Appendix A: Deriving Conditional Intensity of MTPPs

Assume that we have total number of $N([0,T]\times\mathcal{M})$ observations in $\boldsymbol{x}$. For any given $t\in[0,T]$, we assume that n events happened before t and denote the occurrence time of the latest event as t_n. Let $\Omega=[t,t+dt)\times B(m,dm)$ where $m\in\mathcal{M}$. Let $F(t)=\mathbb{P}(x_{n+1},t_{n+1}<t|\mathcal{H}_{t_n}\cup x_n)$ be the conditional cumulative probability function, and $\mathcal{H}_{t_n}\cup x_n$ represents the history events happened up to time t_n and at t_n. Let $f(t,m)\triangleq f(t,m|\mathcal{H}_{t_n}\cup x_n)$ be the corresponding conditional probability density function of new event happening in Ω. As defined in (2), $\lambda(t,m)$ can be expressed as

$$\begin{aligned}\lambda(t,m)&=\mathbb{P}\{x_{n+1}\in\Omega|\mathcal{H}_t\}=\mathbb{P}\{x_{n+1}\in\Omega|\mathcal{H}_{t_n}\cup x_n\cup\{t_{n+1}\geq t\}\}\\&=\frac{\mathbb{P}\{x_{n+1}\in\Omega,t_{n+1}\geq t|\mathcal{H}_{t_n}\cup x_n\}}{\mathbb{P}\{t_{n+1}\geq t|\mathcal{H}_{t_n}\cup x_n\}}\\&=\frac{f(t,m)}{1-F(t)}.\end{aligned}$$

We multiply the differential of time and space $dtdm$ on both side of the equation, and integral over m

$$dt\cdot\int_{\mathcal{M}}\lambda(t,u)du=\frac{dt\cdot\int_{\mathcal{M}}f(t,u)du}{1-F(t)}=\frac{dF(t)}{1-F(t)}=-d\log\left(1-F(t)\right).$$

Hence, integrating over t on (t_n,t) leads to $F(t)=1-\exp(-\int_{t_n}^t\int_{\mathcal{M}}\lambda(\tau,u)dud\tau)$ because $F(t_n)=0$. Then we have

$$f(t,m)=\lambda(t,m)\cdot\exp\left(-\int_{t_n}^t\int_{\mathcal{M}}\lambda(\tau,u)dud\tau\right)\ ,$$

The joint p.d.f. for a realization is then, by the chain rule, $f(x_1,\ldots,x_{N([0,T]\times\mathcal{M})})=\prod_{i=1}^{N([0,T]\times\mathcal{M})}f(t_i,m_i)$. Then the log-likelihood of an observed sequence $\boldsymbol{x}$ can be written as

$$l(\boldsymbol{x})=\sum_{i=1}^{N([0,T]\times\mathcal{M})}\log\lambda(t_i,m_i)-\int_0^T\int_{\mathcal{M}}\lambda(\tau,u)dud\tau.$$

Appendix B: Deriving Log-Likelihood of MTPPs

The likelihood function is defined as:

$$\mathcal{L}=f(t_1,\ldots,t_n;m_1,\ldots,m_n)=f(t_1,m_1|\mathcal{H}_{t_1})\times\cdots\times f(t_n,m_n|\mathcal{H}_{t_n})=\prod_{i=1}^n f(t_i,m_i|\mathcal{H}_{t_i})$$

Note from the definition of conditional intensity function, we get:

$$\int_{\mathcal{M}}\lambda(t,m|\mathcal{H}_t)=\frac{\int_{\mathcal{M}}f(t,m|\mathcal{H}_t)}{1-F(t|\mathcal{H}_t)}=-\frac{\partial}{\partial t}\log(1-F(t|\mathcal{H}_{t_n})).$$

