Supplementary Material

This document is the appendix of the paper "*Offline detection of change-point in the mean for stationary graph signals*", co-authored by Alejandro de la Concha, Nicolas Vayatis, and Argyris Kalogeratos, which was published in the 24th International Conference on Artificial Intelligence and Statistics (AISTATS) 2021. It contains proofs of theorems and technical supplementary material.

1 Proof of Theorems 1 and 2

We present the proofs of Theorem 1 and Theorem 2 of the main text. For completeness, we introduce key components such as basic concepts and results from the model selection literature.

The model selection framework offers an answer to the question: how to choose the function $pen(d)$ and the level of sparsity of the graph signals with respect to the Graph Fourier Transform (GFT) in order to guarantee good performance in practice of the proposed algorithms.

Definition 5. *Given a separable Hilbert space $\mathbb{H}$, a generalized linear Gaussian model is defined by:*

$$Y_\epsilon(g) = \langle f, g\rangle_{\mathbb{H}} + \epsilon W(g), \quad \text{for all } g \in \mathbb{H}, \tag{1}$$

where W is an isonormal process (Definition 6).

Definition 6. *A Gaussian process $(W(g))_{g\in\mathbb{H}}$ is said to be isonormal if it is centered with covariance given by $\mathbb{E}[W(h)W(g)] = \langle h, g\rangle_{\mathbb{H}}$ for all $h, g \in \mathbb{H}$.*

An isonormal process is the natural extension of the notion of standard normal random vector to the infinite dimensional case.

As stated in the main text, the change-point detection problem can be restated as a generalized linear Gaussian model, where $\mathbb{H} = \mathbb{R}^{T\times p}$: the dot product $\langle h, g\rangle_{\mathbb{H}}$ is the one inducing the Frobenius norm divided by T. Finally, the isonormal process $(W(\tilde{\mu}))_{\tilde{\mu}\in\mathbb{R}^{T\times p}}$ is defined by:

$$W(\tilde{\mu}) := \frac{\operatorname{tr}(\eta^\top \tilde{\mu})}{T}, \tag{2}$$

where $\eta \in \mathbb{R}^{T\times p}$ is a matrix whose rows follow a centered multivariate Gaussian distribution with covariance matrix $\mathbb{I}_p$. It is easy to show that $W(\tilde{\mu})$ satisfies Definition 6.

Theorem 3, which can be found as Theorem 4.18 in Massart and Picard (2003), details the model selection procedure and provides us with an oracle-type inequality for this kind of estimators. The result applies for a more general model selection procedure which allows us to deal with non-linear models. Both Theorem 1 and Theorem 2 are a direct consequence of this result.

Theorem 3. *Let $\{S_m\}_{m\in M}$ be some finite or countable collection of closed convex subsets of $\mathbb{H}$. It is assumed that for any $m \in M$, there exits some almost surely continuous version W of the isonormal process on S_m. Assume furthermore the existence of some positive and non-decreasing continuous function ϕ_m defined on $(0, +\infty)$ such that $\phi_m(x)/x$ is non-increasing and*

$$2\mathbb{E}\left[\sup_{g\in S_m}\left(\frac{W(g) - W(h)}{\|g-h\|^2 + x^2}\right)\right] \le x^{-2}\phi_m(x) \tag{3}$$

for any positive x and any point h in S_m. Let define $D_m > 0$ such that

$$\phi_m(\epsilon\sqrt{D_m}) = \epsilon D_m, \tag{4}$$

and consider some family of weights $\{x_m\}_{m\in M}$ *such that*

$$\sum_{m\in M} e^{-x_m} = \Sigma < \infty. \tag{5}$$

Let K *be some constant with* $K > 1$ *and take*

$$pen(m) \geq K\epsilon^2 \left(\sqrt{D_m} + \sqrt{2x_m}\right)^2. \tag{6}$$

Set for all $g \in H$, $\gamma(g) = \|g\|^2 - 2Y_\epsilon(g)$ *and consider some collection of* p_m-*approximate penalized least squares estimators* $\{\hat{f}_m\}_{m\in M}$ *i.e, for any* $m \in M$,

$$\gamma\left(\hat{f}_m\right) \leq \gamma(g) + \rho, \text{ for all } g \in S_m. \tag{7}$$

