lower-triangular and Toeplitz matrices introduced by Dvijotham et al. [1] for streaming differential privacy with correlated noise. The key contribution is proving that the inverse of a BLT matrix is also a BLT matrix, deriving the parameters of this inverse. Specifically, we show that under certain conditions on the original BLT parameters, the inverse BLT parameters exhibit desirable properties for differential privacy applications. Furthermore, we provide a differentiable algorithm for computing the inverse BLT parameters in  $O(d^3)$  time, enabling the optimization of BLT mechanisms for private learning and estimation.

There are several interesting open problems in this space. The first one is to find the largest class of (BLT, inverse-BLT) systems that admit an inversion theorem such as Theorem 1, or an equivalence theorem such as Theorem 2. Moreover, we observe theoretically (from the construction of [1]) and empirically that most practically relevant BLTs (in the context of streaming differential privacy) seem to satisfy the constraint  $\sum_{i=1}^d \alpha_i / \lambda_i < 1$  (or that such a constraint does not hurt). This leads to a practical question: what is the best set of BLTs to optimize over?

The BLT class as introduced by Dvijotham et al. [1, Sec 1.2] is more general than the parameterization we give in Eq. (1) — they allow BLTs to be defined by Toeplitz coefficients given by an arbitrary order- $d$  linear recurrence, or equivalently, an arbitrary degree  $d$  rational generating function, while the more restricted class we consider only captures rational generating functions with distinct roots and (except in degenerate cases) equal degree in the numerator and denominator. It is an interesting open question whether the generalization to arbitrary linear recurrences yields practical benefit.

Here, we have some evidence in the affirmative, in that banded lower-triangular Toeplitz matrices are in fact such BLTs with a trivial recurrence (equivalently: a polynomial ordinary generating function). Such matrices have proved useful in DP with multiple participations and/or privacy amplification via randomization of the data order [7, 9, 17]. On the other hand, banded matrices are generally straightforward to reason about without the machinery of general BLTs or rational generating functions, so showing a strict improvement from this generalization remains an interesting open problem.

## References

- [1] Krishnamurthy Dvijotham, H. Brendan McMahan, Krishna Pillutla, Thomas Steinke, and Abhradeep Thakurta. Efficient and Near-Optimal Noise Generation for Streaming Differential Privacy. In *IEEE Symposium on Foundations of Computer Science (FOCS)*, 2024.
- [2] Chao Li, Gerome Miklau, Michael Hay, Andrew McGregor, and Vibhor Rastogi. The matrix mechanism: optimizing linear counting queries under differential privacy. *The VLDB journal*, 24:757–781, 2015.