Integrating both side from t_n to T (since $\lambda(t,m|\mathcal{H}_t)$ depends on history events, so its support is $[t_n,T)$) where t_n is the last event before T. Integrate over all $\mathcal{M}$, we can get: $\int_{t_n}^T\int_{m\in\mathcal{M}}\lambda(t,m|\mathcal{H}_t)dtdm-\log\left(1-F(t,m|\mathcal{H}_{t_n})\right)$, obviously, using basic calculus we could find:

$$f(t,m|\mathcal{H}_t)=\lambda(t,m|\mathcal{H}_t)\cdot\exp\left(-\int_{t_n}^T\int_{m\in\mathcal{M}}\lambda(t,m|\mathcal{H}_t)dtdm\right).$$

Plugging in the above formula into the definition of likelihood function, we have:

$$\mathcal{L}=\prod_{i=1}^n\left(\lambda(t_i,m_i|\mathcal{H}_t)\right)\cdot\exp\left(-\int_0^T\int_{m\in\mathcal{M}}\lambda(t,m|\mathcal{H}_t)dtdm\right),$$

and the log-likelihood function of marked spatio-temporal point process can be the written as :

$$\ell=\sum_{i=1}^n\log\lambda(t_i,m_i|\mathcal{H}_t)-\int_0^T\int_{m\in\mathcal{M}}\lambda(t,m|\mathcal{H}_t)dtdm.$$

Appendix C: Proof for Proposition 1

For the notational simplicity, we denote Wx as x. First, since both K and p_ω are real-valued, it suffices to consider only the real portion of e^{ix} when invoking Theorem 1. Thus, using $\text{Re}[e^{ix}] = \text{Re}[\cos(x) + i\sin(x)] = cos(x)$, we have

$$K(x,x') = \text{Re}[K(x,x')] = \int_\Omega p_\omega(\omega)\cos(\omega^\top(x-x'))d\omega.$$

Next, we have

$$\begin{aligned}
\int_\Omega p_\omega(\omega)\cos\left(\omega^\top(x-x')\right)d\omega &\overset{(i)}{=} \int_\Omega p_\omega(\omega)\cos\left(\omega^\top(x-x')\right)d\omega + \int_\Omega\int_0^{2\pi}\frac{1}{2\pi}p_\omega(\omega)\cos\left(\omega^\top(x+x')+2u\right)dud\omega \\
&= \int_\Omega\int_0^{2\pi}\frac{1}{2\pi}p_\omega(\omega)\big[\cos\left(\omega^\top(x-x')\right)+\cos\left(\omega^\top(x+x')+2u\right)\big]dud\omega \\
&= \int_\Omega\int_0^{2\pi}\frac{1}{2\pi}p_\omega(\omega)\left[2\cos(\omega^\top x+u)\cdot\cos(\omega^\top x'+u)\right]dud\omega \\
&= \int_\Omega p_\omega(\omega)\int_0^{2\pi}\frac{1}{2\pi}\left[\sqrt{2}\cos(\omega^\top x+u)\cdot\sqrt{2}\cos(\omega^\top x'+u)\right]dud\omega \\
&= \mathbb{E}\left[\phi_\omega(x)\cdot\phi_\omega(x')\right].
\end{aligned}$$

where $\phi_\omega(x) := \sqrt{2}\cos(\omega^\top x+u)$, ω is sampled from p_ω, and u is uniformly sampled from $[0,2\pi]$. The equation (i) holds since the second term equals to 0 as shown below:

$$\int_\Omega\int_0^{2\pi}p_\omega(\omega)\cos\left(\omega^\top(x+x')+2u\right)dud\omega = \int_\Omega p_\omega(\omega)\int_0^{2\pi}\cos\left(\omega^\top(x+x')+2u\right)dud\omega = \int_\Omega p_\omega(\omega)\cdot 0\cdot d\omega = 0.$$

Therefore, we can obtain the result in Proposition 1.

Appendix D: Proof for Proposition 2

Similar to the proof in Appendix C, we denote Wx as $x\in\mathcal{X}$ for the notational simplicity. Recall that we denote R as the radius of the Euclidean ball containing $\mathcal{X}$ in Section 4.2. In the following, we first present two useful lemmas.

