On Nonconvex SVGD

Lukang Sun

March 26, 2021

1 Setup

$$F = \frac{1}{n}\sum_{i=1}^{n} f_i, \qquad \pi \sim e^{-F}$$

$$f_n = \frac{1}{n_0}\sum_{i\in I} f_i, \qquad \#I = n_0, \qquad \sigma_n \sim e^{-f_n}$$

$$\phi_t(x) = \left(I - tP_{\mu_n}\nabla \log\left(\frac{\mu_n}{\sigma_n}\right)\right)(x), \qquad \mu_t = (\phi_t)_{\#}\,\mu_n$$

$$\varphi(t) = \int \log\left(\frac{\mu_t}{\pi}\right) d\mu_t$$

2 PAGE

We define the update rule as:

$$S_{\mu_{n+1}}\nabla\log\left(\frac{\mu_{n+1}}{\sigma_{n+1}}\right) = \begin{cases} S_{\mu_{n+1}}\nabla\log\left(\frac{\mu_{n+1}}{\pi}\right) & p, \\ S_{\mu_n}\nabla\log\left(\frac{\mu_n}{\sigma_n}\right) + S_{\mu_{n+1}}\nabla\log\left(\frac{\mu_{n+1}}{\sigma_{I'}}\right) - S_{\mu_n}\nabla\log\left(\frac{\mu_n}{\sigma_{I'}}\right) & 1-p, \end{cases}$$

where $\sigma_{I'} \sim \exp^{-f_{I'}}$, $f_{I'} = \frac{1}{b'}\sum_{i\in I'} f_i$.

3 Main Proof

$$\phi(\gamma) \leq \phi(0) + \gamma\left\langle S_{\mu_n}\nabla\log\frac{\mu_n}{\pi}, S_{mu_n}\nabla\log\left(\frac{\mu_n}{\sigma_n}\right)\right\rangle_H + \frac{C\cdot\gamma^2}{2}\left\langle S_{\mu_n}\nabla\log\frac{\mu_n}{\sigma_n}, S_{\mu_n}\nabla\log\left(\frac{\mu_n}{\sigma_n}\right)\right\rangle_H$$

then by similar calculation as in PAGE, we have,

$$\begin{aligned}\varphi(\gamma) &\leqslant \varphi(0) - \frac{\gamma}{2}\left\|S_{\mu_n}\nabla\log\left(\frac{\mu_n}{\pi}\right)\right\|_H^2 - \left(\frac{1}{2\gamma} - \frac{c}{2}\right)\cdot\gamma^2\left\|S_{\mu_n}\nabla\log\left(\frac{\mu_n}{\sigma_n}\right)\right\|_H^2 \\ &+ \frac{\gamma}{2}\left\|S_{\mu_n}\nabla\log\left(\frac{\mu_n}{\pi}\right) - S_{\mu_n}\log\left(\frac{\mu_n}{\sigma_n}\right)\right\|_H^2\end{aligned}$$

Let

$$G_{n+1} := \mathbf{E}\left[\left\|S_{\mu_{n+1}}\nabla\log\left(\frac{\mu_{n+1}}{\pi}\right) - S_{\mu_{n+1}}\nabla\log\left(\frac{\mu_{n+1}}{\sigma_{n+1}}\right)\right\|_H^2\right]$$

Then a direct calculation now reveals that

$$
\begin{aligned}
G_{n+1} &= p \cdot 0 + (1-p) \cdot \mathrm{E} \left\| S_{\mu_n} \nabla \log\left(\frac{\mu_n}{\sigma_n}\right) + S_{\mu_{n+1}} \nabla \log\left(\frac{\mu_{n+1}}{\sigma_{I'}}\right) - S_{\mu_n} \nabla \log\left(\frac{\mu_n}{\sigma_{I'}}\right) - S_{\mu_{n+1}} \nabla \log\left(\frac{\mu_{n+1}}{\pi}\right) \right\|_H^2 \\
&= (1-p) \cdot \mathrm{E} \left\| S_{\mu_n} \nabla \log\left(\frac{\mu_n}{\sigma_n}\right) + S_{\mu_{n+1}} \nabla \log\left(\frac{\mu_{n+1}}{\sigma_{I'}}\right) - S_{\mu_n} \nabla \log\left(\frac{\mu_n}{\sigma_{I'}}\right) - S_{\mu_{n+1}} \nabla \log\left(\frac{\mu_{n+1}}{\pi}\right) + S_{\mu_n} \nabla \log\left(\frac{\mu_n}{\pi}\right) \right\|_H^2 \\
&\quad + (1-p) \cdot \left\| S_{\mu_n} \nabla \log\left(\frac{\mu_n}{\pi}\right) - S_{\mu_n} \nabla \log\left(\frac{\mu_n}{\sigma_I}\right) \right\|_H^2
\end{aligned}
$$

