Constraint Relaxation for Bayesian Modeling with Parameter Constraints

Leo Duan, Alex Young, Akihiko Nishimura, David Dunson

Let the approximating function for $\{v_j(\theta)=0\}_{j=1}^s$ take the form $\exp(-\frac{\|v(\theta)\|_1}{\lambda})=\exp(-\frac{\sum_j|v_j(\theta)|}{\lambda})$, with $\lambda>0$ a scalar. Letting $\bar{\mathcal{H}}^{p-s}(d\theta)$ be the $(p-s)$-dimensional Hausdorff measure, we assume $\mathcal{R}$ is chosen so that $0<m(x)=\int_{v^{-1}(x)}\frac{\pi_{\mathcal{R}}(\theta)}{Jv(\theta)}\bar{\mathcal{H}}^{p-s}(d\theta)<\infty$ for any $x\in\mathcal{X}=\{v(\theta);\theta\in\mathcal{R}\}$ and $\mathbf{0}\in\mathcal{X}$.

For any given set $A\subseteq\mathcal{R}$, the exact and approximate measures are

$$
\begin{aligned}
\Pi(A) &= \int_{\mathbb{R}^s}\frac{1}{m_0}\left[\int_{v^{-1}(x)\cap A}\frac{\pi_{\mathcal{R}}(\theta)}{Jv(\theta)}\bar{\mathcal{H}}^{p-s}(d\theta)\right]\delta_{\mathbf{0}}(x)\mu^s(dx)\\
\tilde{\Pi}(A) &= \int_{\mathbb{R}^s}\frac{1}{m_\lambda}\left[\int_{v^{-1}(x)\cap A}\frac{\pi_{\mathcal{R}}(\theta)}{Jv(\theta)}\bar{\mathcal{H}}^{p-s}(d\theta)\right]\exp(-\frac{\|x\|_1}{\lambda})\mathbb{1}_{x\in\mathcal{X}}\mu^s(dx)
\end{aligned}
\tag{1}
$$

where $\delta_{\mathbf{0}}(x)$ is a Dirac measure at $\mathbf{0}$; $\mu^s(dx)$ is the s-dimensional Lebesgue measure; m_0 and m_λ are normalizing constant so that $\Pi(\mathcal{R})=1$ and $\tilde{\Pi}(\mathcal{R})=1$.

Similarly, the expectations of an integrable function $g(\theta)$ are

$$
\begin{aligned}
\mathbb{E}_{\Pi}(g(\theta)) &= \int_{\mathbb{R}^s}\frac{1}{m_0}\left[\int_{v^{-1}(x)}g(\theta)\frac{\pi_{\mathcal{R}}(\theta)}{Jv(\theta)}\bar{\mathcal{H}}^{p-s}(d\theta)\right]\delta_{\mathbf{0}}(x)\mu^s(dx)\\
\mathbb{E}_{\tilde{\Pi}}(g(\theta)) &= \int_{\mathbb{R}^s}\frac{1}{m_\lambda}\left[\int_{v^{-1}(x)}g(\theta)\frac{\pi_{\mathcal{R}}(\theta)}{Jv(\theta)}\bar{\mathcal{H}}^{p-s}(d\theta)\right]\exp(-\frac{\|x\|_1}{\lambda})\mathbb{1}_{x\in\mathcal{X}}\mu^s(dx)
\end{aligned}
\tag{2}
$$

We denote integral $h(g;x)=\int_{v^{-1}(x)}g(\theta)\frac{\pi_{\mathcal{R}}(\theta)}{Jv(\theta)}\bar{\mathcal{H}}^{p-s}(d\theta)$, which is related to conditional expectation $\mathbb{E}(g(\theta)\mid v(\theta)=x)=\frac{1}{m(x)}h(g;x)$.

