Assumption:

① $X_t = \langle \theta^*, A_t \rangle + \eta_t$

② η_t is conditionally 1-subgaussian

Let $V_t = \sum_{s=1}^{t} A_s A_s^T$ and assume V_t is invertible. Let $S_t = \sum_{s=1}^{t} \eta_s X_s$. Since

$$\hat{\theta}_t = V_t^{-1} \sum_{s=1}^{t} X_s A_s = V_t^{-1} \sum_{s=1}^{t} (A_s^T \theta^* + \eta_s) A_s = \theta^* + V_t^{-1} S_t,$$

we have

$$\frac{1}{2} \| \hat{\theta}_t - \theta^* \|_{V_t}^2 = \frac{1}{2} \| S_t \|_{V_t^{-1}}^2 = \max_{x \in \mathbb{R}^d} \left(x^T S_t - \frac{1}{2} \| x \|_{V_t}^2 \right).$$

Lemma 20.2 [Exponential of the "Elliptical Potential" is Supermartingale]

For some fixed $x \in \mathbb{R}^d$, the process $M_t(x) = \exp(x^T S_t - \frac{1}{2} \| x \|_{V_t(\lambda)}^2)$ is an $\mathbb{F}$-adapted non-negative supermartingale with $M_0(x) \leq 1$.

Proof. Let $D_t(x) = \exp(x^T A_s \cdot \eta_s - \frac{1}{2} \| x \|_{A_s A_s^T}^2)$, $\forall t \geq 1$ and $D_0(x) = \exp(- \frac{\lambda}{2} \| x \|_2^2)$. Then $M_t(x) = M_{t-1}(x) D_t(x)$, $\forall t \geq 1$, and

$$\begin{aligned} \mathbb{E}[M_t(x) \mid \mathcal{F}_{t-1}] &= \mathbb{E}[D_t(x) \mid \mathcal{F}_{t-1}] M_{t-1}(x) \\ &= \mathbb{E}_{t-1}[\exp(x^T A_s \cdot \eta_s - \frac{1}{2} \| x \|_{A_s A_s^T}^2)] M_{t-1}(x) \\ &\leq \mathbb{E}_{t-1}[\exp(\frac{x^T A_s A_s^T x}{2} - \frac{1}{2} \| x \|_{A_s A_s^T}^2)] M_{t-1}(x) \\ &\leq M_{t-1}(x). \end{aligned}$$

∎

Laplace's Method / Laplace's Approximation

· Assumption:

① f is twice-differentiable

② f has a unique maximum at $x_0 \in (a, b)$ (hence $f'(x_0) = 0$, $f''(x_0) < 0$).

· Intuition:

① Function $h(x) = e^{sf(x)}$ with a large s can be well-approximated by a Gaussian function since $\frac{h(x)}{h(x_0)} = e^{s(f(x)-f(x_0))}$ will approach 0 exponentially fast for a large s. Thus,

$$I_s = \int_a^b e^{sf(x)}\,dx \approx \int_{-\infty}^{+\infty} e^{-\frac{(x-x_0)^2}{2\sigma^2}}$$

integral of an unnormalized "Gaussian" function

② Further, for a large enough s, $\int_a^b e^{sf(x)}\,dx$ will be dominated by the $f(x_0)$ since the significant contributions to the integral $I_s = \int_a^b e^{sf(x)}\,dx$ only come from points x in a neighborhood of x_0.

· Application:

① approximate $I_s = \int_a^b e^{sf(x)}\,dx$ using $f(x_0)$

② approximate $f(x_0)$ using $I_s = \int_a^b e^{sf(x)}\,dx$

· Derivation:

①
Consider the 2nd-order Taylor expansion of $f(x)$ at x_0:

$$f(x) \approx f(x_0) + f'(x_0)(x-x_0) + \frac{f''(x_0)}{2}(x-x_0)^2$$
$$= f(x_0) - \frac{|f''(x_0)|}{2}(x-x_0)^2,$$

where the equality comes from $f'(x_0)=0$ and $f''(x_0)<0$. Hence

$$\int_a^b e^{sf(x)}\,dx \approx \int_a^b e^{s(f(x_0) - \frac{|f''(x_0)|}{2}(x-x_0)^2)}\,dx$$
$$= e^{sf(x_0)}\sqrt{\frac{2\pi}{s|f''(x_0)|}}\int_a^b \sqrt{\frac{s|f''(x_0)|}{2\pi}}\, e^{-\frac{s|f''(x_0)|}{2}(x-x_0)^2}\,dx \quad \left(\sigma^2 = \frac{1}{s|f''(x_0)|}\right)$$
$$\approx e^{sf(x_0)}\sqrt{\frac{2\pi}{s|f''(x_0)|}}\int_{-\infty}^{+\infty} \sqrt{\frac{s|f''(x_0)|}{2\pi}}\, e^{-\frac{s|f''(x_0)|}{2}(x-x_0)^2}\,dx$$
$$= e^{sf(x_0)}\sqrt{\frac{2\pi}{s|f''(x_0)|}},$$

where the first $\approx$ comes from that $e^{sf(x)}$ can be well-approximated by the unnormalized "Gaussian" function $e^{-\frac{s|f''(x_0)|}{2}(x-x_0)^2}$ for large enough s, and the second $\approx$ comes from that $\sigma^2 = \frac{1}{s|f''(x_0)|}$ is small again due to large enough s.