Lemma 1. *Assume $\mathcal{X}\subset\mathbb{R}^d$ is compact. Let R denote the radius of the Euclidean ball containing $\mathcal{X}$, then for the kernel-induced feature mapping Φ defined in (8), the following holds for any $0<r\le 2R$ and $\epsilon>0$:*

$$\mathbb{P}\left\{\sup_{x,x'\in\mathcal{X}}\left|\Phi(x)^\top\Phi(x') - K(x,x')\right|\ge\epsilon\right\}\le\ 2\mathcal{N}(2R,r)\exp\left\{-\frac{D\epsilon^2}{8}\right\}+\frac{4r\sigma_p}{\epsilon}.$$

where $\sigma_p^2 = \mathbb{E}_{\omega\sim p_\omega}[\omega^\top\omega]<\infty$ is the second moment of the Fourier features, and $\mathcal{N}(R,r)$ denotes the minimal number of balls of radius r needed to cover a ball of radius R.

Proof of Lemma 1 Now, define $\Delta = \{\delta:\delta = x-x',\ ,x,x'\in\mathcal{X}\}$ and note that Δ is contained in a ball of radius at most $2R$. Δ is a closed set since $\mathcal{X}$ is closed and thus Δ is a compact set. Define $B=\mathcal{N}(2R,r)$ the number of balls of radius r needed to cover Δ and let δ_j, for $j\in[B]$ denote the center of the covering balls. Thus, for any $\delta\in\Delta$ there exists a j such that $\delta=\delta_j+r'$ where $|r'|<r$.

Next, we define $S(\delta)=\Phi(x)^\top\Phi(x^\top)-K(x,x')$, where $\delta = x-x'$. Since S is continuously differentiable over the compact set Δ, it is L-Lipschitz with $L=\sup_{\delta\in\Delta}||\nabla S(\delta)||$. Note that if we assume $L<\epsilon/2r$ and for all $j\in[B]$ we have $|S(\delta_j)|<\epsilon/2$, then the following inequality holds for all $\delta=\delta_j+r'\in\Delta$:

$$|S(\delta)| = |S(\delta_j+r')|\le L|\delta_j-(\delta_j+r')|+|S(\delta_j)|\le rL+\frac{\epsilon}{2}<\epsilon. \tag{12}$$

The remainder of this proof bounds the probability of the events $L > \epsilon/(2r)$ and $|S(\delta_j)| \geq \epsilon/2$. Note that all following probabilities and expectations are with respect to the random variables $\omega_1, \dots, \omega_D$.

To bound the probability of the first event, we use Proposition 1 and the linearity of expectation, which implies the key fact $\mathbb{E}[\nabla(\Phi(x)^\top \Phi(x'))] = \nabla K(x, x^\top)$. We proceed with the following series of inequalities:

$$\begin{aligned}
\mathbb{E}\left[L^2\right] &= \mathbb{E}\left[\sup_{\delta\in\Delta} ||\nabla S(\delta)||^2\right] \\
&= \mathbb{E}\left[\sup_{x,x'\in\mathcal{X}} ||\nabla(\Phi(x)^\top \Phi(x')) - \nabla K(x,x')||^2\right] \\
&\overset{(i)}{\leq} 2\mathbb{E}\left[\sup_{x,x'\in\mathcal{X}} ||\nabla(\Phi(x)^\top \Phi(x'))||^2\right] + 2\sup_{x,x'\in\mathcal{X}} ||\nabla K(x,x')||^2 \\
&= 2\mathbb{E}\left[\sup_{x,x'\in\mathcal{X}} ||\nabla(\Phi(x)^\top \Phi(x'))||^2\right] + 2\sup_{x,x'\in\mathcal{X}} ||\mathbb{E}\left[\nabla(\Phi(x)^\top \Phi(x'))\right]||^2 \\
&\overset{(ii)}{\leq} 4\mathbb{E}\left[\sup_{x,x'\in\mathcal{X}} ||\nabla(\Phi(x)^\top \Phi(x'))||^2\right],
\end{aligned}$$