Remark 1. *The 1-Wasserstein distance between the measures* Π *and* $\tilde{\Pi}$ *has*

$$\lim_{\lambda\to 0}W_1(\Pi,\tilde{\Pi})=0.$$

Further,

$$W_1(\Pi,\tilde{\Pi})\le(2\lambda)^s(\frac{k_1k_3}{m_0^2}+\frac{k_1}{m_0})+\exp(-\lambda^{-1}t)(\frac{k_1}{m_0^2}+\frac{k_2}{m_0}),\tag{3}$$

where $k_1 = \sup_{g:\|g\|_L \le 1} \sup_{x^*:\|x^*\|_1 \in [0,t]\cap\mathcal{X}} |h(g;x^*)|$, $k_2 = \sup_{g:\|g\|_L\le 1} \mathbb{E}(|g(\theta)|)$ *with expectation taken over* $\pi_{\mathcal{R}}(\theta)$, $k_3 = \sup_{x^*:\|x^*\|_1\in(0,t]\cap\mathcal{X}} m(x^*)$.

Proof. Let $g:\mathbb{R}^p \to \mathbb{R}$ be a 1-Lipschitz continuous function, i.e. $\|g(x)-g(y)\| \le \|x-y\|$, denoted by $\|g\|_L \le 1$. By Kantorovich-Rubinstein duality, the 1-Wasserstein distance based on Euclidean metric equals to:

$$W_1(\Pi,\tilde{\Pi}) = \sup_{g:\|g\|_L\le 1} \mathbb{E}_{\Pi}(g(\theta)) - \mathbb{E}_{\tilde{\Pi}}(g(\theta)) \tag{4}$$

The two normalizing constants are:

$$\begin{aligned} m_0 &= \int_{\mathbb{R}^s} m(x)\delta_{\mathbf{0}}(x)\mu^s(dx),\\ m_\lambda &= \int_{\mathbb{R}^s} m(x)\exp(-\frac{\|x\|_1}{\lambda})\mathbb{1}_{x\in\mathcal{X}}\mu^s(dx). \end{aligned} \tag{5}$$

Noting $\delta_{\mathbf{0}}(x) = \lim_{\lambda_0\to 0}\exp(-\frac{\|x\|_1}{\lambda_0})$, we have $\exp(-\frac{\|x\|_1}{\lambda_0}) \le \exp(-\frac{\|x\|_1}{\lambda})$ for any $\lambda \ge \lambda_0$, therefore $m_0 \le m_\lambda$.

We have

$$\begin{aligned} \sup_{g:\|g\|_L\le 1} \mathbb{E}_{\Pi}(g(\theta)) - \mathbb{E}_{\tilde{\Pi}}(g(\theta)) &= \sup_{g:\|g\|_L\le 1} \mathbb{E}_{\tilde{\Pi}}(g(\theta)) - \mathbb{E}_{\Pi}(g(\theta))\\ &= \sup_{g:\|g\|_L\le 1} \int_{\mathbb{R}^s} h(g;x)\left[\frac{\exp(-\lambda^{-1}\|x\|_1)}{m_\lambda} - \frac{\delta_{\mathbf{0}}(x)}{m_0}\right]\mathbb{1}_{x\in\mathcal{X}}\mu^s(dx)\\ &= \sup_{g:\|g\|_L\le 1} \int_{\mathbb{R}^s} h(g;x)\left[\mathbb{1}_{x=\mathbf{0}}(\frac{1}{m_\lambda} - \frac{1}{m_0}) + \mathbb{1}_{x\in\mathcal{X}\setminus\mathbf{0}}(\frac{\exp(-\lambda^{-1}\|x\|_1)}{m_\lambda})\right]\mu^s(dx)\\ &\le \sup_{g:\|g\|_L\le 1} \int_{\mathbb{R}^s} |h(g;x)|\left[\mathbb{1}_{x=\mathbf{0}}(\frac{1}{m_0} - \frac{1}{m_\lambda}) + \mathbb{1}_{x\in\mathcal{X}\setminus\mathbf{0}}(\frac{\exp(-\lambda^{-1}\|x\|_1)}{m_\lambda})\right]\mu^s(dx)\\ &\le \sup_{g:\|g\|_L\le 1} \int_{\mathbb{R}^s} |h(g;x)|\left[\mathbb{1}_{x=\mathbf{0}}(\frac{1}{m_0} - \frac{1}{m_\lambda}) + \frac{1}{m_0}\mathbb{1}_{x\in\mathcal{X}\setminus\mathbf{0}}(\exp(-\lambda^{-1}\|x\|_1))\right]\mu^s(dx) \end{aligned} \tag{6}$$