4 Key Inequality

Theorem 1 *There exists stepsize γ_n(in the following for simplicity we write as γ) only depending on $\int |\nabla \log(\pi(x))\mu_n(x)dx|$ and $\int |\nabla \log(\sigma_{I'}(x))|\mu_n(x)dx, such that \boldsymbol{E}(\Phi_{n+1}) \leq \boldsymbol{E}(\Phi_n) - \frac{\gamma}{2}\boldsymbol{E}\left\|S_{\mu_n}\nabla\log(\frac{\mu_n}{\pi})\right\|_H^2$, where*

$\Phi_n := KL(\mu_n|\pi) + \frac{\gamma}{2p}\left\|S_{\mu_n}\nabla\log(\frac{\mu_n}{\pi}) - S_{\mu_n}\nabla\log(\frac{\mu_n}{\sigma_n})\right\|_H^2$

Proof 1 *We need to estimate*

$$\left\|S_{\mu_{n+1}}\nabla\log\left(\frac{\mu_{n+1}}{\pi}\right) - S_{\mu_n}\nabla\log\left(\frac{\mu_n}{\pi}\right)\right\|_H^2$$

and

$$\left\|S_{\mu_{n+1}}\nabla\log\left(\frac{\mu_{n+1}}{\sigma_{I'}}\right) - S_{\mu_n}\nabla\log\left(\frac{\mu_n}{\sigma_{I'}}\right)\right\|_H^2$$

$S_{\mu_{n+1}}\nabla\log\left(\frac{\mu_{n+1}}{\pi}\right) - S_{\mu_n}\nabla\log\left(\frac{\mu_n}{\pi}\right)$
$= \int k(y,\cdot)\nabla\log(\mu_{n+1}(y))\mu_{n+1}(y)dy - \int k(\phi_\gamma(x),\cdot)\nabla\log(\mu_n(x))\mu_n(x)dx$
$+ \int k(\phi_\gamma(x),\cdot)\nabla\log(\pi(x))\mu_n(x)dx - \int k(y,\cdot)\nabla\log(\pi(y))\mu_{n+1}(y)dy + \int k(\phi_\gamma(x),\cdot)\nabla\log(\frac{\mu_n(x)}{\pi(x)})\mu_n(x)dx - \int k(x,\cdot)\nabla\log(\frac{\mu_n(x)}{\pi(x)})\mu_n(x)dx = I + II + III$
$\|II\|_H^2 = \int\int k(\phi_\gamma(x),\phi_\gamma(x'))(\nabla\log(x) - \nabla\log(\pi(\phi_\gamma(x))))(\nabla\log(x') - \nabla\log(\pi(\phi_\gamma(x'))))\mu(x)\mu(x')dxdx'$
since F is L-smooth, so
$|\nabla\log(x') - \nabla\log(\pi(\phi_\gamma(x')))| \leq L\cdot|x-\phi_\gamma(x)| = L\cdot\gamma\cdot|P_{\mu_\gamma}\nabla\log(\frac{\mu_n(x)}{\sigma_n(x)})| \leq C\cdot\gamma\cdot\left\|S_{\mu_n}\nabla\log(\frac{\mu_n}{\sigma_n})\right\|_H$
then by Cauchy-Schwartz inequality and $k(x,x)$ is bounded, we derive

$$\|II\|_H^2 \leq C\cdot\gamma^2\cdot\left\|S_{\mu_n}\nabla\log(\frac{\mu_n}{\sigma_n})\right\|_H^2$$