Defining a penalized $\rho-LSE$ *as* $\hat{f} = \hat{f}_{\hat{m}}$, *the following risk bounds holds for all* $f \in \mathbb{H}$

$$\mathbb{E}\left[\left\|\hat{f} - f\right\|^2\right] \leq C(K)\left[\inf_{m\in M}\left(d(f, S_m)^2 + pen(m)\right) + \epsilon(\Sigma + 1) + \rho\right]. \tag{8}$$

Theorem 3 requires us to have a predefined list of estimators that will be related with a list of closed convex subsets of $\mathbb{H}$. It states that we are able to recover a penalization term *pen*(m) which allows us to find a model satisfying an oracle kind inequality if we manage to control a kind of standardized version of the isonormal process and to design a set of weights for the elements in our list of candidate models. Theorem 4 is a restricted version of Theorem 3 which is more handy when dealing with the ℓ_1-penalization term. This version of the theorem appears as Theorem A.1 in Massart and Meynet (2011).

Theorem 4. *Let* $\{S_m\}_{m\in M}$ *be a countable collection of convex and compact subsets of a Hilbert space* $\mathbb{H}$*: lets define for any* $m \in M$,

$$\Delta_m = \mathbb{E}\left[\sup_{h\in S_m} W(h)\right], \tag{9}$$

and consider weights $\{x_m\}_{m\in M}$ *such that*

$$\Sigma := \sum_{m\in M} e^{-x_m} < \infty.$$

Let $K > 1$ *and assume that, for any* $m \in M$,

$$pen(m) \geq 2K\epsilon\left(\Delta_m + \epsilon x_m + \sqrt{\Delta_m \epsilon x_m}\right). \tag{10}$$

Given a non-negative $\rho_m, m \in M$,*define a* ρ_m*-approximate penalized least squares estimator as any* $\hat{f} \in S_{\hat{m}}, \hat{m} \in M$, *such that*

$$\gamma(\hat{f}) + pen(\hat{m}) \leq \inf_{m\in M}\left(\inf_{h\in S_m} \gamma(h) + pen(m) + \rho_m\right). \tag{11}$$

Then, there is a positive constant C(K) such that for all $f \in \mathbb{H}$ *and* $z > 0$, *with probability larger than* $1 - \Sigma e^{-z}$,

$$\left\|f - \hat{f}\right\|^2 + pen(\hat{m}) \leq C(K)\left[\inf_{m\in M}\left(\inf_{h\in S_m}\|f - h\|^2 + pen(m) + \rho_m\right) + (1 + z)\epsilon^2\right]. \tag{12}$$

After integrating the inequality with respect to z *leads to the following risk bound:*

$$\mathbb{E}\left[\left\|f - \hat{f}\right\|^2 + pen(\hat{m})\right] \leq C(K)\left[\inf_{m\in M}\left(\inf_{h\in S_m}\|f - h\|^2 + pen(m) + \rho_m\right) + (1 + \Sigma)\epsilon^2\right]. \tag{13}$$

Finally, we will make use of the following lemma that can be found as Lemma 2.3 in Massart and Meynet (2011), a concentration inequality for real-valued random variables.

Lemma 2. *Let $\{Z_i, i \in I\}$ be a finite family of real-valued random variables. Let ψ be some convex and continuously differentiable function on $[0,b)$, with $0 < b \leq \infty$, such that $\psi(0) = \psi'(0) = 0$. Assume that $\forall \gamma \in (0,b)$ and $\forall i \in I$, $\psi_{Z_i}(\gamma) \leq \psi(\gamma)$. Then, using any measurable set B with $\mathbb{P}[B > 0]$ we have:*

$$\frac{\mathbb{E}[\sup_{i\in I} Z_i \mathbf{1}_B]}{\mathbb{P}[B]} \leq \psi^{*-1}\left(\log \frac{|I|}{\mathbb{P}[B]}\right).$$

