A Proving lemma 5, concentration result

Lemma 10. *Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \Pr)$ be a filtered space, with $\mathcal{F}_0$ the trivial sigma algebra. Let $(x_t)_{t\geq 1}$ be a previsible sequence $x_t \colon \Omega \mapsto \mathbb{R}^d$ and let $(\epsilon_t)_{t\geq 1}$ with $\epsilon_t \colon \Omega \mapsto \mathbb{R}$ be a sequence of random variables adapted to the filtration, with ϵ_t 1-subGaussian conditionally on $\mathcal{F}_{t-1}$ for all t. Let $(N_t)_{t\geq 1}$ be a non-decreasing sequence of integers. Let $(\mathcal{A}_t)_{t\geq 0}$ be a sequence of random sets $\mathcal{A}_t \colon \Omega \mapsto 2^{2^{\mathcal{X}}}$, such that $\mathcal{A}_0$ is $\mathcal{F}_0$-measurable, $(\mathcal{A}_t)_{t\geq 1}$ previsible and $1 \leq |\mathcal{A}_t| < N_t$ almost surely for all $t \geq 0$. Let $k \colon \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$ be a symmetric, positive-semidefinite kernel. Then for any given $\delta \in (0,1)$ and $\eta > 0$, for all $t \geq 0$ and all $A \in \mathcal{A}_t$ we have*

$$\|\epsilon^A_{1:t}\|_{(I+(K^A_t+\eta I)^{-1})^{-1}} \leq 2\log\left(\det(K^A_t + I + \eta I)^{\frac{1}{2}} N_t/\delta\right),$$

with probability $1-\delta$, where $\epsilon^A_{1:t}$ for the random vector that is the concatenation of $(\epsilon_z \colon x_z \in A)_{z=1}^t$.

Proof. For a function $g \colon \mathcal{X} \mapsto \mathbb{R}$ and a sequence of real numbers $(a_t)_{t\geq 1}$, define

$$\Delta^{g,n}_t = \exp\left\{(g(x_t) + a_t)\epsilon_t - \tfrac{1}{2}(g(x_t)+a_t)^2\right\},$$

with $\Delta^{g,n}_0$ defined as equal to 1 almost surely. Then $\Delta^{g,n}_t$ is $\mathcal{F}_t$ measurable for all $t \geq 0$. By the conditional subGaussianity of ϵ_t, we have that $\mathbb{E}[\Delta^{g,n}_t | \mathcal{F}_{t-1}] \leq 1$ for all $t \geq 0$ almost surely. For a set $A \in 2^{\mathcal{X}}$, define

$$\mathcal{M}^{g,n}_t(A) = \Delta^{g,n}_0 \prod_{z=1}^t (\Delta^{g,n}_z)^{1\{x_z \in A\}}.$$

Then, for any $A \in 2^{\mathcal{X}}$ and all $t \geq 1$, $\mathbb{E}[\mathcal{M}^{g,n}_t(A)|\mathcal{F}_{t-1}] \leq \mathcal{M}^{g,n}_{t-1}(A)$ and $\mathbb{E}[\mathcal{M}^{g,n}_t(A)] \leq 1$.

Let $\zeta = (\zeta_t)_{t\geq 1}$ be a sequence of independent and identically distributed Gaussian random variables with mean 0 and variance $\eta > 0$, independent of $\mathcal{F}_\infty = \bigcup_{t\geq 0} \mathcal{F}_t$. Let h be a random real valued function on A distributed according to the Gaussian process measure $\mathcal{GP}(0, k|_A)$, where $k|_A$ is the restriction of k to A. Define

$$M^A_t = \mathbb{E}[\mathcal{M}^{h,\zeta}_t(A) | \mathcal{F}_\infty].$$

Then M^A_t is itself a non-negative supermartingale bounded in expectation by 1. Define $\widetilde{M}^A_t = M^A_t / N_t$. Since $N_t \geq 1$ for all $t \geq 0$ and is non-decreasing, $\widetilde{M}^A_t$ is a non-negative supermartingale bounded in expectation by $1/N_t$.