$\|I\|_H^2 = \left\|\int k(y,\cdot)\nabla\log(\mu_{n+1}(y))\mu_{n+1}(y)dy - \int k(\phi_\gamma(x),\cdot)\nabla\log(\mu_n(x))\mu_n(x)dx\right\|_H^2$
$= \int\int k(y,y')\nabla\log(\mu_{n+1}(y))\nabla\log(\mu_{n+1}(y'))\mu_{n+1}(y)\mu_{n+1}(y')dydy'$
$- \int\int k(y,\phi_\gamma(x))\nabla\log(\mu_{n+1}(y))\nabla\log(\mu_n(x'))\mu_{n+1}(y)\mu_n(x)dydx'$
$- \int\int k(\phi_\gamma(x),y')\nabla\log(\mu_{n+1}(y'))\nabla\log(\mu_n(x))\mu_{n+1}(y')\mu_n(x)dy'dx$
$+ \int\int k(\phi_\gamma(x),\phi_\gamma(x'))\nabla\log(\mu_n(x))\nabla\log(\mu_n(x'))\mu_n(x)\mu_n(x')dxdx'$
$= \int\int tr\nabla_{1,2}^2k(y,y')\mu_{n+1}(y)\mu_{n+1}(y')dydy'$
$- \int\int tr(\nabla_{1,2}^2k(y,\phi_\gamma(x'))J\phi_\gamma(x'))\mu_{n+1}(y)\mu_n(x')dydx'$
$- \int\int tr(\nabla_{1,2}^2k(\phi_\gamma(x),y')J\phi_\gamma(x))\mu_n(x)\mu_{n+1}(y')dxdy'$
$+ \int\int tr(\nabla_{1,2}^2K(\phi_\gamma(x),\phi_\gamma(x'))J\phi_\gamma(x)J\phi_\gamma(x'))\mu_n(x)\mu_n(x')dxdx'$
$= \int\int tr(\nabla_{1,2}^2K(\phi_\gamma(x),\phi_\gamma(x'))(I-J\phi_\gamma(x))(I-J\phi_\gamma(x')))\mu_n(x)\mu_n(x')dxdx'$
assume $|\nabla_{1,2}^2K(\cdot,\cdot)| \leq C$, then $tr(\nabla_{1,2}^2K(\phi_\gamma(x),\phi_\gamma(x'))(I-J\phi_\gamma(x))(I-J\phi_\gamma(x'))) \leq C\cdot\gamma^2\cdot\left\|S_{\mu_n}\nabla\log(\frac{\mu_n}{\sigma_n})\right\|_H^2$
so $\|I\|_H^2 \leq C\cdot\gamma^2\cdot\left\|S_{\mu_n}\nabla\log(\frac{\mu_n}{\sigma_n})\right\|_H^2$

$\|III\|_H^2 = \int\int (k(\phi_\gamma(x),\phi_\gamma(x')) - k(\phi_\gamma(x),x') - k(x,\phi_\gamma(x')) + k(x,x'))\nabla\log(\frac{\mu_n(x)}{\pi(x)})\nabla\log(\frac{\mu_n(x')}{\pi(x')})$
$\mu_n(x)\mu_n(x')dxdx'$.
denote $F(x,x') := k(\phi_\gamma(x),\phi_\gamma(x')) - k(\phi_\gamma(x),x') - k(x,\phi_\gamma(x')) + k(x,x')$
then by integration by part, we have
$\|III\|_H^2 = \int\int (tr(\nabla^2_{1,2}F(x,x')) + \nabla_2 F(x,x')\nabla\log(\pi(x)) + \nabla_1 F(x,x')\nabla\log(\pi(x')) + F(x,x')\nabla\log(\pi(x))\nabla\log(\pi(x')))\mu_n(x)\mu$
We need to know the order of $F(x,x'), \nabla_1 F(x,x'), \nabla_2 F(x,x'), \nabla^2_{1,2}F(x,x')$
in the following we assume supremum norm the second,third and fourth order derivatives of k *are finite, and we miss* $O(|\phi_\gamma(x)-x|^2)$ *and* $O(|\phi_\gamma(x')-x'|^2)$ *in the derivation.*