In particular, if one assumes that for some non-negative number ϵ, $\psi(\gamma) = \frac{\gamma^2\epsilon^2}{2} \forall \gamma \in (0,\infty)$, then:

$$\frac{\mathbb{E}\left[\sup_{i\in I} Z_i \mathbf{1}_B\right]}{\mathbb{P}\left[B\right]} \leq \epsilon\sqrt{2\log\frac{|I|}{\mathbb{P}(B)}} \leq \epsilon\sqrt{2\log|I|} + \epsilon\sqrt{2\log\frac{1}{\mathbb{P}(B)}}. \tag{14}$$

Proof of Theorem 1. Let us define the set $S_{(m,\tau)}$:

$$S_{(m,\tau)} := \left\{\tilde{\mu} \in F_\tau, \|\tilde{\mu}\|_{[\tau]} \leq m\epsilon\right\}, \tag{15}$$

where $\|\tilde{\mu}\|_{[\tau]} = \frac{\sum_{l=1}^{d_\tau} I_{\tau_l} \|\tilde{\mu}_{\tau_l}\|_1}{T}$.

And $M := \mathbb{N}^* \times \mathcal{T}$, where $\mathcal{T}$ is the set of all possible segmentations of a stream of length T.

We denote by $\hat{\tau}$ and $\hat{\tilde{\mu}}_{\hat{\tau}}$ the estimators obtained by solving the Problem of Eq 5 of the main text and we will define $d_{\hat{\tau}} := |\hat{\tau}| - 1$. Denote by $\hat{m}$ the smallest integer such that $\hat{\tilde{\mu}}_{\hat{\tau}}$ belongs to $S_{\hat{m}}$, i.e.

$$\hat{m} = \left\lceil \frac{\|\hat{\tilde{\mu}}_{\hat{\tau}}\|_{[\tau]}}{\epsilon} \right\rceil, \tag{16}$$

then,

$$\begin{aligned}
\gamma(\hat{\tilde{\mu}}_{\hat{\tau}}) + \lambda\hat{m}\epsilon + pen(d_{\hat{\tau}}) &\leq \gamma(\hat{\tilde{\mu}}_{\hat{\tau}}) + \lambda\|\hat{\tilde{\mu}}_{\hat{\tau}}\|_{[\hat{\tau}]} + \lambda\epsilon + pen(d_{\hat{\tau}}) \\
&\leq \inf_{\tau\in\mathcal{T}} \inf_{\tilde{\mu}\in S_{(m,\tau)}} \left[\gamma(\tilde{\mu}) + \lambda\|\tilde{\mu}\|_{[\tau]} + pen(d_\tau)\right] + \lambda\epsilon \quad (\text{Definition of } \hat{\tilde{\mu}}_{\hat{\tau}} \text{ and } \hat{\tau}) \\
&\leq \inf_{(m,\tau)\in M} \inf_{\tilde{\mu}\in S_{(m,\tau)}} \left[\gamma(\tilde{\mu}) + \lambda m\epsilon + pen(d_\tau)\right] + \lambda\epsilon.
\end{aligned}$$

In conclusion, we have the following result:

$$\gamma(\hat{\tilde{\mu}}_{\hat{\tau}}) + pen(\hat{m},\hat{\tau}) \leq \inf_{(m,\tau)\in M}\left[\inf_{\hat{\mu}\in S_{(m,\tau)}} \gamma(\tilde{\mu}) + pen(m,\tau) + \rho\right], \tag{17}$$

where $\rho = \lambda\epsilon > 0$ and $pen(m,\tau) = \lambda m\epsilon + pen(d_\tau) > 0$.

Ineq. 17 implies $\hat{\tilde{\mu}}_{\hat{\tau}}$ is a ρ-approximated least squares estimator. Then, the only hypothesis that remains to be proved is Expression 10.

