Fast and Flexible Temporal Point Processes with Triangular Maps

Oleksandr Shchur, Nicholas Gao, Marin Biloš, Stephan Günnemann

- Introduction
- Defining temporal point processes using triangular maps
- Differentiable sampling-based losses
- Experiments

Challenges

The existing TPP models resort to generating the samples one by one because of the sequential dependency, so they can't benefit from the parallelism of modern hardware.

Contributions

The paper's contributions are:

- A new parametrization for several classic TPPs was proposed.
- TriTPP — a new class of non-recurrent TPPs was proposed.
- A differentiable relaxation for non-differentiable sampling-based TPP losses was derived.

- Introduction
- Defining temporal point processes using triangular maps
- Differentiable sampling-based losses
- Experiments

Temporal point processes (TPP)

$p(t)$:density of a point process, also called likelihood.
$\boldsymbol{t} = (t_1, t_2, \ldots, t_N)$: a variable-length sequence of strictly increasing arrival times
$\lambda^*(t) := \lambda(t|\mathrm{H}_t)$:conditional intensity function
$\Lambda^*(t) := \Lambda(t|\mathrm{H}_t) = \int_0^t \lambda^*(u)du$:cumulative conditional intensity function

$$\begin{aligned} p(t) &= \left(\prod_{i=1}^{N} \lambda^*(t_i)\right) \exp\left(-\int_0^T \lambda^*(u)du\right) \\ &= \left(\prod_{i=1}^{N} \frac{\partial}{\partial t_i}\Lambda^*(t_i)\right) \exp\left(-\Lambda^*(T)\right) \end{aligned} \tag{1}$$

Triangular maps

$\boldsymbol{F} = (f_1, \ldots, f_N) : \mathbb{R}^N \to \mathbb{R}^N$:an increasing differentiable triangular map that pushes forward p into $\widetilde{p}$. If there is a random variable $x \sim p$, then $\boldsymbol{z} := \boldsymbol{F}(x)$ with a density $\widetilde{p}$.
$\det \mathrm{J}_{\boldsymbol{F}}(x)$:the Jacobian determinant of $\boldsymbol{F}$ at x.

$$\begin{aligned} p(x) &= \left|\det \mathrm{J}_{\boldsymbol{F}}(x)\right| \, \widetilde{p}(\boldsymbol{F}(x)) \\ &= \left(\prod_{i=1}^{N} \frac{\partial}{\partial x_i} f_i(x_1, \ldots, x_N)\right) \widetilde{p}(\boldsymbol{F}(x)) \end{aligned} \tag{2}$$

Defining temporal point processes using triangular maps

$$p(t) = \left(\prod_{i=1}^{N} \frac{\partial}{\partial t_i} \Lambda^*(t_i)\right) \exp\left(-\Lambda^*(T)\right)$$

$$p(x) = \left(\prod_{i=1}^{N} \frac{\partial}{\partial x_i} f_i(x_1, \ldots, x_N)\right) \widetilde{p}(\boldsymbol{F}(x))$$

Defining temporal point processes using triangular maps

$$p(t) = \left(\prod_{i=1}^{N} \frac{\partial}{\partial t_i} \Lambda^*(t_i) \right) \exp\left(-\Lambda^*(T)\right)$$

$$p(x) = \left(\prod_{i=1}^{N} \frac{\partial}{\partial x_i} f_i(x_1, \dots, x_N) \right) \widetilde{p}(\boldsymbol{F}(x))$$

- $\boldsymbol{t} = (t_1, \dots, t_N)$ a realization of a TPP on $[0, T]$ with compensator Λ^* (i.e. with density p(t)).
- $\boldsymbol{z} = (\Lambda^*(t_1), \dots, \Lambda^*(t_N))$ a realization of a homogeneous Poisson process (HPP) with unit rate on the interval $[0, \Lambda^*(T)]$
- $\boldsymbol{F} = (f_1, \dots, f_N) : \boldsymbol{t} \to \boldsymbol{z}$
- $f_i(t) = \Lambda(t_i | t_1, \dots, t_{i-1})$
- $\widetilde{p}(z) = \widetilde{p}(\boldsymbol{F}(t)) = \exp(-\Lambda^*(T))$

Sampling

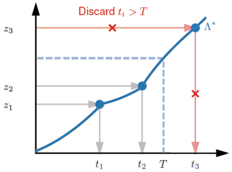

Figure 2: Sampling is done by applying $\boldsymbol{F}^{-1}$ to a sample $\boldsymbol{z}$ from a HPP with unit rate.