1. Asymptotic result:

Note $m_\lambda \le \int_{\mathbb{R}^s} m(x)\mathbb{1}_{x\in\mathcal{X}}\mu^s(dx) = \int_{\mathbb{R}^p}\pi_{\mathcal{R}}(\theta)\mathbb{1}_{v(\theta)\in\mathcal{X}}\mu^p(d\theta) = 1$. By dominated convergence theorem,

$$\lim_{\lambda\to 0} m_\lambda = \int_{\mathbb{R}^s} m(x)\lim_{\lambda\to 0}\exp(-\lambda^{-1}\|x\|_1)\mathbb{1}_{x\in\mathcal{X}}\mu^s(dx) = m_0. \tag{7}$$

Since

$$\sup_{g:\|g\|_L\le 1}\int_{\mathbb{R}^s}|h(g;x)|\left[\mathbb{1}_{x=\mathbf{0}}(\frac{1}{m_0}-\frac{1}{m_\lambda})+\mathbb{1}_{x\in\mathcal{X}\setminus\mathbf{0}}(\frac{\exp(-\lambda^{-1}\|x\|_1)}{m_0})\right]\mu^s(dx)$$
$$\le\int_{\mathbb{R}^s}\sup_{g:\|g\|_L\le 1}|h(g;x)|\left[\mathbb{1}_{x=\mathbf{0}}(\frac{1}{m_0}-\frac{1}{m_\lambda})+\mathbb{1}_{x\in\mathcal{X}\setminus\mathbf{0}}(\frac{\exp(-\lambda^{-1}\|x\|_1)}{m_0})\right]\mu^s(dx), \tag{8}$$

letting $q_\lambda = \sup_{g:\|g\|_L\le 1}|h(g;x)|\left[\mathbb{1}_{x=\mathbf{0}}(\frac{1}{m_0}-\frac{1}{m_\lambda})+\mathbb{1}_{x\in\mathcal{X}\setminus\mathbf{0}}(\frac{\exp(-\lambda^{-1}\|x\|_1)}{m_0})\right]$, for fixed x, a decreasing series $1\ge\lambda_1\ge\lambda_2\ge\ldots$, has monotone increasing $0\le q_1-q_{\lambda_1}\le q_1-q_{\lambda_2}$, by monotone convergence theorem, $\lim_{\lambda\to 0}\int[q_1(x)-q_\lambda(x)]dx=\int[q_1(x)-q_0(x)]dx$ hence $\lim_{\lambda\to 0}\int q_\lambda(x)dx=0$. Combining the results yields