- [3] Aleksandar Nikolov, Kunal Talwar, and Li Zhang. The Geometry of Differential Privacy: the Sparse and Approximate Cases. *SIAM Journal on Computing*, 45(2):575–616, 2016.
- [4] Peter Kairouz, Brendan McMahan, Shuang Song, Om Thakkar, Abhradeep Thakurta, and Zheng Xu. Practical and Private (Deep) Learning without Sampling or Shuffling. In *International Conference on Machine Learning*, pages 5213–5225. PMLR, 2021.
- [5] Sergey Denisov, H Brendan McMahan, John Rush, Adam Smith, and Abhradeep Guha Thakurta. Improved Differential Privacy for SGD via Optimal Private Linear Operators on Adaptive Streams. In *Advances in Neural Information Processing Systems*, volume 35, pages 5910–5924, 2022.
- [6] Mark Bun and Thomas Steinke. Concentrated differential privacy: Simplifications, extensions, and lower bounds. In Martin Hirt and Adam Smith, editors, *Theory of Cryptography*, pages 635–658, Berlin, Heidelberg, 2016. Springer Berlin Heidelberg. ISBN 978-3-662-53641-4.
- [7] Christopher A Choquette-Choo, Arun Ganesh, Ryan McKenna, H Brendan McMahan, John Rush, Abhradeep Guha Thakurta, and Zheng Xu. (Amplified) Banded Matrix Factorization: A unified approach to private training. *Advances in Neural Information Processing Systems*, 36, 2023.
- [8] Hendrik Fichtenberger, Monika Henzinger, and Jalaj Upadhyay. Constant Matters: Fine-grained Error Bound on Differentially Private Continual Observation. In *International Conference on Machine Learning*, 2023.
- [9] Ryan McKenna. Scaling up the Banded Matrix Factorization Mechanism for Differentially Private ML. *arXiv preprint arXiv:2405.15913*, 2024.
- [10] Monika Henzinger and Jalaj Upadhyay. Improved Differentially Private Continual Observation Using Group Algebra. In *Proceedings of the 2025 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. SIAM, 2025.
- [11] H. Brendan McMahan, Zheng Xu, and Yanxiang Zhang. A Hassle-free Algorithm for Private Learning in Practice: Don’t Use Tree Aggregation, Use BLTs. In *EMNLP: Industry Track*, 2024.
- [12] Raymond G. Ayoub. Paolo Ruffini’s Contributions to the Quintic. *Archive for History of Exact Sciences*, 23(3):253–277, 1980. ISSN 00039519, 14320657. URL <http://www.jstor.org/stable/41133596>.
- [13] Paul Ramond. Abel–Ruffini’s Theorem: Complex but Not Complicated. *The American Mathematical Monthly*, 129(3):231–245, 2022.
- [14] MathWords. Polynomial roots - MATLAB roots. <https://www.mathworks.com/help/matlab/ref/roots.html>.
- [15] Roger A Horn and Charles R Johnson. *Matrix Analysis*. Cambridge University Press, 2012.
- [16] Tosio Kato. *Perturbation Theory for Linear Operators*, volume 132. Springer Science & Business Media, 2013.

- [17] Christopher A. Choquette-Choo, Arun Ganesh, Thomas Steinke, and Abhradeep Guha Thakurta. Privacy Amplification for Matrix Mechanisms. In *The Twelfth International Conference on Learning Representations*, 2024. URL <https://openreview.net/forum?id=xUzWmFdg1P>.

## A Notation Summary

**Notation Table** We summarize our main notation in Table 1.

**Generation Function Summary** The following lemma summarizes the key results of the generating function that we use throughout.

**Lemma 18.** *We show two characterizations of the generating functions of BLTs.*

- For non-zero scale parameters  $\alpha \in \mathbb{R}_{++}^d$  and decay parameters  $\lambda \in (0, 1)^d$ , we define the polynomials

$$p(x) := \prod_{i=1}^d (1 - \lambda_i x) \quad \text{and} \quad r(x) := \sum_{i=1}^d \frac{\alpha_i p(x)}{1 - \lambda_i x}.$$

Then, there exist  $\hat{\alpha} \in \mathbb{R}^d, \hat{\lambda} \in \mathbb{R}^d$  such that the statements below hold.

- Given non-zero distinct decay parameters  $\lambda, \hat{\lambda} \in (\mathbb{R} \setminus \{0\})^d$ , we define the polynomials

$$p(x) := \prod_{i=1}^d (1 - \lambda_i x) \quad \text{and} \quad q(x) := \prod_{i=1}^D (1 - \hat{\lambda}_i x).$$

Assuming  $\lambda_i \neq \hat{\lambda}_j$  for all  $i, j \in [d]$ , there exist  $\alpha, \hat{\alpha} \in \mathbb{R}^d$  such that the statements below hold.

Then, in either of the above scenarios, the following equalities hold:

$$p(x) = \prod_{i=1}^d (1 - \lambda_i x) = \frac{(-1)^d}{M} \prod_{i=1}^d (x - \mu_i) \quad \text{of degree } d \text{ with roots } \mu_i = \lambda_i^{-1} \text{ and } M = \prod_{i=1}^d \mu_i$$

$$r(x) = \sum_{i=1}^d \frac{\alpha_i p(x)}{1 - \lambda_i x} \quad \text{of degree } d - 1, \text{ and}$$

$$q(x) = p(x) + xr(x) = \prod_{i=1}^d (1 - \hat{\lambda}_i x) \quad \text{of degree } D \leq d \text{ with roots } \nu_i = \hat{\lambda}_i^{-1} \text{ for } i \in [D].$$