where the first inequality (i) holds due to the the inequality $||a+b||^2 \leq 2||a||^2 + 2||b||^2$ (which follows from Jensen's inequality) and the subadditivity of the supremum function. The second inequality (ii) also holds by Jensen's inequality (applied twice) and again the subadditivity of supremum function. Furthermore, using a sum-difference trigonometric identity and computing the gradient with respect to $\delta = x - x'$, yield the following for any $x, x' \in \mathcal{X}$:

$$\nabla(\Phi(x)^\top \Phi(x')) = \nabla\left(\frac{1}{D}\sum_{k=1}^{D}\cos(\omega_k^\top(x-x'))\right) = \frac{1}{D}\sum_{k=1}^{D}\omega_k \sin(\omega_k^\top(x-x')).$$

Combining the two previous results gives

$$\begin{aligned}
\mathbb{E}[L^2] &\leq 4\mathbb{E}\left[\sup_{x,x'\in\mathcal{X}} ||\frac{1}{D}\sum_{k=1}^{D}\omega_k \sin(\omega_k^\top(x-x'))||^2\right] \\
&\leq 4 \mathop{\mathbb{E}}_{\omega_1,\dots,\omega_D}\left[\left(\frac{1}{D}\sum_{k=1}^{D}||\omega_k||\right)^2\right] \\
&\leq 4 \mathop{\mathbb{E}}_{\omega_1,\dots,\omega_D}\left[\frac{1}{D}\sum_{k=1}^{D}||\omega_k||^2\right] = 4\mathop{\mathbb{E}}_{\omega}[||\omega||^2] = 4\sigma_p^2,
\end{aligned}$$

which follows from the triangle inequality, $|\sin(\cdot)| \leq 1$, Jensen's inequality and the fact that the ω_ks are drawn i.i.d. derive the final expression. Thus, we can bound the probability of the first event via Markov's inequality:

$$\mathbb{P}\left[L \geq \frac{\epsilon}{2r}\right] \leq \left(\frac{4r\sigma_p}{\epsilon}\right)^2. \tag{13}$$

To bound the probability of the second event, note that, by definition, $S(\delta)$ is a sum of D i.i.d. variables, each bounded in absolute value by $\frac{2}{D}$ (since, for all x and x', we have $|K(x,x')| \leq 1$ and $|\Phi(x)^\top \Phi(x')| \leq 1$), and $\mathbb{E}[S(\delta)] = 0$. Thus, by Hoeffding's inequality and the union bound, we can write

$$\mathbb{P}\left[\exists j \in [B] : |S(\delta_j)| \geq \frac{\epsilon}{2}\right] \leq \sum_{j=1}^{B}\mathbb{P}\left[|S(\delta_j)| \geq \frac{\epsilon}{2}\right] \leq 2B\exp\left(-\frac{D\epsilon^2}{8}\right). \tag{14}$$

Finally, combining (12), (13), (14), and the definition of B we have the result in Proposition 2, i.e.,

$$\mathbb{P}\left[\sup_{\delta\in\Delta}|S(\delta_j)| \geq \epsilon\right] \leq 2\mathcal{N}(2R,r)\exp\left\{-\frac{D\epsilon^2}{8}\right\} + \left(\frac{4r\sigma_p}{\epsilon}\right)^2,$$

As we can see now, a key factor in the bound of the proposition is the covering number $N(2R, r)$, which strongly depends on the dimension of the space d. In the following proof, we make this dependency explicit for one especially simple case, although similar arguments hold for more general scenarios as well.