We start by getting an upper bound for Δ_m. By the definition of the isonormal process $(W(\tilde{\mu}))_{\tilde{\mu}\in\mathbb{R}^{T\times p}}$, we know it is continuous. This implies that it achieves its maximum at $S_{(m,\tau)}$, a compact set, let call $\hat{g}$ this point, then:

$$\begin{aligned}
\mathbb{E}[|W(\hat{g})|] = \quad \mathbb{E}\left[\left|\frac{\mathrm{tr}(\zeta^\top\hat{g})}{T}\right|\right] &= \mathbb{E}\left[\left|\sum_{i=1}^{p}\sum_{t=1}^{T}\frac{\zeta_t^{(i)}\hat{g}_t^{(i)}}{T}\right|\right] \\
&\leq \sum_{t=1}^{T}\sum_{i=1}^{p}\left|\frac{\hat{g}_t^{(i)}}{T}\right| \mathbb{E}\left[\max_{\{i=1,..,p\}} |\zeta_t^{(i)}|\right] \\
&\leq \sum_{l=1}^{D_\tau}\frac{I_{\tau_l}}{T}\|\hat{g}_{\tau_l}\|_1 \,\mathbb{E}\left[\max_{\{i=1,..,p\}}\{\zeta_t^{(i)}, -\zeta_t^{(i)}\}\right] \\
&\leq \|\hat{g}\|_{[\tau]}\sqrt{2\log 2p} \quad (\text{Lemma 2}) \\
&\leq \sqrt{2}m\epsilon\sqrt{\log 2 + \log p}. \quad (\text{Eq. 15})
\end{aligned} \tag{18}$$

Let us define the $x_{(m,\tau)} = \gamma m + d_\tau L(d_\tau)$, where $\gamma > 0$. $L(d_\tau) = 2 + \log \frac{T}{d_\tau}$ is a constant that just depends on the cardinality of the segmentation induced by τ. Then:

$$\begin{aligned}
\Sigma &= \left(\sum_{m\in N^*} e^{-\gamma m}\right)\left(\sum_{\tau\in\mathcal{T}} e^{-d_\tau L(d_\tau)}\right) \\
&= \left(\frac{1}{e^\gamma - 1}\right)\left(\sum_{d=1}^{T} e^{-dL(d)} |\{\tau\in\mathcal{T}, |d_\tau| = d\}|\right) \\
&\leq \left(\frac{1}{e^\gamma - 1}\right)\left(\sum_{d=1}^{T} e^{-dL(d)} \binom{T}{d}\right) \\
&\leq \left(\frac{1}{e^{\frac{\gamma}{T}} - 1}\right)\left(\sum_{d=1}^{T} e^{-\frac{dL(d)}{T}} \left(\frac{eT}{d}\right)^d\right) \\
&\leq \left(\frac{1}{e^\gamma - 1}\right)\left(\sum_{d=1}^{T} e^{-d\left(L(d) - 1 - \log\frac{T}{d}\right)}\right) \\
&\leq \left(\frac{1}{e^\gamma - 1}\right)\left(\frac{1}{e-1}\right) < \infty.
\end{aligned} \tag{19}$$

Finally, let us fix $\eta = (3\sqrt{2} - 2)^{-1} > 0$, $K = \frac{3}{2+\eta} > 1$, $\gamma = \frac{\sqrt{\log p + L} - \sqrt{\log p + \log 2}}{K}$. It is clear $\gamma > 0$ since $L > \log 2$. Then by the expressions of Eq. 18 and Eq. 19, and the useful inequality $2\sqrt{ab} \leq a\eta^{-1} + b\eta$, we have:

$$\begin{aligned}
2\frac{K\epsilon}{T}\left[\Delta_{(m,\tau)} + \epsilon x_{(m,\tau)} + \sqrt{\Delta_{(m,\tau)}\epsilon x_{(m,\tau)}}\right] &\leq 2\frac{K\epsilon}{T}\left[\left(1+\frac{\eta}{2}\right)\Delta_{(m,\tau)} + \left(1 + \frac{\eta^{-1}}{2}\right) x_{(m,\tau)}\epsilon\right] \\
&\leq 2\frac{K\epsilon^2}{T}\left[\left(1+\frac{\eta}{2}\right)\left(\sqrt{2}m(\sqrt{\log p + \log 2})\right) + \right. \\
&\left.\left(1 + \frac{\eta^{-1}}{2}\right)(\gamma m + d_\tau L(d_\tau))\right] \\
&\leq 3\sqrt{2}\frac{\epsilon^2}{T}\left[\left(\sqrt{\log p + \log 2} + K\gamma\right) m + d_\tau L(d_\tau)\right] \\
&\leq 3\sqrt{2}\frac{\epsilon^2}{T}\left[\left(\sqrt{\log p + L}\right) m + d_\tau L(d_\tau)\right] \\
&\leq 3\sqrt{2}\frac{\epsilon^2}{T}\left(\sqrt{\log p + L}\right) m + \frac{d_\tau}{T}\left(c_1 + c_2 \log\frac{T}{d_\tau}\right) \\
&\leq \lambda m\epsilon + pen(d_\tau) = pen(m,\tau).
\end{aligned} \tag{20}$$

Then Eq. 10 is satisfied.

We can conclude by Eq. 17, 19 and 20 that, if the hypotheses of Theorem 4 are satisfied, then there exists a positive constant $C(K)$ such that $\mu^* \in \mathbb{R}^{T\times p}$ and $z > 0$, with probability larger that $1 - \Sigma e^{-z}$,

$$\begin{aligned}
\frac{\left\|\hat{\hat{\mu}}_{\hat{\tau}} - \mu^*\right\|_F^2}{T} + pen(\hat{m}) + pen(d_{\hat{\tau}}) &\leq C(K)\left[\inf_{(\tau,m)\in M}\inf_{\tilde{\mu}\in S_{(m,\tau)}}\left(\frac{\|\tilde{\mu} - \mu^*\|_F^2}{T} + \lambda m\epsilon + pen(d_\tau)\right) + \lambda\epsilon + (1+z)\epsilon^2\right] \\
&\leq C(K)\left[\inf_{\tau\in\mathcal{T}}\inf_{\tilde{\mu}\in F_\tau}\left(\frac{\|\tilde{\mu} - \mu^*\|_F^2}{T} + \lambda\|\tilde{\mu}\|_{[\tau]} + pen(d_\tau)\right) + 2\lambda\epsilon + (1+z)\epsilon^2\right].
\end{aligned} \tag{21}$$

Thanks to the last expression, we have that:

$$\frac{\left\|\hat{\hat{\mu}}_{\hat{\tau}} - \mu^*\right\|_F^2}{T} + \lambda\left\|\hat{\hat{\mu}}_{\hat{\tau}}\right\|_{[\tau]} + pen(d_{\hat{\tau}}) \leq C(K)\left[\inf_{\tau\in\mathcal{T}}\inf_{\tilde{\mu}\in F_\tau}\left(\frac{\|\tilde{\mu} - \mu^*\|_F^2}{T} + \lambda\|\tilde{\mu}\|_{[\tau]} + pen(d_\tau)\right) + 2\lambda\epsilon + (1+z)\epsilon^2\right]. \tag{22}$$

After integrating this inequality, we get the desired result.

Proof of Theorem 2. We will call S_{D_m} the space generated by m specific elements of the standard basis of $\mathbb{R}^p$ and let us define the set $S_{(D_m,\tau)}$ as:

$$S_{(D_m,\tau)} := \{\tilde{\mu} \in F_\tau | \tilde{\mu}_{\tau_l} \in S_{D_m} \text{ for all } l \in \{1,...,d_\tau\}\}, \tag{23}$$

This implies that we restrict the means defined in each of the segments to be elements of S_{D_m}.