Further,  $f(x)$  and  $\hat{f}(x)$  are the ordinary generating functions for  $\text{BLT}(\alpha, \lambda)$  and  $\text{BLT}(\hat{\alpha}, \hat{\lambda}) = \text{BLT}(\alpha, \lambda)^{-1}$ :

$$f(x) = 1 + x \frac{r(x)}{p(x)} = \frac{p(x) + xr(x)}{p(x)} = \frac{q(x)}{p(x)} = 1 + \sum_{i=1}^d \frac{\alpha_i x}{1 - \lambda_i x}, \quad (20)$$

$$\hat{f}(x) = \frac{1}{f(x)} = 1 + x \frac{-r(x)}{q(x)} = \frac{p(x)}{p(x) + xr(x)} = \frac{p(x)}{q(x)} = 1 + \sum_{i=1}^d \frac{\hat{\alpha}_i x}{1 - \hat{\lambda}_i x}. \quad (21)$$

<b>Symbol</b>	<b>Meaning</b>
$\mathbf{C}[j, k]$	The $(j, k)^{\text{th}}$ entry of the matrix $\mathbf{C}$ .
$\mathbf{Z} \sim \mathcal{N}_{n \times m}(0, \sigma^2)$	A random matrix $\mathbf{Z} \in \mathbb{R}^{n \times m}$ whose entries are i.i.d. $\mathcal{N}(0, \sigma^2)$ .
$d$	Number of buffers (positive integer)
$[d]$	The set $\{1, 2, \dots, d\}$ .
$\boldsymbol{\alpha} \in \mathbb{R}^d$	Scale parameters of the BLT. We assume $\alpha_i \geq 0$ and $\sum_{i=1}^d \alpha_i < 1$
$\boldsymbol{\lambda} \in \mathbb{R}^d$	Decay parameters of the BLT. We assume distinct $\lambda_i \in (0, 1)$ for each $i$
$\text{BLT}(\boldsymbol{\alpha}, \boldsymbol{\lambda})$	A semi-infinite lower triangular Toeplitz matrix whose first column is given by $1, \sum_{i=1}^d \alpha_i, \sum_{i=1}^d \alpha_i \lambda_i, \sum_{i=1}^d \alpha_i \lambda_i^2, \sum_{i=1}^d \alpha_i \lambda_i^3, \dots$
$\text{BLT}_n(\boldsymbol{\alpha}, \boldsymbol{\lambda})$	An $n \times n$ lower triangular and Toeplitz matrix which is the principal sub-matrix of $\text{BLT}(\boldsymbol{\alpha}, \boldsymbol{\lambda})$
$\hat{\boldsymbol{\alpha}} \in \mathbb{R}^d, \hat{\boldsymbol{\lambda}}^d$	Scale and decay parameters such that $\text{BLT}(\boldsymbol{\alpha}, \boldsymbol{\lambda})^{-1} = \text{BLT}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\lambda}})$ (whose existence is posited by Theorem 1)
$p(x)$	The degree- $d$ polynomial $\prod_{i=1}^d (1 - \lambda_i x)$ ; $\lambda_i$ 's are assumed distinct throughout
$r(x)$	The degree- $(d-1)$ polynomial $\sum_{i=1}^d \frac{\alpha_i p(x)}{1 - \lambda_i x}$
$q(x)$	The polynomial $q(x) = p(x) + x r(x)$ ; its degree- $D$ can be $d-1$ or $d$
$D$	Degree of the polynomial $q$
$\nu_1, \dots, \nu_D$	Roots of $q$ ; the decay parameter $\hat{\boldsymbol{\lambda}}$ of the inverse BLT satisfies $\hat{\lambda}_i = \nu_i^{-1}$ for $i \in [D]$
$f$	Generating function of the first column of $\text{BLT}(\boldsymbol{\alpha}, \boldsymbol{\lambda})$ . It satisfies $f(x) = 1 + x \frac{r(x)}{p(x)} = \frac{q(x)}{p(x)} = 1 + \sum_{i=1}^d \frac{\alpha_i x}{1 - \lambda_i x}$
$\hat{f}$	Generating function of the first column of $\text{BLT}(\boldsymbol{\alpha}, \boldsymbol{\lambda})^{-1} = \text{BLT}(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\lambda}})$ . It satisfies $\hat{f}(x) = 1 + x \frac{-r(x)}{q(x)} = \frac{p(x)}{q(x)} = 1 + \sum_{i=1}^d \frac{\hat{\alpha}_i x}{1 - \hat{\lambda}_i x}$
$\mu_1, \dots, \mu_d$	Roots of the polynomial $p(x)$ ; satisfies $\mu_i = \lambda_i^{-1}$ and sorted in ascending order
$M$	Shorthand for $\prod_{i=1}^d \mu_i$
$\beta_1, \dots, \beta_d$	Constants that satisfy $\beta_i = r(\mu_i)$

