Importance Sampling. Minimization of Renyi Divergence. Strong convexity investigation. Updated version from 22.08.2017

Alena Shilova

Skoltech

September 06,2017

Optimization problems

Optimization problem with $\alpha = 1$

$$\max_{\theta \in \Theta} \frac{1}{I} \int \varphi(x) f(x) \left[\theta^T T(x) - A(\theta)\right] \, dx$$

Optimization problem with $\alpha = 2$

$$\min_{\theta \in \Theta} \int \frac{\varphi^2(x) f^2(x)}{\exp(\theta^T T(x) - A(\theta)) h(x)} \, dx.$$

Normal distribution

For the normal distribution $\mathcal{N}(\mu, \Sigma)$ to look like the one from exponential family, let's make a change of variables: $S = \Sigma^{-1}$ and $m = \Sigma^{-1}\mu$. Then the density of distribution will look like the following:

$$g_{m,S}(x) = \frac{1}{2\pi} \exp\Big(m^T x - \frac{1}{2}\operatorname{tr}(Sxx^T)\Big) \exp\Big(-\frac{1}{2}\big(m^T S^{-1} m - \log|S|\big)\Big).$$

So in this case $A(m, S) = \frac{1}{2} m^T S^{-1} m - \frac{1}{2}\log|S|,\ T(x) = (x, -\frac{1}{2}xx^T)$ and $h(x) = \frac{1}{2\pi}$

Strong convexity of the first problem

Problem

$$\min_{m,S} \int \varphi(x) f(x) \left[\frac{1}{2} x^T S x - m^T x + \frac{1}{2} m^T S^{-1} m - \frac{1}{2} \log |S| \right] dx$$

As it was shown earlier and exploiting the equivalency of the norms, this objective function is strongly convex:

- with the coefficient $\frac{\lambda_{min}^2}{2}$ in the norm $\| \cdot \|_2$;
- with the coefficient $\frac{\lambda_{min}^2}{2}$ in the norm $\| \cdot \|_\infty$;
- with the coefficient $\frac{\lambda_{min}^2}{2\sqrt{n}}$ in the norm $\| \cdot \|_1$.

Juditsky and Nemirovski

Lemma

Let δ be an open interval. Suppose $\phi : \delta \to \mathbb{R}_*$ is a twice differentiable convex function such that ϕ'' is monotonically non-decreasing. Let $\mathbb{S}_n(\delta)$ be the set of all symmetric $n \times n$ matrices with eigenvalues in δ. Define the function $F : \mathbb{S}_n(\delta) \dagger' \mathbb{R}_*$

$$F(X) = \sum_{i=1}^{n} \phi(\lambda_i(X))$$

and let

$$f(t) = F(X + tH)$$

for some $X \in \mathbb{S}_n(\delta)$, $H \in \mathbb{S}_n$. Then, we have,

$$f''(0) \leq 2 \sum_{i=1}^{n} \phi''(\lambda_i(X)) \lambda_i^2(H)$$

Strong convexity of the first problem

- $\frac{1}{2}x^T Sx - m^T x + \frac{1}{2}m^T S^{-1}m$ is convex.
- $-\frac{1}{2}\log|S|$; its dual is the function $-2n + \frac{1}{2}\ln 2 - \frac{1}{2}\sum_{i=1}^{n}\ln|\lambda_i(S)|$. Exploiting the lemma of Juditsky and Nemirovski and considering $\tilde{f}(t) = \frac{1}{2}\ln 2 - 2n - \frac{1}{2}\sum_{i=1}^{n}\ln|\lambda_i(X+tH)|$ for some fixed X, s.t. $\lambda(X) \subseteq [-\frac{\lambda_{max}}{2}, -\frac{\lambda_{min}}{2}]$ and H, then we have the following:

$$\tilde{f}''(0) \leq 2\sum_{i=1}^{n}\frac{\lambda_i^2(H)}{2\lambda_i^2(X)} \leq \frac{4}{\lambda_{min}^2}\sum_{i=1}^{n}\lambda_i^2(H) \leq \frac{4}{\lambda_{min}^2}\left(\sum_{i=1}^{n}|\lambda_i(H)|\right)^2$$

and

$$\tilde{f}''(0) \leq 2\sum_{i=1}^{n}\frac{\lambda_i^2(H)}{2\lambda_i^2(X)} \leq \lambda_{max}^2(H)\sum_{i=1}^{n}\frac{1}{\lambda_i^2(X)} \leq \frac{4n}{\lambda_{min}^2}\lambda_{max}^2(H)$$

Strong convexity of the first problem

Problem

$$\min_{m,S} \int \varphi(x) f(x) \left[\frac{1}{2} x^T S x - m^T x + \frac{1}{2} m^T S^{-1} m - \frac{1}{2} \log |S| \right] dx$$

This objective function is strongly convex:

- with the coefficient $\frac{\lambda_{min}^2}{4}$ in the spectral norm;
- with the coefficient $\frac{\lambda_{min}^2}{4n}$ in the nuclear norm;

Strong convexity of the second problem

Problem

$$\min_{m,S} \int \varphi^2(x) f^2(x) \sqrt{(2\pi)^n} \exp\left[\frac{1}{2} x^T S x - m^T x + \frac{1}{2} m^T S^{-1} m - \frac{1}{2} \log |S|\right] dx$$

$e^{f(x)}$ is strongly convex $\Leftrightarrow$ f(x) is strongly convex and $f(X) \subseteq [-C, +\infty]$. Indeed, for any x and y

$$\begin{aligned} &\langle e^{f(x)} \nabla f(x) - e^{f(y)} \nabla f(y), x - y \rangle \geq \\ \geq &e^{-C} \langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \\ \geq &e^{-C} \beta \|x - y\|^2 \end{aligned}$$

Strong convexity of the second problem

Problem

$$\min_{m,S} \int \varphi^2(x) f^2(x) \sqrt{(2\pi)^n} \exp\left[\frac{1}{2} x^T S x - m^T x + \frac{1}{2} m^T S^{-1} m - \frac{1}{2} \log |S|\right] dx$$

From the things mentioned earlier and exploiting the equivalency of the norms, this objective function is strongly convex:

- with the coefficient $\frac{\lambda_{min}^{\frac{n}{2}+2}}{2}$ in the norm $\|\cdot\|_2$;
- with the coefficient $\frac{\lambda_{min}^{\frac{n}{2}+2}}{2}$ in the norm $\|\cdot\|_\infty$;
- with the coefficient $\frac{\lambda_{min}^{\frac{n}{2}+2}}{2\sqrt{n}}$ in the norm $\|\cdot\|_1$.

Strong convexity of the second problem

Problem

$$\min_{m,S} \int \varphi^2(x) f^2(x) \sqrt{(2\pi)^n} \exp\left[\frac{1}{2} x^T S x - m^T x + \frac{1}{2} m^T S^{-1} m - \frac{1}{2} \log |S|\right] dx$$

This objective function is strongly convex:

- with the coefficient $\frac{\lambda_{min}^{\frac{n}{2}+2}}{4}$ in the spectral norm;
- with the coefficient $\frac{\lambda_{min}^{\frac{n}{2}+2}}{4n}$ in the nuclear norm;