NA-DPVI AISTATS 2025 Rebuttal

Anonymous

November 2024

1 NA-DPVI Algorithm

$(\boldsymbol{x}_i)_{i=1}^N \leftarrow$ data;
$(\epsilon, \delta) \leftarrow$ DP privacy level;
$C \leftarrow$ clipping threshold;
$T \leftarrow$ training iterations;
$\lambda \leftarrow$ learning rate;
$\kappa \leftarrow$ Poisson sampling rate;
$\boldsymbol{\phi}_0 \leftarrow$ initial values for $\boldsymbol{\phi}$;
$d \leftarrow \dim(\phi_0)$
$\sigma_{\mathrm{DP}} \leftarrow$ PRV-Accountant$(\epsilon, \delta, T, \kappa)$;

Algorithm 1 DPVI

for $t = 1, \ldots, T-1$ **do**
  $\boldsymbol{g}_{t+1} \leftarrow \sum_{i \in \mathcal{B}_{t+1}} \mathrm{clip}\left(\nabla_{\boldsymbol{\phi}} \ell(\boldsymbol{\phi}_t; \boldsymbol{x}_i), C\right)$;
  Sample $\boldsymbol{\eta}_{t+1} \sim \mathcal{N}(0, \mathbf{I}_d)$;
  $\widetilde{\boldsymbol{g}}_{t+1} \leftarrow \boldsymbol{g}_{t+1} + \sigma_{\mathrm{DP}} C \boldsymbol{\eta}_{t+1}$
  $\boldsymbol{\phi}_{t+1} \leftarrow \boldsymbol{\phi}_t - \lambda \widetilde{\boldsymbol{g}}$;
end for
return $\mathcal{T} = (\boldsymbol{\phi}_t)_{t=0}^T$, $\widetilde{\boldsymbol{g}} = (\widetilde{\boldsymbol{g}}_t)_{t=1}^T$;

Algorithm 2 NA-DPVI

$M \leftarrow$ number of samples;
$\boldsymbol{\mu}_{\boldsymbol{\phi}^*} \leftarrow \boldsymbol{\phi}^*$ prior distribution mean;
$\Sigma_{\boldsymbol{\phi}^*} \leftarrow \boldsymbol{\phi}^*$ prior distribution covariance matrix;
$\boldsymbol{\mu}_{\mathbf{A}} \leftarrow \mathbf{A}$ entries prior distribution mean;
$\Sigma_{\mathbf{A}} \leftarrow \mathbf{A}$ entries prior distribution covariance matrix;
$\boldsymbol{\mu}_{\Sigma_{\mathrm{sub}}} \leftarrow \Sigma_{\mathrm{sub}}$ entries prior distribution mean;
$\Sigma_{\Sigma_{\mathrm{sub}}} \leftarrow \Sigma_{\mathrm{sub}}$ entries prior distribution covariance matrix;
$\mathcal{T}, \widetilde{\boldsymbol{g}} \leftarrow$ DPVI;
model-definition $\widetilde{\boldsymbol{g}} \mid \mathcal{T}, \boldsymbol{\phi}^*, \mathbf{A}, \Sigma_{\mathrm{sub}}$
  $\boldsymbol{\phi}^* \sim \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{\phi}^*}, \Sigma_{\boldsymbol{\phi}^*})$;
  $\mathbf{A} \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{A}}, \Sigma_{\mathbf{A}})$;
  $\Sigma_{\mathrm{sub}} \sim \mathcal{N}(\boldsymbol{\mu}_{\Sigma_{\mathrm{sub}}}, \Sigma_{\Sigma_{\mathrm{sub}}})$;
  $\widetilde{\boldsymbol{g}}_{t+1} \mid \boldsymbol{\phi}_t, \mathbf{A}, \boldsymbol{\phi}^*, \Sigma_{\mathrm{sub}} \sim \mathcal{N}\left(\kappa \mathbf{A}\left(\boldsymbol{\phi}_t - \boldsymbol{\phi}^*\right), \sigma_{\mathrm{DP}}^2 C^2 \mathbf{I}_d + \Sigma_{\mathrm{sub}}\right)$;
end model-definition
Sample $(\boldsymbol{\phi}_i^*, \mathbf{A}_i, \Sigma_{\mathrm{sub},i})_{i=1}^M \sim \boldsymbol{\phi}^*, \mathbf{A}, \Sigma_{\mathrm{sub}} \mid \widetilde{\boldsymbol{g}}, \mathcal{T}$; ▷ using any approximate inference method.
return $\widetilde{p}(\boldsymbol{\theta} \mid \mathcal{T}) = \frac{1}{M} \sum_{i=1}^M q_{\mathrm{VI}}(\boldsymbol{\theta}; \boldsymbol{\phi}_i^*)$;