LEMMA 2. *Let $\mathcal{X} \subset \mathbb{R}^d$ be a compact and let R denote the radius of the smallest enclosing ball. Then, the following inequality holds:*

$$\mathcal{N}(R, r) \leq \left(\frac{3R}{r}\right)^d.$$

Proof of Lemma 2 By using the volume of balls in $\mathbb{R}^d$, we already see that $R^d/(r/3)^d = (3R/r)^d$ is a trivial upper bound on the number of balls of radius $r/3$ that can be packed into a ball of radius R without intersecting. Now, consider a maximal packing of at most $(3R/r)^d$ balls of radius $r/3$ into the ball of radius R. Every point in the ball of radius R is at distance at most r from the center of at least one of the packing balls. If this were not true, we would be able to fit another ball into the packing, thereby contradicting the assumption that it is a maximal packing. Thus, if we grow the radius of the at most $(3R/r)^d$ balls to r, they will then provide a (not necessarily minimal) cover of the ball of radius R.

Finally, by combining the two previous lemmas, we can present an explicit finite sample approximation bound. We use lemma 1 in conjunction with lemma 2 with the following choice of r:

$$r = \left[\frac{2(6R)^d \exp(-\frac{D\epsilon^2}{8})}{\left(\frac{4\sigma_p}{\epsilon}\right)^2}\right]^{\frac{2}{d+2}},$$

which results in the following expression

$$\mathbb{P}\left[\sup_{\delta\in\Delta} |S(\delta)| \geq \epsilon\right] \leq 4\left(\frac{24R\sigma_p}{\epsilon}\right)^{\frac{2d}{d+2}} \exp\left(-\frac{D\epsilon^2}{4(d+2)}\right).$$

Since $32R\sigma_p/\epsilon \geq 1$, the exponent $2d/(d+2)$ can be replaced by 2, which completes the proof.

Appendix E: Proof for Proposition 3

To calculate the integral (the second term of the log-likelihood function defined in (4)), we first need consider the time and the mark of events separately, i.e., $x = [t, m]^\top$. Denote $\boldsymbol{i}$ as the imaginary unit. Hence the integral can be written as

$$\begin{aligned}
&\int_0^T \int_{\mathcal{M}} \left(\mu + \sum_{t_j < t} \widetilde{K}([t, m]^\top, [t_j, m_j]^\top)\right) dmdt \\
&= \int_0^T \int_{\mathcal{M}} \left(\mu + \sum_{t_j < t} \frac{1}{D} \sum_{k=1}^{D} e^{\boldsymbol{i}\omega_k^\top W([t,m]^\top - [t_j, m_j]^\top)}\right) dmdt \\
&= \mu T|\mathcal{M}| + \frac{1}{D} \sum_{k=1}^{D} \int_0^T \int_{\mathcal{M}} \sum_{t_j < t} e^{\boldsymbol{i}\omega_k^\top W([t,m]^\top - [t_j, m_j]^\top)} dmdt \\
&= \mu T|\mathcal{M}| + \frac{1}{D} \sum_{k=1}^{D} \sum_{i=0}^{N_T} \int_{t_i}^{t_{i+1}} \int_{\mathcal{M}} \sum_{t_j < t} e^{\boldsymbol{i}\omega_k^\top W([t,m]^\top - [t_j, m_j]^\top)} dmdt \\
&= \mu T|\mathcal{M}| + \frac{1}{D} \sum_{k=1}^{D} \sum_{i=0}^{N_T} \sum_{t_j < t_i} \int_{t_i}^{t_{i+1}} \int_{\mathcal{M}} e^{\boldsymbol{i}\omega_k^\top W([t,m]^\top - [t_j, m_j]^\top)} dmdt
\end{aligned}$$

(15)