$|F(x,x')| = |<\phi_\gamma(x'-x'), \nabla_2 k(\phi_\gamma(x),x')> + <x'-\phi_\gamma(x')), \nabla_2 k(x,x')>|$
$\leq |\nabla^2_{1,2}k(x,x')(\phi_\gamma(x)-x)(\phi_\gamma(x')-x')| \leq C\cdot\gamma^2\cdot\left\|S_{\mu_n}\nabla\log(\frac{\mu_n}{\sigma_n})\right\|_H^2$
$|\nabla_1 F(x,x')| = |\nabla_1 k(\phi_\gamma(x),\phi_\gamma(x'))J\phi_\gamma(x) - \nabla_1 k(\phi_\gamma(x),x')J\phi_\gamma(x) - \nabla_1 k(x,\phi_\gamma(x')) + \nabla_1 k(x,x')|$
$= |\nabla^2_{1,2}k(\phi_\gamma(x),x')(\phi_\gamma(x'-x'))J\phi_\gamma(x) - \nabla^2_{1,2}k(x,x')(\phi_\gamma(x')-x')|$
$= |\nabla^3_{11,2}k(x,x')(\phi_\gamma(x')-x')(\phi_\gamma(x)-x) + \nabla^2_{1,2}k(\phi_\gamma(x),x)(\phi_\gamma(x')-x')(J\phi_\gamma(x)-I)| \leq C\cdot\gamma^2\left\|S_{\mu_n}\nabla\log(\frac{\mu_n}{\sigma_n})\right\|_H^2$
similarly $|\nabla_2 F(x,x')| \leq C\cdot\gamma^2\left\|S_{\mu_n}\nabla\log(\frac{\mu_n}{\sigma_n})\right\|_H^2$
$tr\nabla^2_{1,2}F(x,x') = tr(J\phi_\gamma(x)\nabla^2_{1,2}k(\phi_\gamma(x),\phi_\gamma(x'))J\phi_\gamma(x') - J\phi_\gamma(x)\nabla^2_{1,2}k(\phi_\gamma(x),x') - \nabla^2_{1,2}k(x,\phi_\gamma(x'))J\phi_\gamma(x') + \nabla^2_{1,2}k(x,x'))$,
denote $J\phi_\gamma(x') = I - \gamma JP_{\mu_n}\nabla\log(\frac{\mu_n}{\sigma_n})(x') = I - \gamma P', J\phi_\gamma(x) = I - \gamma JP_{\mu_n}\nabla\log(\frac{\mu_n}{\sigma_n})(x) = I - \gamma P$
then
$tr\nabla^2_{1,2}F(x,x') = tr((I-\gamma P)\nabla^3_{1,22}k(\phi_\gamma(x),x')(\phi_\gamma(x')-x') - (I-\gamma P)\nabla^2_{1,2}k(\phi_\gamma(x),\phi_\gamma(x'))(\gamma P')$
$-\nabla^3_{1,22}k(x,x')(\phi_\gamma(x')-x') + (\gamma P')\nabla^2_{1,2}k(x,\phi_\gamma(x')))$
$= tr(\nabla^4_{11,22}k(x,x')(\phi_\gamma(x)-x)(\phi_\gamma(x')-x') - (\gamma P')\nabla^3_{11,2}k(x,\phi_\gamma(x'))(\phi_\gamma(x)-x) - (\gamma P)\nabla^3_{1,22}k(\phi_\gamma(x),x')(\phi_\gamma(x')-x') + (\gamma P)\nabla^2_{1,2}k(x,x')(\gamma P'))$
$|tr\nabla^2_{1,2}F(x,x')| \leq C\cdot\gamma^2\left\|S_{\mu_n}\nabla\log(\frac{\mu_n}{\sigma_n})\right\|_H^2$
then, we know,
$\|III\|_H^2 \leq (C_1 + C_2\cdot\int|\nabla\log(\pi(x))|\mu_n(x)dx + C_3\cdot(\int|\nabla\log(\pi(x))|\mu_n(x)dx)^2)\cdot\gamma^2\left\|S_{\mu_n}\nabla\log(\frac{\mu_n}{\sigma_n})\right\|_H^2$
similarly, since the set of $\{\sigma_{I'}\}\}$ *is finite and have similar property with* π*, so*
$\left\|S_{\mu_{n+1}}\nabla\log(\frac{\mu_{n+1}}{\sigma_{I'}}) - S_{\mu_n}\nabla\log(\frac{\mu_n}{\sigma_{I'}})\right\|_H^2 \leq (C_1' + C_2'\cdot\int|\nabla\log(\sigma_{I'}(x))|\mu_n(x)dx + C_3'\cdot(\int|\nabla\log(\sigma_{I'}(x))|\mu_n(x)dx)^2)\cdot\gamma^2\left\|S_{\mu_n}\nabla\log(\frac{\mu_n}{\sigma_n})\right\|_H^2$
Finally,
$\boldsymbol{E}\left\|S_{\mu_{n+1}}\nabla\log(\frac{\mu_{n+1}}{\pi}) - S_{\mu_{n+1}}\nabla\log(\frac{\mu_{n+1}}{\sigma_{n+1}})\right\|_H^2 \leq (1-p)\cdot\left\|S_{\mu_n}\nabla\log(\frac{\mu_n}{\pi}) - S_{\mu_n}\nabla\log(\frac{\mu_n}{\sigma_n})\right\|_H^2 + (C_1 + C_2\cdot\int|\nabla\log(\pi(x))|\mu_n(x)dx + C_3\cdot(\int|\nabla\log(\pi(x))|\mu_n(x)dx)^2 + C_1' + C_2'\cdot\int|\nabla\log(\sigma_{I'}(x))|\mu_n(x)dx + C_3'\cdot(\int|\nabla\log(\sigma_{I'}(x))|\mu_n(x)dx)^2)\cdot\gamma^2\left\|S_{\mu_n}\nabla\log(\frac{\mu_n}{\sigma_n})\right\|_H^2$
So if γ *is small enough,s.t* $\frac{\gamma}{2p}\cdot(C_1 + C_2\cdot\int|\nabla\log(\pi(x))|\mu_n(x)dx + C_3\cdot(\int|\nabla\log(\pi(x))|\mu_n(x)dx)^2 + C_1' + C_2'\cdot\int|\nabla\log(\sigma_{I'}(x))|\mu_n(x)dx + C_3'\cdot(\int|\nabla\log(\sigma_{I'}(x))|\mu_n(x)dx)^2) - (\frac{1}{2\gamma} - \frac{C}{2}) \leq 0$*, then*
$\boldsymbol{E}(\Phi_{n+1}) \leq \boldsymbol{E}(\Phi_n) - \frac{\gamma}{2}\boldsymbol{E}\left\|S_{\mu_n}\nabla\log(\frac{\mu_n}{\pi})\right\|_H^2$,