2 NA-DPVI vs Gibbs-SS-Noisy (Bernstein & Sheldon, 2019)

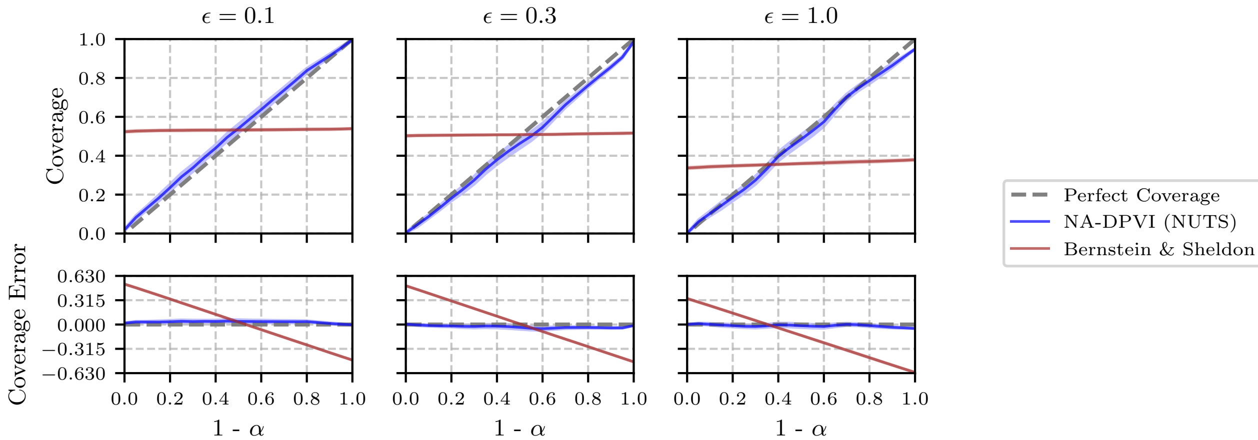

Figure 1: The first row in the figure shows the TARP coverages for the 10d Bayesian linear regression experiment for both NA-DPVI (NUTS) and Bernstein & Sheldon's method. The second row shows the error for the coverages $(C(\alpha) - (1 - \alpha))$. The solid lines show the average performance over 20 independent TARP repetitions and the error bars show the corresponding std. The parameters for NA-DPVI (NUTS) are $\delta = 10^{-5}$, $N = 5000$, $\kappa = 0.1$ and $T = 10^4$. On the other hand, the parameters for Gibbs-SS-Noisy (Bernstein & Sheldon, 2019) were $n = 10^4$, 2×10^4 burnin iterations and 2×10^4 posterior samples.

Table 1: The RMSE errors corresponding to the linear regression coverages in Figure 1.

Method	$\epsilon = 0.1$	$\epsilon = 0.3$	$\epsilon = 1.0$
NA-DPVI (NUTS)	0.036 ± 0.011	0.035 ± 0.011	0.027 ± 0.009
Bernstein & Sheldon	0.301 ± 0.001	0.299 ± 0.001	0.323 ± 0.003