Let define $M \subset \{1,...,p\} \times \mathcal{T}$ and let us denote $\hat{\tilde{\mu}}^{\mathrm{LSE}}_{\hat{\tau}^{\mathrm{LSE}}}$ and $\hat{\tau}^{\mathrm{LSE}}$ the solutions to the following optimization problem:

$$\begin{aligned}(\hat{\tilde{\mu}}^{\mathrm{LSE}}_{\hat{\tau}^{\mathrm{LSE}}}, \hat{\tau}^{\mathrm{LSE}}) := \underset{(\tau\in\mathcal{T},\tilde{\mu}\in S_{(D_m,\tau)})}{\operatorname{argmin}} \Bigg\{ &\sum_{l=1}^{d_\tau}\left(\sum_{t=\tau_{l-1}+1}^{\tau_l}\sum_{i=1}^{p}\frac{(\tilde{y}_t^{(i)} - \tilde{\mu}_{\tau_l}^{(i)})^2}{T}\right) + K_1\frac{D_m}{T} \\ &+\frac{d_\tau}{T}\left(K_2 + K_3\log\frac{T}{d_\tau}\right)\Bigg\}\end{aligned} \tag{24}$$

In order to obtain a oracle inequality for this estimator, we will rely on the result stated in Theorem 3. This means that we need to verify Ineq. 3 and Ineq. 6 for a set of weights satisfying Ineq. 5. We will begin by proving Ineq. 3. Let $\hat{g}, \hat{f} \in S_{(D_m,\tau)}$, then we have:

$$\begin{aligned}W(\hat{g}) - W(\hat{h}) &= \frac{\mathrm{tr}(\eta^\top\hat{g})}{T} - \frac{\mathrm{tr}(\eta^\top\hat{h})}{T} \\ &\le \sum_{i\in Supp_m}\sum_{t=1}^{T}\frac{\zeta_t^{(i)}(\hat{g}_t^{(i)} - \hat{h}_t^{(i)})}{T} \\ &\le \sum_{i\in Supp_m}\sqrt{\sum_{t=1}^{T}\frac{(\zeta_t^{(i)})^2}{T}}\sqrt{\sum_{t=1}^{T}\frac{(\hat{g}_t^{(i)} - \hat{h}_t^{(i)})^2}{T}} \quad \left(\text{Cauchy-Schwarz Ineq.}\right) \\ &\le \sqrt{\sum_{i\in Supp_m}\sum_{t=1}^{T}\frac{(\zeta_t^{(i)})^2}{T}}\sqrt{\sum_{i\in Supp_m}\sum_{t=1}^{T}\frac{(\hat{g}_t^{(i)} - \hat{h}_t^{(i)})^2}{T}} \quad \left(\text{Cauchy-Schwarz Ineq.}\right) \\ &= \left\|\hat{g} - \hat{h}\right\|_{\mathbb{H}}\sqrt{\sum_{i\in Supp_m}\sum_{t=1}^{T}\frac{(\zeta_t^{(i)})^2}{T}}.\end{aligned} \tag{25}$$

Thanks to this inequality and the fact that $\zeta_t^{(i)}$ follows a standard Gaussian distribution, we derive the following expression for each $h \in S_{(D_m,\tau)}$:

$$\begin{aligned}2\mathbb{E}\left[\sup_{\hat{g}\in S_{(D_m,\tau)}}\left(\frac{W(\hat{g}) - W(\hat{h})}{\left\|\hat{g} - \hat{h}\right\|_{\mathbb{H}}^2 + x^2}\right)\right] &\le x^{-1}\mathbb{E}\left[\sup_{\hat{g}\in S_{(D_m,\tau)}}\left(\frac{W(\hat{g}) - W(\hat{h})}{\left\|\hat{g} - \hat{h}\right\|_{\mathbb{H}}}\right)\right] \\ &\le x^{-1}\left[\mathbb{E}\left[\sup_{g\in S_{(D_m,\tau)}}\left(\frac{W(\hat{g}) - W(\hat{h})}{\left\|\hat{g} - \hat{h}\right\|_{\mathbb{H}}}\right)^2\right]\right]^{1/2} \quad \left(\text{Jensen's ineq.}\right) \\ &\le x^{-1}\left[\mathbb{E}\left[\sum_{i\in Supp_m}\frac{\sum_{t=1}^{T}(\zeta_t^{(i)})^2}{T}\right]\right]^{1/2} \\ &= x^{-1}\sqrt{D_m}. \quad \left((\zeta_t^{(i)}) \text{ follows a standard Gaussian distribution}\right).\end{aligned} \tag{26}$$

We can conclude that Ineq. 3 with $\phi_m(x) = x\sqrt{D_m}$, from which is straightforward to derive D_m.