where
$\Phi_n := KL(\mu_n|\pi) + \frac{\gamma}{2p}\left\|S_{\mu_n}\nabla\log(\frac{\mu_n}{\pi}) - S_{\mu_n}\nabla\log(\frac{\mu_n}{\sigma_n})\right\|_H^2$

Lemma 1 *If there exists a constant* $a > 0$ *and a point* x_0 *such that* $\int e^{ad(x_0,x)^2}\pi(dx) < +\infty$*, then we have we have* T_1 *inequality*

$$W_1(\mu,\pi) \leq C\cdot\sqrt{KL(\mu|\pi)}$$

.

Proof 2 *see "optimal transport old and new" page 592.*

We assume there exists a constant $a > 0$ and a point x_0 such that $\int e^{ad(x_0,x)^2}\pi(dx) < +\infty$, so the last lemma holds.

Lemma 2 *if we have $\Phi_n \leq C$, then we have $\int |\nabla \log(\pi(x))|\mu_n(x)dx \leq \tilde{C}$*

Proof 3 *$KL(\mu_n|\pi) \leq \Phi_n \leq C$, so $W_1(\mu_n, \pi) \leq C \cdot \sqrt{KL(\mu_n, \pi)} \leq C_1$, so by L-smoothness of F, we have $\int |\nabla \log(\pi(x))|\mu_n(x)dx \leq C \int |x|\mu_n(x)dx = CW_1(\mu_n, \delta_0) \leq C(W_1(\mu_n, \pi) + W_1(\pi, \delta)) \leq \tilde{C}$*

Lemma 3 *if $\Phi_n \leq A$, then we can choose$\gamma_n \geq \gamma_0$, where γ_0 depend on A(as you see in proof of the main theorem, it depends on $\int ||\nabla \log(\pi)\mu_n dx$, as you see in the last lemma ,$\int ||\nabla \log(\pi)\mu_n dx$ depends on Φ_n), such that $\Phi_{n+1} \leq A$*

Proof 4 *(note $\mu_{n+1} = (I - \gamma_n P_{\mu_n} \nabla \log(\frac{\mu_n}{\sigma_n}))_\sharp \mu_n$, so you see μ_n is independent of γ_n,but depend on γ_{n-1}).from the last lemma we can choose $\gamma_n \geq \gamma_0$, such that $\Phi_{n+1} \leq \Phi_n - \frac{\gamma_n}{2} I^2_{stein}(\mu_n|\pi)$, so $KL(\mu_{n+1}|\pi) \leq \Phi_{n+1} \leq \Phi_n \leq A$*

Theorem 2 *if we set a constant A, s.t $\Phi_0 \leq A$, then we can set $\gamma_n = \gamma_0$, for all n, such that,*

$\Phi_{n+1} \leq \Phi_n - \frac{\gamma_0}{2} I^2_{Stein}(\mu_n|\pi)$

Proof 5 *we can use the last lemma recursively, so finish the proof.*