- draw $\boldsymbol{z}$ from an HPP on $[0, \Lambda^*(T)]$

Sampling

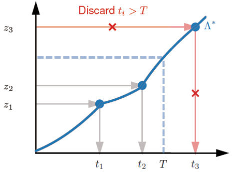

Figure 2: Sampling is done by applying $\boldsymbol{F}^{-1}$ to a sample $\boldsymbol{z}$ from a HPP with unit rate.

- draw $\boldsymbol{z}$ from an HPP on $[0, \Lambda^*(T)]$
- apply the inverse map $\boldsymbol{t} = \boldsymbol{F}^{-1}(\boldsymbol{z})$

Sampling

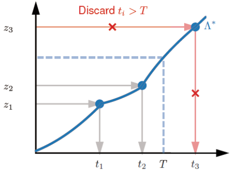

Figure 2: Sampling is done by applying $\boldsymbol{F}^{-1}$ to a sample $\boldsymbol{z}$ from a HPP with unit rate.

- draw $\boldsymbol{z}$ from an HPP on $[0, \Lambda^*(T)]$
- apply the inverse map $\boldsymbol{t} = \boldsymbol{F}^{-1}(\boldsymbol{z})$
- discard the points $t_i > T$

Sampling

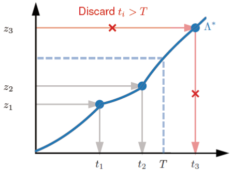

Figure 2: Sampling is done by applying $\boldsymbol{F}^{-1}$ to a sample $\boldsymbol{z}$ from a HPP with unit rate.

- draw $\boldsymbol{z}$ from an HPP on $[0, \Lambda^*(T)]$
- apply the inverse map $\boldsymbol{t} = \boldsymbol{F}^{-1}(\boldsymbol{z})$
- discard the points $t_i > T$
- get the samples $(t_1, \ldots, t_N)$ from a TPP density $p(t)$

Fast temporal point process model

Modulated renewal process (MRP) generalizes both inhomogeneous Poisson and renewal processes.

$$\Lambda(t|\mathrm{H}_t) = \Phi(\Lambda(t) - \Lambda(t_i)) + \sum_{j=1}^{i} \Phi(\Lambda(t_j) - \Lambda(t_{j-1})) \quad (3)$$

$$\boldsymbol{F} = \boldsymbol{C} \circ \boldsymbol{\Phi} \circ \boldsymbol{D} \circ \boldsymbol{\Lambda} \quad (4)$$

- $\boldsymbol{C}$:The $N \times N$ cumulative sum matrix,$C_{ij} = \begin{cases} 1, & i \leq j, \\ 0, & else \end{cases}$
- $\boldsymbol{D} \equiv \boldsymbol{C}^{-1}$:The $N \times N$ difference matrix,$D_{ij} = \begin{cases} 1, & i = j, \\ -1, & i = j+1, \\ 0, & else \end{cases}$
- $\boldsymbol{\Phi}, \boldsymbol{\Lambda}$ applies Φ, Λ elementwise, respectively

Fast temporal point process model

$$\boldsymbol{F} = \boldsymbol{C} \circ \boldsymbol{\Phi} \circ \boldsymbol{D} \circ \boldsymbol{\Lambda}$$

- $\boldsymbol{t} = (t_1, \ldots, t_N)$

Fast temporal point process model

$$\boldsymbol{F} = \boldsymbol{C} \circ \boldsymbol{\Phi} \circ \boldsymbol{D} \circ \boldsymbol{\Lambda}$$

- $\boldsymbol{t} = (t_1, \ldots, t_N)$
- $\boldsymbol{\Lambda} :\to (\Lambda(t_1), \ldots, \Lambda(t_N))$

Fast temporal point process model

$$\boldsymbol{F} = \boldsymbol{C} \circ \boldsymbol{\Phi} \circ \boldsymbol{D} \circ \boldsymbol{\Lambda}$$

- $\boldsymbol{t} = (t_1, \ldots, t_N)$
- $\boldsymbol{\Lambda} :\to (\Lambda(t_1), \ldots, \Lambda(t_N))$
- $\boldsymbol{D} :\to (\Lambda(t_1), \Lambda(t_2) - \Lambda(t_1), \ldots, \Lambda(t_N) - \Lambda(t_{N-1}))$

Fast temporal point process model

$$\boldsymbol{F} = \boldsymbol{C} \circ \boldsymbol{\Phi} \circ \boldsymbol{D} \circ \boldsymbol{\Lambda}$$