Table 1: Summary of main notation. Matrices and vectors are denoted in boldface.

## B More Illustrations and Details

We give examples illustrating the behaviors of the polynomials  $r, p, q$  defined in Lemma 4 for degree  $d = 5$  in Figures 1 and 2 and for degree  $d = 4$  in Figures 3 and 4. They use the following BLT parameters:

- Figure 1 and the top row of Figure 2:  $\alpha = (0.2, 0.15, 0.1, 0.1, 0.1)$  and  $\lambda = (0.9, 0.8, 0.7, 0.6, 0.5)$ . We have  $\sum_{i=1}^d \alpha_i/\lambda_i \approx 0.919 < 1$ .
- Bottom row of Figure 2:  $\alpha = (0.2, 0.15, 0.2, 0.2, 0.2)$  and  $\lambda = (0.9, 0.8, 0.7, 0.6, 0.5)$ . We have  $\sum_{i=1}^d \alpha_i = 0.95 < 1$  and  $\sum_{i=1}^d \alpha_i/\lambda_i \approx 1.43 > 1$ .
- Figure 3 and the top row of Figure 4:  $\alpha = (0.25, 0.2, 0.15, 0.1)$  and  $\lambda = (0.9, 0.8, 0.7, 0.6)$ . We have  $\sum_{i=1}^d \alpha_i/\lambda_i \approx 0.909 < 1$ .
- Bottom row of Figure 4:  $\alpha = (0.24, 0.24, 0.24, 0.24)$  and  $\lambda = (0.9, 0.8, 0.7, 0.6)$ . We have  $\sum_{i=1}^d \alpha_i = 0.96 < 1$  and  $\sum_{i=1}^d \alpha_i/\lambda_i \approx 1.31 > 1$ .