Then the remainder of the proof calculates the integral $\int_{t_i}^{t_{i+1}}\int_{\mathcal{M}} e^{\boldsymbol{i}\omega_k^\top W([t,m]^\top - [t_j,m_j]^\top)} dm dt$. First, let the linear mapping matrix $W = [w_0|w_1|\ldots|w_d]$ be split into $d+1$ column vectors, where $w_0 \in \mathbb{R}^{r\times 1}$, $w_\ell \in \mathbb{R}^{r\times 1}, \ell = 1,\ldots,d$ correspond to the linear mappings for the time and mark subspace, respectively. Denote the matrix formed by first ℓ column vectors of matrix W as $W_{1:\ell} := [w_1|\ldots|w_\ell]$. Denote the ℓ-th dimension of $\mathcal{M}$ as $\mathcal{M}_\ell \in \mathbb{R}, \ell = 1,\ldots,d$. Assume each dimension of the mark space $\mathcal{M}_\ell, \ell = 1,\ldots,d$ is normalized to range $[a,b]$. Denote the sub-space of $\mathcal{M}$ with the first ℓ dimensions as $\mathcal{M}_{1:\ell} := \mathcal{M}_1 \times,\ldots,\times \mathcal{M}_\ell, \ell = 1,\ldots,d$. Denote the mark vector with first ℓ elements as $m_{1:\ell} = [m_1,\ldots,m_\ell]$. Then the integral can be written as

$$\begin{aligned}\int_{t_i}^{t_{i+1}}\int_{\mathcal{M}} e^{\boldsymbol{i}\omega_k^\top W([t,m]^\top - [t_j,m_j]^\top)} dm dt &= e^{-\boldsymbol{i}\omega_k^\top W[t_j,m_j]^\top}\int_{t_i}^{t_{i+1}}\int_{\mathcal{M}} e^{\boldsymbol{i}\omega_k^\top W[t,m]^\top} dm dt \\ &= e^{-\boldsymbol{i}\omega_k^\top W[t_j,m_j]^\top}\int_{t_i}^{t_{i+1}} e^{\boldsymbol{i}\omega_k^\top w_0 t} dt \int_{\mathcal{M}} e^{\boldsymbol{i}\omega_k^\top W_{1:d} m^\top} dm. \end{aligned} \tag{16}$$

To avoid notational overload, let $f_{k,\ell}(m)$ denote $e^{\boldsymbol{i}\omega_k^\top w_\ell m_\ell}$, and $F_{k,\ell}(m)$ denote $e^{\boldsymbol{i}\omega_k^\top W_{1:\ell} m_{1:\ell}^\top}$. Note that

$$\int_a^b f_{k,\ell}(m) dm_\ell = \frac{e^{\boldsymbol{i}\omega_k^\top w_\ell b} - e^{\boldsymbol{i}\omega_k^\top w_\ell a}}{\boldsymbol{i}\omega_k^\top w_\ell},$$

and $F_{k,\ell}(m) = f_{k,\ell}(m) F_{k,\ell-1}(m)$. Then $\int_{\mathcal{M}} e^{\boldsymbol{i}\omega_k^\top W_{1:d} m^\top} dm$ can be written as

$$\begin{aligned}\int_{\mathcal{M}} e^{\boldsymbol{i}\omega_k^\top W_{1:d} m^\top} dm &= \int_{\mathcal{M}} F_{k,d}(m) dm = \int_{\mathcal{M}_{1:d-1}} F_{k,d-1}(m) dm_{1:d-1} \int_a^b f_{k,d}(m) dm_d \\ &= \prod_{\ell=1}^d \left(\int_a^b f_{k,\ell}(m) dm_\ell\right) = \prod_{\ell=1}^d \left(\frac{e^{\boldsymbol{i}\omega_k^\top w_\ell b} - e^{\boldsymbol{i}\omega_k^\top w_\ell a}}{\boldsymbol{i}\omega_k^\top w_\ell}\right). \end{aligned} \tag{17}$$