- $\boldsymbol{t} = (t_1, \ldots, t_N)$
- $\boldsymbol{\Lambda} :\to (\Lambda(t_1), \ldots, \Lambda(t_N))$
- $\boldsymbol{D} :\to (\Lambda(t_1), \Lambda(t_2) - \Lambda(t_1), \ldots, \Lambda(t_N) - \Lambda(t_{N-1}))$
- $\boldsymbol{\Phi} :\to (\Phi(\Lambda(t_1)), \Phi(\Lambda(t_2) - \Lambda(t_1)), \ldots, \Phi(\Lambda(t_N) - \Lambda(t_{N-1})))$

Fast temporal point process model

$$\boldsymbol{F} = \boldsymbol{C} \circ \boldsymbol{\Phi} \circ \boldsymbol{D} \circ \boldsymbol{\Lambda}$$

- $\boldsymbol{t} = (t_1, \dots, t_N)$
- $\boldsymbol{\Lambda} :\to (\Lambda(t_1), \dots, \Lambda(t_N))$
- $\boldsymbol{D} :\to (\Lambda(t_1), \Lambda(t_2) - \Lambda(t_1), \dots, \Lambda(t_N) - \Lambda(t_{N-1}))$
- $\boldsymbol{\Phi} :\to (\Phi(\Lambda(t_1)), \Phi(\Lambda(t_2) - \Lambda(t_1)), \dots, \Phi(\Lambda(t_N) - \Lambda(t_{N-1})))$
- $\boldsymbol{C} :\to$ $(\Phi(\Lambda(t_1)), \Phi(\Lambda(t_1)) + \Phi(\Lambda(t_2) - \Lambda(t_1)), \dots, \sum_{j=1}^{N} \Phi(\Lambda(t_j) - \Lambda(t_{j-1})))$

$\boldsymbol{C}, \boldsymbol{\Phi}, \boldsymbol{D}, \boldsymbol{\Lambda}$ can be applied in $O(N)$ parallel operations.

TriTPP

To make the map $\boldsymbol{F}$ more expressive, we can add learnable lower-triangular matrices into the composition. $\boldsymbol{B}_l$ is a block-diagonal matrix, where each block is a repeated $H \times H$ lower-triangular matrix. Computing $\boldsymbol{B}_l^{-1}$ takes $O(H^2)$, and multiplication by $\boldsymbol{B}_l$ or $\boldsymbol{B}_l^{-1}$ can be done in $O(NH)$ in parallel.

$$\boldsymbol{F} = \boldsymbol{C} \circ \Phi_2 \circ \boldsymbol{B}_L \circ \ldots \circ \boldsymbol{B}_1 \circ \Phi_1 \circ \boldsymbol{D} \circ \boldsymbol{\Lambda} \quad (5)$$

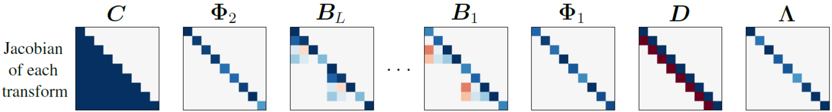

Figure 3: TriTPP defines an expressive map $\boldsymbol{F}$ as a composition of easy-to-invert transformations.

- Introduction
- Defining temporal point processes using triangular maps
- Differentiable sampling-based losses
- Experiments

Entropy maximization

- $p_\lambda(\boldsymbol{t})$:a homogeneous Poisson process on $[0, T]$ with rate $\lambda > 0$
- $-\mathbb{E}_p[\log p_\lambda(\boldsymbol{t})]$:objective function
- draw a sequence $\boldsymbol{z} = (z_1, z_2, \ldots)$ from a HPP with unit rate
- apply the inverse map $\boldsymbol{t} = \boldsymbol{F}_\lambda^{-1}(\boldsymbol{z}) = \frac{1}{\lambda}\boldsymbol{z}$

$$-\mathbb{E}_p[\log p_\lambda(\boldsymbol{t})] \approx \lambda T - \sum_{i=1}^{\infty} 1(t_i \leq T) log \lambda = \lambda T - \sum_{i=1}^{\infty} 1(\frac{1}{\lambda} z_i \leq T) log \lambda \tag{6}$$

where $1(x)$ is the indicator function, $1(x) = \begin{cases} 1, & x \text{ is true}, \\ 0, & else \end{cases}$

Relaxation

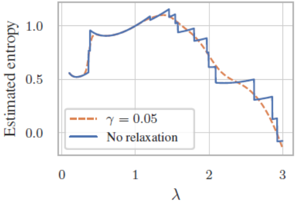

Figure 4: Monte Carlo estimate of the entropy.

$$-\mathbb{E}_p[\log p_\lambda(\boldsymbol{t})] \approx \lambda T - \sum_{i=1}^{\infty} 1(\frac{1}{\lambda} z_i \leq T) log \lambda$$

We can see that the Equation 6 is not continuous at point $\lambda = \frac{1}{T} z_i$.