Next, we define $x_{(m,\tau)} = \gamma D_m + d_\tau L(d_\tau)$, where $\gamma > 0$ and $L(d_\tau) = 2 + \log\frac{T}{d_\tau}$, which is a constant that only depends on the cardinality of the segmentation induced by τ. Then:

$$\begin{aligned}
\Sigma = \sum_{(m,\tau)\in M} e^{-x_{(m,\tau)}} &= \left(\sum_{m\in N^*, m\le p} e^{-\gamma D_m}\right)\left(\sum_{\tau\in\mathcal{T}} e^{-d_\tau L(d_\tau)}\right) \\
&\le \left(\frac{1}{e^\gamma - 1}\right)\left(\sum_{d=1}^{T} e^{-dL(d)} |\{\tau \in \mathcal{T}, |d_\tau| = d\}|\right) \\
&\le \left(\frac{1}{e^\gamma - 1}\right)\left(\sum_{d=1}^{T} e^{-dL(d)} \binom{T}{d}\right) \\
&\le \left(\frac{1}{e^\gamma - 1}\right)\left(\sum_{d=1}^{T} e^{-dL(d)} \left(\frac{eT}{d}\right)\right)^d \\
&\le \left(\frac{1}{e^\gamma - 1}\right)\left(\sum_{d=1}^{T} e^{-d\left(L(d)-1-\log\frac{T}{d}\right)}\right) \\
&\le \left(\frac{1}{e^\gamma - 1}\right)\left(\frac{1}{e-1}\right) < \infty.
\end{aligned} \tag{27}$$

Let fix $\eta > 0$, $C > 2 + \frac{2}{\eta}$, then $K = \frac{C\eta}{2(1+\eta)} > 1$. And fix $0 < \delta < 1$ such that $\gamma = 1 - \delta > 0$. By using the useful inequality $2\sqrt{ab} \le a\eta^{-1} + b\eta$.

$$\begin{aligned}
\frac{K\epsilon^2}{T}\left(\sqrt{D_m} + \sqrt{2(\gamma D_m + d_\tau L(d_\tau)}\right)^2 &\le \frac{K\epsilon^2}{T}\left(\sqrt{(1+\gamma)D_m} + \sqrt{2d_\tau L(d_\tau)}\right)^2 \qquad \left(\text{Triangle inequality}\right) \\
&\le \frac{K\epsilon^2}{T}\Big((1+\gamma)D_m + 2\sqrt{2(1+\gamma)D_m d_\tau L(d_\tau)} \\
&+2d_\tau L(d_\tau)) \\
&\le \frac{K\epsilon^2}{T}\left((1+\gamma)D_m + 2d_\tau L(d_\tau)\right. \\
&\left.+(1+\gamma)D_m\eta + 2d_\tau L(d_\tau)\eta^{-1}\right) \\
&\le \frac{K\epsilon^2}{T}\left((1+\gamma)(1+\eta)D_m + (2+\eta^{-1})d_\tau L(d_\tau)\right) \\
&\le \left(C\eta(2-\delta)\epsilon^2\frac{D_m}{T} + C\epsilon^2\frac{d_\tau}{T}L(d_\tau)\right) \\
&= K_1\frac{D_m}{T} + \frac{d_\tau}{T}\left(c_1 + c_2\log\frac{T}{d_\tau}\right) \\
&= pen(m,\tau).
\end{aligned} \tag{28}$$

As the hypotheses of Theorem 3 are satisfied, we obtain the desired result.

References

Massart, P. and Meynet, C. (2011). The Lasso as an $\ell 1$-ball model selection procedure. *Electr. J. of Statistics*, 5.

Massart, P. and Picard, J. (2003). *Concentration Inequalities and Model Selection: Ecole d'Eté de Probabilités de Saint-Flour XXXIII - 2003*. Lecture Notes in Mathematics. Springer Berlin Heidelberg.