$$\lim_{\lambda\to 0}W_1(\Pi,\tilde{\Pi})=0. \tag{9}$$

2. Non-asymptotic result:

$$\begin{aligned}
\frac{1}{m_0}-\frac{1}{m_\lambda}&\le\frac{m_\lambda-m_0}{m_0^2}\\
&=\frac{\int_{\mathbb{R}^s}m(x)\exp(-\lambda^{-1}\|x\|_1)\mathbb{1}_{x\in\mathcal{X}\setminus\mathbf{0}}\mu^s(dx)}{m_0^2}\\
&=\frac{1}{m_0^2}\left[\int_{\mathbb{R}^s}\mathbb{1}_{\|x\|_1\in(0,t]}m(x)\exp(-\lambda^{-1}\|x\|_1)\mathbb{1}_{x\in\mathcal{X}}\mu^s(dx)+\int_{\mathbb{R}^s}\mathbb{1}_{\|x\|_1\in(t,\infty)}m(x)\exp(-\lambda^{-1}\|x\|_1)\mathbb{1}_{x\in\mathcal{X}}\mu^s(dx)\right]\\
&\le\frac{1}{m_0^2}\left[\sup_{x^*:\|x^*\|_1\in(0,t]\cap\mathcal{X}}m(x^*)\int_{\mathbb{R}^s}\mathbb{1}_{\|x\|_1\in(0,t]}\exp(-\lambda^{-1}\|x\|_1)\mathbb{1}_{x\in\mathcal{X}}\mu^s(dx)\right.\\
&\left.+\exp(-\lambda^{-1}t)\int_{\mathbb{R}^s}\mathbb{1}_{\|x\|_1\in(t,\infty]}m(x)\mathbb{1}_{x\in\mathcal{X}}\mu^s(dx)\right]\\
&\le\frac{1}{m_0^2}\left[(2\lambda)^s\sup_{x^*:\|x^*\|_1\in(0,t]\cap\mathcal{X}}m(x^*)+\exp(-\lambda^{-1}t)\right]
\end{aligned} \tag{10}$$

where

$$\begin{aligned}
\int_{\mathbb{R}^s}\mathbb{1}_{\|x\|_1\in(0,t]}\exp(-\lambda^{-1}\|x\|_1)\mathbb{1}_{x\in\mathcal{X}}\mu^s(dx)&=\int_{\mathbb{R}^s}\mathbb{1}_{x\in\mathcal{X}}\mathbb{1}_{\|x\|_1\in(0,t]}\prod_{i=1}^s\exp(-\lambda^{-1}|x_i|)\mu^s(dx)\\
&\le\int_{\mathbb{R}^s}\prod_{i=1}^s\exp(-\lambda^{-1}|x_i|)\mu^s(dx)\\
&=(2\lambda)^s
\end{aligned} \tag{11}$$

$$
\begin{aligned}
&\sup_{g:\|g\|_L\le 1}\int_{\mathbb{R}^s}|h(g;x)|\exp(-\lambda^{-1}\|x\|_1)\mathbb{1}_{x\neq\mathbf{0}}\mathbb{1}_{x\in\mathcal{X}}\mu^s(dx)\\
&\le \sup_{g:\|g\|_L\le 1}\quad\sup_{x^*:\|x^*\|_1\in(0,t]\cap\mathcal{X}}|h(g;x^*)|\int_{\mathbb{R}^s}\mathbb{1}_{\|x\|_1\in(0,t]}\exp(-\lambda^{-1}\|x\|_1)\mathbb{1}_{x\in\mathcal{X}}\mu^s(dx)\\
&+\exp(-\lambda^{-1}t)\sup_{g:\|g\|_L\le 1}\int_{\mathbb{R}^s}\mathbb{1}_{\|x\|_1\in(t,\infty)}|h(g;x)|\mathbb{1}_{x\in\mathcal{X}}\mu^s(dx)\\
&\le \sup_{g:\|g\|_L\le 1}\quad\sup_{x^*:\|x^*\|_1\in(0,t]\cap\mathcal{X}}|h(g;x^*)|\int_{\mathbb{R}^s}\mathbb{1}_{\|x\|_1\in(0,t]}\exp(-\lambda^{-1}\|x\|_1)\mathbb{1}_{x\in\mathcal{X}}\mu^s(dx)\\
&+\exp(-\lambda^{-1}t)\sup_{g:\|g\|_L\le 1}\int_{\mathbb{R}^s}\mathbb{1}_{\|x\|_1\in(t,\infty)}h(|g|;x)\mathbb{1}_{x\in\mathcal{X}}\mu^s(dx)\\
&\le \sup_{g:\|g\|_L\le 1}\quad\sup_{x^*:\|x^*\|_1\in(0,t]\cap\mathcal{X}}|h(g;x^*)|(2\lambda)^s+\exp(-\lambda^{-1}t)\sup_{g:\|g\|_L\le 1}\mathbb{E}(|g(\theta)|\mathbb{1}_{v(\theta)\in\mathcal{X}}),
\end{aligned}
\tag{12}
$$