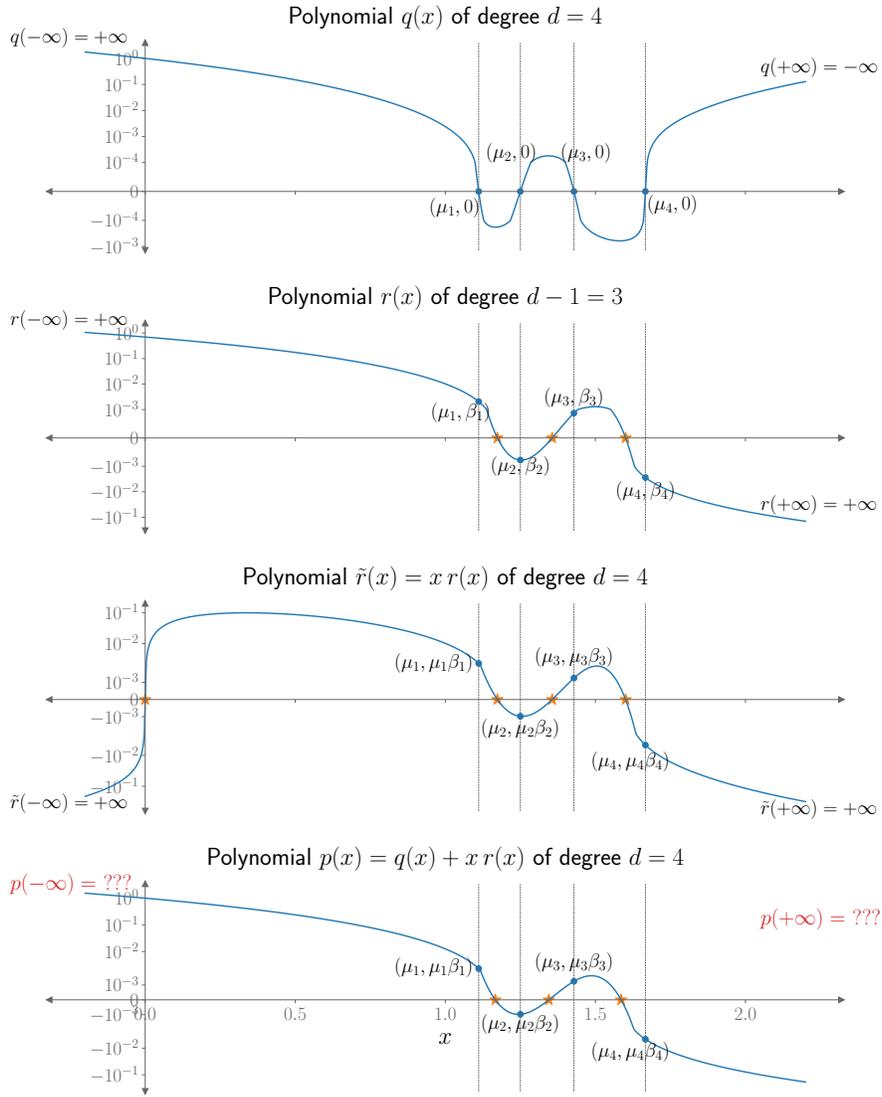


Figure 3: Illustrations of the polynomials  $r, p, q$  for  $d = 4$  in symmetrical log scale. This is the counterpart of Figure 1 for even degree  $d$ .

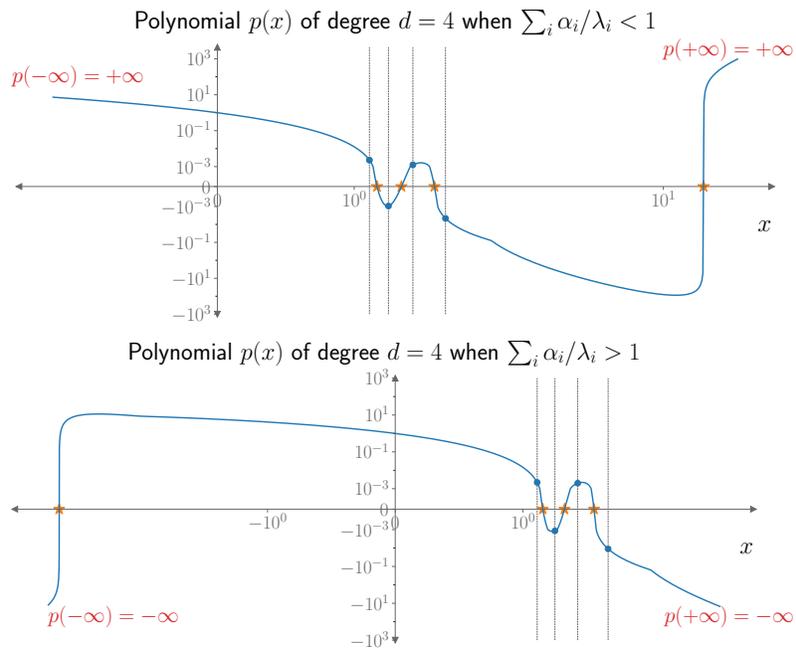


Figure 4: The counterpart of Figure 2 for even degree: this plot shows examples for  $d = 4$  and is continued from Figure 3, which shows  $d - 1$  roots of  $q(x) = p(x) + x r(x)$ . This figure illustrates how the final  $d^{\text{th}}$  root of  $q(x)$  depends on the BLT parameters  $\alpha, \lambda$ . As previously, the dotted lines denote the roots  $\mu_1, \dots, \mu_d$  of  $p(x)$  (where  $\mu_i = 1/\lambda_i$ ) and the orange star denotes the roots of  $q(x)$ .