Relaxation

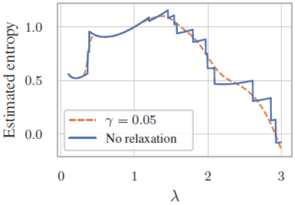

Figure 4: Monte Carlo estimate of the entropy.

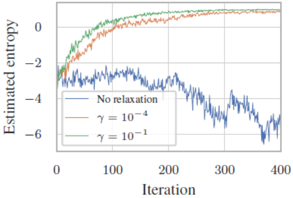

Figure 5: Maximizing the entropy with different values of γ.

$$1(t_i \leq T) \approx \sigma_\gamma(T - t_i) \tag{7}$$

where $\sigma_\gamma(x) = \frac{1}{1+\exp(\frac{-x}{\gamma})}$ is the sigmoid function.

- Introduction
- Defining temporal point processes using triangular maps
- Differentiable sampling-based losses
- Experiments

Scalability

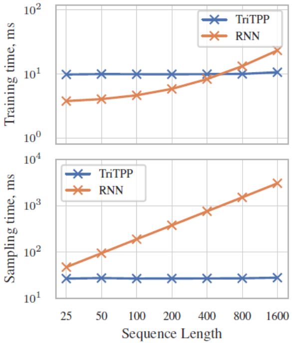

Figure 7: Scalability analysis. Standard devs. are below 1ms.

Density estimation

Table 1: Average test set NLL on synthetic and real-world datasets (lower is better). Best NLL in **bold**, second best underlined. Results with standard deviations can be found in Appendix F.1.

	Hawkes1	Hawkes2	SC	IPP	MRP	RP	PUBG	Reddit-C	Reddit-S	Taxi	Twitter	Yelp1	Yelp2
IPP	1.06	1.03	1.00	**0.71**	0.70	0.89	-0.06	-1.59	-4.08	**-0.68**	1.60	<u>0.62</u>	-0.05
RP	0.65	0.08	0.94	0.85	0.68	**0.24**	0.12	-2.08	-4.00	-0.58	1.20	0.67	-0.02
MRP	0.65	0.07	0.93	**0.71**	0.36	0.25	-0.83	-2.13	-4.38	**-0.68**	1.23	**0.61**	**-0.10**
Hawkes	**0.51**	0.06	1.00	0.86	0.98	0.39	0.11	**-2.40**	-4.19	-0.64	**1.04**	0.69	0.01
RNN	<u>0.52</u>	**-0.03**	**0.79**	0.73	<u>0.37</u>	**0.24**	<u>-1.96</u>	**-2.40**	**-4.89**	-0.66	1.08	0.67	-0.08
TriTPP	0.56	<u>0.00</u>	<u>0.83</u>	**0.71**	**0.35**	**0.24**	**-2.41**	-2.36	<u>-4.49</u>	-0.67	<u>1.06</u>	0.64	<u>-0.09</u>

Table 2: MMD between the hold-out test set and the generated samples (lower is better).

	Hawkes1	Hawkes2	SC	IPP	MRP	RP	PUBG	Reddit-C	Reddit-S	Taxi	Twitter	Yelp1	Yelp2
IPP	0.08	0.09	0.58	**0.02**	0.15	0.07	**0.01**	0.10	0.21	0.10	0.16	0.15	<u>0.16</u>
RP	0.06	0.06	1.13	0.34	1.24	**0.01**	0.46	0.07	0.18	0.57	0.14	0.16	0.23
MRP	0.05	0.06	0.50	**0.02**	<u>0.11</u>	0.02	<u>0.12</u>	0.09	0.20	<u>0.09</u>	0.13	<u>0.13</u>	<u>0.16</u>
Hawkes	<u>0.02</u>	0.04	0.58	0.36	0.65	0.05	0.16	**0.04**	0.35	0.20	0.20	0.20	0.32
RNN	**0.01**	**0.02**	**0.19**	0.09	0.17	**0.01**	0.23	**0.04**	**0.09**	0.13	**0.08**	0.19	0.18
TriTPP	0.03	<u>0.03</u>	<u>0.23</u>	**0.02**	**0.08**	**0.01**	0.16	0.07	<u>0.16</u>	**0.08**	**0.08**	**0.12**	**0.14**