Thm 20.5. [Concentration of Least Squares Estimators] (Some significant contents are lost now) V.0.1

$\forall \delta \in (0,1)$, w.p. $1-\delta$, it holds that

$$\mathbb{P}(\exists t \in \mathbb{N}^+ : \theta_* \notin C_t) \le \delta$$

where

$$C_t = \left\{ \theta \in \mathbb{R}^d : \|\hat{\theta}_{t-1} - \theta\|_{V_{t-1}(\lambda)} < \sqrt{\lambda}\cdot\|\theta_*\|_2 + \sqrt{2\cdot\log\left(\frac{1}{\delta}\right) + \log\left(\frac{\det V_t(\lambda)}{\lambda^d}\right)} \right\}.$$

Proof.

$$\begin{aligned}
\hat{\theta}_t &= V_t^{-1}(\lambda)\cdot X_t^T\cdot Y_t \\
&= V_t^{-1}(\lambda)\cdot X_t^T(X_t\cdot\theta_* + \underbrace{\epsilon_t}_{Y_t - X_t\cdot\theta_*.}) \\
&= V_t^{-1}(\lambda)\cdot X_t^T\cdot X_t\cdot\theta_* + V_t^{-1}(\lambda)\cdot X_t^T\cdot\epsilon_t \\
&= V_t^{-1}(\lambda)\cdot(V_t(\lambda) - \lambda I)\cdot\theta_* + V_t^{-1}(\lambda)\cdot X_t^T\cdot\epsilon_t \\
&= \theta_* - \lambda\cdot V_t^{-1}(\lambda)\cdot\theta_* + V_t^{-1}(\lambda)\cdot X_t^T\cdot\epsilon_t.
\end{aligned}$$

$$\begin{aligned}
\|\hat{\theta}_t - \theta_*\|_{V_t(\lambda)} &\le \lambda\cdot\|\theta_*\|_{V_t^{-1}(\lambda)} + \|X_t^T\cdot\epsilon_t\|_{V_t^{-1}(\lambda)} \\
&\le \lambda\cdot\sqrt{\sigma_{max}(V_t^{-1}(\lambda))}\cdot\|\theta_*\|_2 + R\cdot\sqrt{2\cdot\log\left(\frac{\det(V_t(\lambda))^{\frac{1}{2}}}{\det(V_0(\lambda))^{-\frac{1}{2}}\,\delta}\right)} \\
&= \sqrt{\lambda}\|\theta_*\|_2 + R\cdot\sqrt{2\cdot\log\left(\frac{\det(V_t(\lambda))^{\frac{1}{2}}}{\det(V_0(\lambda))^{-\frac{1}{2}}\,\delta}\right)},
\end{aligned}$$

where the second inequality follows from the self-normalized bound for vector-valued martingales.

Thm 20.4. [Concentration of Self-normalized Vector-valued Martingales]

Let $\{\mathcal{F}_t\}_{t=0}^{+\infty}$ be a filtration. Let $\{\eta_t\}_{t=0}^{+\infty}$ be a real-valued stochastic process s.t. η_t is $\mathcal{F}_t$-measurable and η_t is conditionally R-subgaussian for some $R \geq 0$. Let $\{A_t\}_{t=0}^{+\infty}$ be an $\mathbb{R}^d$-valued stochastic process s.t. X_t is $\mathcal{F}_{t-1}$-measurable. Let

$$V_t(\lambda) = X_t^T \cdot X_t + \lambda I,$$

where $X_t \in \mathbb{R}^{t \times d}$ and $X_{t[s,:]} = A_s^T$. Let $\ell_t \in \mathbb{R}^t$ and $\ell_{t[s]} = \eta_s$. Then

$$\| X_t^T \cdot \ell_t \|^2_{V_t^{-1}(\lambda)} \leq 2R^2 \cdot \log\left(\frac{1}{\delta} \cdot \frac{\det(V_t(\lambda))^{\frac{1}{2}}}{\det(V_0(\lambda))^{\frac{1}{2}}} \right),$$

holds w.p. $1-\delta$ $\forall \delta \in (0,1)$, $t \geq 0$.