② d-dimensional case:

$$\int_a^b e^{sf(x)}\,dx \approx \int_a^b e^{s\left(f(x_0)+\frac{\|x-x_0\|^2_{H(x_0)}}{2}\right)}\,dx \qquad \left(\Sigma = (-sH(x_0))^{-1} = -\tfrac{1}{s}H^{-1}(x_0)\right)$$

$$= e^{sf(x_0)}\sqrt{\left|-\tfrac{2}{s}\pi H^{-1}(x_0)\right|}\int_a^b \frac{1}{\sqrt{\left|-\frac{2}{s}\pi H^{-1}(x_0)\right|}}\, e^{-\frac{\|x-x_0\|^2_{(-sH(x_0))}}{2}}\,dx$$

$$\approx e^{sf(x_0)}\sqrt{\left|-\tfrac{2}{s}\pi H^{-1}(x_0)\right|}\int_{\mathbb{R}^d} \frac{1}{\sqrt{\left|-\frac{2}{s}\pi H^{-1}(x_0)\right|}}\, e^{-\frac{\|x-x_0\|^2_{(-sH(x_0))}}{2}}\,dx$$

$$= e^{sf(x_0)}\sqrt{\left|-\tfrac{2}{s}\pi H^{-1}(x_0)\right|}. \qquad \left(e^{sf(x_0)}\left(\tfrac{2\pi}{s}\right)^{\frac{d}{2}}|-H(x_0)|^{-\frac{1}{2}}\right)$$

③ Note that for any h on $\mathbb{R}^d$, we also have

$$e^{sf(x_0)} \underset{\text{相差一个 multiplicative constant (depend on } s\text{?)}}{\approx} \int_{\mathbb{R}^d} e^{sf(x)}\,dh(x).$$

Lemma 20.3

Let h be some probability measure on $\mathbb{R}^d$. Then $\bar{M}_t = \int_{\mathbb{R}^d} M_t(x)\,dh(x)$ is also an $\mathbb{F}$-adapted non-negative supermartingale with $\bar{M}_0 = 1$.

Proof. TBD. 2022.09.07

Theorem. For all $\lambda > 0$ and $\delta \in (0,1)$,

$$\mathbb{P}\left(\sup_{t\in\mathbb{N}} \frac{1}{2}\|S_t\|^2_{V_t(\lambda)^{-1}} \geq \log\frac{1}{\delta} + \log\sqrt{\frac{|V_t(\lambda)|}{|H|}}\right) \leq \delta.$$

Proof. Let $H = \lambda I \in \mathbb{R}^{d\times d}$, and $h = \mathcal{N}(0, H^{-1})$. And

$$\bar{M}_t = \int_{\mathbb{R}^d} M_t(x)\, dh(x)$$

$$= \int_{\mathbb{R}^d} \frac{1}{\sqrt{|2\pi H^{-1}|}} \exp\left(S_t^T x - \frac{1}{2}\|x\|^2_{V_t} - \frac{1}{2}\|x\|^2_H\right) dx$$

$$= |2\pi H|^{\frac{1}{2}} \int_{\mathbb{R}^d} \exp\left(S_t^T x - \frac{1}{2}\|x\|^2_{V_t(\lambda)}\right) dx \quad (\text{abbreviate } V_t + H = V_t(\lambda))$$

$$= |2\pi H|^{\frac{1}{2}} \int_{\mathbb{R}^d} \exp\left(-\frac{1}{2}\|x - V_t(\lambda)^{-1}S_t^T\|^2_{V_t(\lambda)} + \frac{1}{2}\|S_t\|^2_{V_t(\lambda)^{-1}}\right) dx$$

$$= |2\pi H|^{\frac{1}{2}} \exp\left(\frac{1}{2}\|S_t\|^2_{V_t(\lambda)^{-1}}\right) \int_{\mathbb{R}^d} \exp\left(-\frac{1}{2}\|x - V_t(\lambda)^{-1}S_t^T\|^2_{V_t(\lambda)}\right) dx$$

$$= \sqrt{|2\pi V_t^{-1}(\lambda)|} \int_{\mathbb{R}^d} \frac{1}{\sqrt{|2\pi V_t^{-1}(\lambda)|}} \exp\left(-\frac{1}{2}\|x - V_t^{-1}(\lambda)S_t^T\|^2_{V_t(\lambda)}\right) dx$$

$$= \exp\left(\frac{1}{2}\|S_t\|^2_{V_t(\lambda)^{-1}}\right)\sqrt{\frac{|2\pi H|}{|2\pi V_t(\lambda)|}}$$

$$= \exp\left(\frac{1}{2}\|S_t\|^2_{V_t(\lambda)^{-1}}\right)\sqrt{\frac{|H|}{|V_t(\lambda)|}}$$

is a non-negative supermartingale with $\bar{M}_0 = 1$. Applying supermartingale maximal inequality shows that

$$\mathbb{P}\left(\sup_{t\in\mathbb{N}} \bar{M}_t \geq \frac{1}{\delta}\right) = \mathbb{P}\left(\sup_{t\in\mathbb{N}} \log \bar{M}_t \geq \log\frac{1}{\delta}\right)$$

$$= \mathbb{P}\left(\sup_{t\in\mathbb{N}} \frac{1}{2}\|S_t\|^2_{V_t(\lambda)^{-1}} \geq \log\frac{1}{\delta} + \log\sqrt{\frac{|V_t(\lambda)|}{|H|}}\right)$$

$$\leq \frac{\mathbb{E}[\bar{M}_0]}{\frac{1}{\delta}}$$

$$= \delta.$$