where we used $h(|g|;x)=\int_{v^{-1}(x)}|g(\theta)|\frac{\pi_{\mathcal{R}}(\theta)}{Jv(\theta)}\bar{\mathcal{H}}^{p-s}(d\theta)\ge h(g;x)$ and $\int_{\mathbb{R}^s}h(|g|;x)\mathbb{1}_{x\in\mathcal{X}}\mu^s(dx)=\mathbb{E}(|g(\theta)|)$ with expectation taken over $\pi_{\mathcal{R}}(\theta)$.

Combining (6)(10)(12), $k_1=\sup_{g:\|g\|_L\le 1}\sup_{x^*:\|x^*\|_1\in[0,t]\cap\mathcal{X}}|h(g;x^*)|$, $k_2=\sup_{g:\|g\|_L\le 1}\mathbb{E}(|g(\theta)|)$, $k_3=\sup_{x^*:\|x^*\|_1\in(0,t]\cap\mathcal{X}}m(x^*)$

$$
\begin{aligned}
&\sup_{g:\|g\|_L\le 1}\mathbb{E}_{\Pi}(g(\theta))-\mathbb{E}_{\tilde{\Pi}}(g(\theta))\\
&\le \sup_{g:\|g\|_L\le 1}\int_{\mathbb{R}^s}|h(g;x)|\left[\mathbb{1}_{x=\mathbf{0}}(\frac{1}{m_0}-\frac{1}{m_\lambda})+\mathbb{1}_{x\in\mathcal{X}\setminus\mathbf{0}}(\frac{\exp(-\lambda^{-1}\|x\|_1)}{m_0})\right]\mu^s(dx)\\
&\le \sup_{g:\|g\|_L\le 1}|h(g;\mathbf{0})|(\frac{1}{m_0}-\frac{1}{m_\lambda})+\sup_{g:\|g\|_L\le 1}\int_{\mathbb{R}^s}|h(g;x)|\mathbb{1}_{x\in\mathcal{X}\setminus\mathbf{0}}(\frac{\exp(-\lambda^{-1}\|x\|_1)}{m_0})\Big]\mu^s(dx)\\
&\le k_1\frac{1}{m_0^2}\left[(2\lambda)^s k_3+\exp(-\lambda^{-1}t)\right]+\frac{1}{m_0}\left[k_1(2\lambda)^s+\exp(-\lambda^{-1}t)k_2\right]\\
&\le (2\lambda)^s(\frac{k_1k_3}{m_0^2}+\frac{k_1}{m_0})+\exp(-\lambda^{-1}t)(\frac{k_1}{m_0^2}+\frac{k_2}{m_0})
\end{aligned}
\tag{13}
$$

□

The first part shows the asymptotic accuracy of the approximation. The second part shows the rate with non-asymptotic λ under mild assumptions. The interpretation for these assumptions is that if in a small space expansion of $\mathcal{D}$, defined as $\{\theta^*:\|v(\theta^*)\|_1\in[0,t]\}$, the marginal density of $v(\theta^*)$ and the integral of Lipschitz functions are bounded $k_1,k_2=\mathcal{O}(1)$, and the expected norm of Lipschitz function are smaller than a bound that grows near exponentially $k_3 = \mathcal{O}(\lambda \exp(t/\lambda))$, then the distance $W_1(\Pi, \tilde{\Pi})$ converges to 0 in $\mathcal{O}(\lambda^s)$ as $\lambda \to 0$.

References