For $A \in \mathcal{B}(\mathcal{X})$, let $B^A_t = \{\omega \colon \widetilde{M}^A_t > 1/\delta\}$ and $B_t = \bigcup_{A\in\mathcal{A}_t} B^A_t$. Define the stopping time $\tau(\omega) = \inf\{t \colon \omega \in B_t\}$. Then

$$\Pr[B^A_\tau | \mathcal{F}_{\tau-1}] \leq \delta\mathbb{E}[\widetilde{M}^A_\tau | \mathcal{F}_{\tau-1}] = \delta\mathbb{E}[M^A_\tau|\mathcal{F}_{\tau-1}]/N_\tau \leq \delta/N_\tau M^A_{\tau-1} \quad \text{a.s.}$$

We now examine the probability of B_τ. We have

$$\Pr[B_\tau] = \mathbb{E}\left[\Pr[B_\tau | F_{\tau-1}]\right] \leq \sum_{A\in\mathcal{B}(\mathcal{X})} \mathbb{E}\left[1\{A \in \mathcal{A}_\tau\} \Pr[B^A_\tau|\mathcal{F}_{\tau-1}]\right] \leq \delta/N_\tau \sum_{A\in\mathcal{B}(\mathcal{X})} \mathbb{E}\left[1\{A\in\mathcal{A}_\tau\} M^A_{\tau-1}\right].$$

The final expectation is complicated by the fact that the event $\{A \in \mathcal{A}_t\}$ is not independent of M^A_{t-1}. However,

$$\{A \in \mathcal{A}_t\} \subset \{A \in \mathcal{Z} \colon \mathcal{Z} \subset \mathcal{B}(X), |\mathcal{Z}| \leq N_t\}.$$

The latter event holds with probability 1 for all $t \geq 1$, and is therefore independent of M^A_t. This gives,

$$\Pr[B_\tau] \leq \delta/N_\tau \sum_{A\in\mathcal{B}(\mathcal{X})} \mathbb{E}\left[1\{A\in\mathcal{A}_\tau\}M^A_{\tau-1}\right] \leq \delta/N_\tau \sum_{A\in\mathcal{B}(\mathcal{X})} \mathbb{E}\left[1\{A \in \mathcal{Z} \colon |\mathcal{Z}| \leq N_t\} M^A_{\tau-1}\right] \tag{5}$$

$$= \delta/N_\tau \mathbb{E}[M^A_{\tau-1}] \sum_{A\in\mathcal{B}(\mathcal{X})} \mathbb{E}\left[1\{A\in\mathcal{Z} \colon |\mathcal{Z}| \leq N_t\}\right] \leq \delta, \tag{6}$$

and consequently

$$\Pr\left[\cup_{t\geq 0} B_t\right] = \Pr[\tau < \infty] = \Pr[B_\tau, \tau < \infty] \leq \Pr[B_\tau] \leq \delta. \tag{7}$$

Finally, by comparing with the proof of Theorem 1 in Chowdhury and Gopalan (2017), it can be verified that

$$M_t^A = \det(K_t^A + I + \eta I)^{-\frac{1}{2}} \exp\left\{\frac{1}{2}\|\epsilon_{1:t}^A\|_{(I+(K_t^A+\eta I)^{-1})^{-1}}\right\}.$$

The statement of the lemma follows from using this expression with equation (7), and noting that logarithms preserve order. □

Proof of lemma 5. To prove lemma 5, first since $|\mathcal{A}_t| \leq |\widetilde{\mathcal{A}}_t| \leq \widetilde{N}_t$, we can use $\widetilde{N}_t$ from lemma 6 as the bound N_t required for lemma 10. Then the proof of lemma 5 follows the proof of theorem 2 in Chowdhury and Gopalan (2017), with our concentration inequality, lemma 10, used instead of their theorem 1. □