Importance Sampling

Alena Shilova

Skoltech

May 30,2017

Strong Convexity

For differential functions one can formulate the following definition of strong convexity:

$$\forall x, y \hookrightarrow \langle \nabla f(x) - \nabla f(y), (x - y) \rangle \geq \alpha \|x - y\|^2$$

Let's check if $\exp(A(\theta) - \theta^T T(x))$ is strongly convex function for Gaussian distribution for some Θ, i.e. if it satisfies the inequality above

Strong convexity

For Gaussian function:

Let $C = \frac{1}{\sqrt{(2\pi)^n}}$ and $u_i(x) = \frac{\sqrt{|\Sigma_i|}}{\exp\left(\frac{1}{2}(x-\mu_i)^T\Sigma_i^{-1}(x-\mu_i)\right)}$ then

$$
\begin{aligned}
&\left(Cu_1(\Sigma_1^{-1}\mu_1 - \Sigma_1^{-1}x) - Cu_2(\Sigma_2^{-1}\mu_2 - \Sigma_2^{-1}x)\right)^T(\mu_1 - \mu_2) + \\
&+\operatorname{tr}\Big((Cu_1(\frac{1}{2}\Sigma_1^{-1} - \frac{1}{2}\Sigma_1^{-1}(x-\mu_1)(x-\mu_1)^T\Sigma_1^{-1}) - \\
&-Cu_2(\frac{1}{2}\Sigma_2^{-1} - \frac{1}{2}\Sigma_2^{-1}(x-\mu_2)(x-\mu_2)^T\Sigma_2^{-1}))^T(\Sigma_1 - \Sigma_2)\Big) \geq \\
&\geq \alpha\|\mu_1 - \mu_2\|_2^2 + \alpha\|\Sigma_1 - \Sigma_2\|_F^2
\end{aligned}
$$

First simplification

Let's suppose that we have an optimal μ from beforehand, therefore let's put $\mu_1 = \mu_2$, then it would be

$$\begin{aligned} &\operatorname{tr}\Big(Cu_1(\frac{1}{2}I - \frac{1}{2}\Sigma_1^{-1}\Sigma_2 - \frac{1}{2}\Sigma_1^{-1}xx^T + \frac{1}{2}\Sigma_1^{-1}xx^T\Sigma_1^{-1}\Sigma_2) - \\ -&Cu_2(-\frac{1}{2}I + \frac{1}{2}\Sigma_2^{-1}\Sigma_1 + \frac{1}{2}\Sigma_2^{-1}xx^T - \frac{1}{2}\Sigma_2^{-1}xx^T\Sigma_2^{-1}\Sigma_1)\Big) \geq \\ \geq &\alpha\|\Sigma_1 - \Sigma_2\|_F^2 \end{aligned}$$

We want this inequality to be true for any $\Sigma_1, \Sigma_2 \in \mathbf{S}$ for some $\mathbf{S}$ and $\alpha(x) \geq 0$

Second Simplification

Let's check if we can get a strongly convex function if we do the next simplifications:

One more

Let's consider Σ such that $\exists\vec{\lambda} : \Sigma = \mathrm{diag}(\vec{\lambda})$

Still more

Let our covariation matrix may be written as $\Sigma = \lambda I$

Result

Due to the mentioned simplifications, one can come to the following inequality:

$$\left(\frac{n^2}{2} - n\right)\lambda^{\frac{n}{2}-2}\exp\left(\frac{x^T x}{2\lambda}\right) - (n-2)\lambda^{\frac{n}{2}-3}x^T x\exp\left(\frac{x^T x}{2\lambda}\right)$$
$$+\frac{1}{2}\lambda^{\frac{n}{2}-4}(x^T x)^2\exp\left(\frac{x^T x}{2\lambda}\right) \geq \alpha n$$

For $n = 2$:

$$\frac{1}{2}\lambda^{-3}(x^T x)^2\exp\left(\frac{x^T x}{2\lambda}\right) \geq \frac{1}{2}\lambda_{max}^{-3}(x^T x)^2\exp\left(\frac{x^T x}{2\lambda_{max}}\right) \geq$$
$$\geq \frac{1}{2}\lambda_{max}^{-3}(x^T x)^2 \geq 2\alpha$$

If we take $\alpha = \frac{C}{4}(x^T x)^2$, then we have the following constraint for strong convexity $\lambda_{max} \leq \frac{1}{C^{\frac{1}{3}}}$

Closing remark

If we have a proper $\alpha(x)$ and a set $\boldsymbol{\Theta}$, then for all $\theta_1, \theta_2 \in \boldsymbol{\Theta}$ we will have:

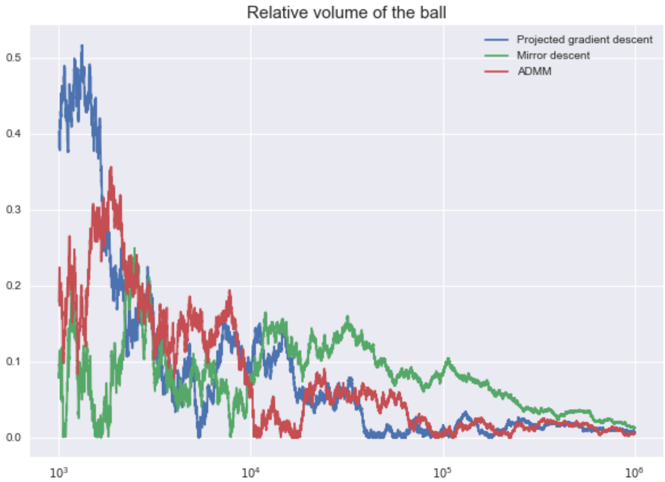
$$\int \exp(A(t(\theta_1) + (1-t)\theta_2) - (t\theta_1 + (1-t)\theta_2)^T T(x)) f^2(x)\phi^2(x)h(x)\, dx \leq$$

$$\leq t \int \exp(A(\theta_1) - \theta_1^T T(x)) f^2(x)\phi^2(x)h(x)\, dx +$$

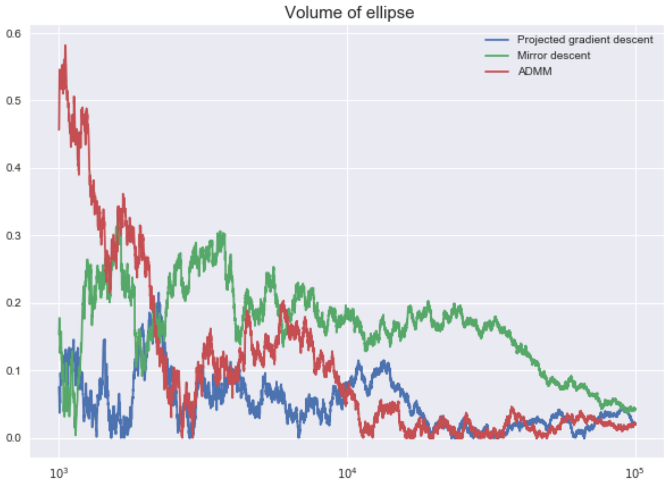
$$+ (1-t) \int \exp(A(\theta_2) - \theta_2^T T(x)) f^2(x)\phi^2(x)h(x)\, dx -$$

$$- t(1-t) \int \alpha(x) f^2(x)\phi^2(x)h(x)\, dx \|\theta_1 - \theta_2\|_2^2$$

About algotithms



About algotithms



About algotithms