Substitute (17) into (16), we have

$$\begin{aligned}& e^{-\boldsymbol{i}\omega_k^\top W[t_j,m_j]^\top}\int_{t_i}^{t_{i+1}} e^{\boldsymbol{i}\omega_k^\top w_0 t} dt \int_{\mathcal{M}} e^{\boldsymbol{i}\omega_k^\top W_{1:d} m^\top} dm \\ &= e^{-\boldsymbol{i}\omega_k^\top W[t_j,m_j]^\top}\left(\frac{e^{\boldsymbol{i}\omega_k^\top w_0 t_{i+1}} - e^{\boldsymbol{i}\omega_k^\top w_0 t_i}}{\boldsymbol{i}\omega_k^\top w_0}\right)\prod_{\ell=1}^d \left(\frac{e^{\boldsymbol{i}\omega_k^\top w_\ell b} - e^{\boldsymbol{i}\omega_k^\top w_\ell a}}{\boldsymbol{i}\omega_k^\top w_\ell}\right). \end{aligned}$$

Due to the fact that, for any $b > a$ where $a,b \in \mathbb{R}$,

$$\begin{aligned}\frac{e^{\boldsymbol{i}\omega_k^\top w_\ell b} - e^{\boldsymbol{i}\omega_k^\top w_\ell a}}{\boldsymbol{i}\omega_k^\top w_\ell} &= \frac{e^{\omega_k^\top w_\ell}}{\omega_k^\top w_\ell}\cdot\frac{e^{\boldsymbol{i}b} - e^{\boldsymbol{i}a}}{\boldsymbol{i}} \\ &= \frac{e^{\omega_k^\top w_\ell}}{\omega_k^\top w_\ell}\cdot e^{\boldsymbol{i}\frac{b+a}{2}}\cdot\frac{e^{\boldsymbol{i}\frac{b-a}{2}} - e^{\boldsymbol{i}\frac{a-b}{2}}}{\boldsymbol{i}} \\ &\overset{(i)}{=} \frac{2e^{\omega_k^\top w_\ell}}{\omega_k^\top w_\ell}\left(\cos\left(\frac{b+a}{2}\right) + \boldsymbol{i}\sin\left(\frac{b+a}{2}\right)\right)\sin\left(\frac{b-a}{2}\right) \\ &\overset{(ii)}{=} \frac{2e^{\omega_k^\top w_\ell}}{\omega_k^\top w_\ell}\cos\left(\frac{b+a}{2}\right)\sin\left(\frac{b-a}{2}\right). \end{aligned}$$

The equality (i) holds because of Euler's formula. The equality (ii) holds, since both K and p_ω are real-valued, it suffices to consider only the real portion. Let x_j denote $[t_j,m_j]^\top$ and substitute $\exp\{-\boldsymbol{i}\omega_k^\top W[t_j,m_j]^\top\}$ with $\cos(-\omega_k^\top W x_j)$. Thus, the integral (16) can be written as

$$\cos\left(-\omega_k^\top W x_j\right)\cos\left(\frac{t_{i+1}+t_i}{2}\right)\sin\left(\frac{t_{i+1}-t_i}{2}\right)\cos^d\left(\frac{b+a}{2}\right)\sin^d\left(\frac{b-a}{2}\right)\prod_{\ell=1}^{d+1}\frac{2e^{\omega_k^\top w_\ell}}{\omega_k^\top w_\ell}. \tag{18}$$

Finally, combining previous results in (15), (16), and (18) gives the result in Proposition 3, i.e.,

$$\int_{\mathcal{X}} \lambda(x|\mathcal{H}_t;\theta_0)dx = \mu T(b-a)^d + \frac{1}{D}\sum_{k=1}^{D}\sum_{i=0}^{N_T}\sum_{t_j<t_i}$$
$$\cos\left(-\omega_k^\top W x_j\right)\cos\left(\frac{t_{i+1}+t_i}{2}\right)\sin\left(\frac{t_{i+1}-t_i}{2}\right)\cos^d\left(\frac{b+a}{2}\right)\sin^d\left(\frac{b-a}{2}\right)\prod_{\ell=1}^{d+1}\frac{2e^{\omega_k^\top w_\ell}}{\omega_k^\top w_\ell}.$$