Appendix
Curse of Dimensionality on Randomized Smoothing for Certifiable Robustness

Aounon Kumar [1] **Alexander Levine** [1] **Tom Goldstein** [1] **Soheil Feizi** [1]

A. Proof for lemma 3

Proof. Applying the series expansion of e^{tZ}, we get,

$$\begin{aligned}
E[e^{tZ}] &= \sum_{n=0}^{\infty} \frac{t^n E[Z^n]}{n!} \\
E[Z^n] &= \frac{1}{C}\int_{-\infty}^{\infty} z^n e^{-(|z|/b)^q} dz \\
&= \frac{1}{C}\int_{0}^{\infty} (1+(-1)^n) z^n e^{-z^q/b^q} dz \\
&= \begin{cases} 0, & n \text{ is odd} \\ \frac{2}{C}\int_0^\infty z^n e^{-z^q/b^q} dz, & n \text{ is even} \end{cases}
\end{aligned}$$

When n is even:

$$\begin{aligned}
E[Z^n] &= \frac{2}{C}\int_{0}^{\infty} z^n e^{-z^q/b^q} dz \\
&= \frac{2b^{n+1}\Gamma\left(\frac{n+1}{q}\right)}{Cq}
\end{aligned}$$

Substituting C,

$$\begin{aligned}
E[Z^n] &= \frac{b^n \Gamma\left(\frac{n+1}{q}\right)}{\Gamma(1/q)} \leq b^n \Gamma(n+1) \qquad \text{for } q \geq 1 \\
E[Z^n] &\leq b^n n!
\end{aligned}$$

Therefore, keeping only the terms with even n in the expansion of $E[e^{tZ}]$, we get:

$$\begin{aligned}
E[e^{tZ}] &\leq \sum_{m=0}^{\infty} (t^2 b^2)^m \\
&= \sum_{m=0}^{\infty} \left(\frac{t^2\sigma^2\Gamma(1/q)}{\Gamma(3/q)}\right)^m \quad \text{using } \sigma^2 = \tfrac{b^2\Gamma(3/p)}{\Gamma(1/p)} \\
&\leq \sum_{m=0}^{\infty} (c^2 t^2 \sigma^2)^m
\end{aligned}$$

for some positive constant $c < 1.85$, because,

$$\begin{aligned}
\frac{\Gamma(1/q)}{\Gamma(3/q)} &= \frac{3q\Gamma(1+1/q)}{q\Gamma(1+3/q)} \qquad (\text{using } \Gamma(z+1) = z\Gamma(z)) \\
&= \frac{3\Gamma(1+1/q)}{\Gamma(1+3/q)} \\
&< 1.85^2 \\
&\quad (\text{for } q \geq 1, \Gamma(1+1/q) \leq 1 \text{ and } \Gamma(1+3/q) > 0.88)
\end{aligned}$$

□

B. Proof for lemma 6

Proof. The points in V_1 satisfy the following 2^d constraints:

$$\begin{aligned}
x_1 + x_2 + \cdots + x_d &\leq b \\
-x_1 + x_2 + \cdots + x_d &\leq b \\
x_1 - x_2 + \cdots + x_d &\leq b \\
-x_1 - x_2 + \cdots + x_d &\leq b \\
&\vdots \\
-x_1 - x_2 - \cdots - x_d &\leq b
\end{aligned}$$

Similarly, points in V_2 satisfy,

$$\begin{aligned}
(x_1 - \epsilon) + x_2 + \cdots + x_d &\leq b \\
-(x_1 - \epsilon) + x_2 + \cdots + x_d &\leq b \\
(x_1 - \epsilon) - x_2 + \cdots + x_d &\leq b \\
-(x_1 - \epsilon) - x_2 + \cdots + x_d &\leq b \\
&\vdots \\
-(x_1 - \epsilon) - x_2 - \cdots - x_d &\leq b
\end{aligned}$$

Then, the points in $V_1 \cap V_2$ must satisfy the following set of constraints constructed by picking constraints that have a $+$ sign for x_1 in the first set of constraints and a $-$ sign for x_1

[1]University of Maryland, College Park, Maryland, USA. Correspondence to: Aounon Kumar <aounon@umd.edu>, Soheil Feizi <sfeizi@cs.umd.edu>.

Proceedings of the 37th International Conference on Machine Learning, Online, PMLR 119, 2020.

in the second set.

$$\begin{aligned} x_1 + x_2 + \cdots + x_d &\leq b \\ -(x_1 - \epsilon) + x_2 + \cdots + x_d &\leq b \\ x_1 - x_2 + \cdots + x_d &\leq b \\ -(x_1 - \epsilon) - x_2 + \cdots + x_d &\leq b \\ &\vdots \\ -(x_1 - \epsilon) - x_2 - \cdots - x_d &\leq b \end{aligned}$$

They may be rewritten as,

$$\begin{aligned} (x_1 - \epsilon/2) + x_2 + \cdots + x_d &\leq b - \epsilon/2 \\ -(x_1 - \epsilon/2) + x_2 + \cdots + x_d &\leq b - \epsilon/2 \\ (x_1 - \epsilon/2) - x_2 + \cdots + x_d &\leq b - \epsilon/2 \\ -(x_1 - \epsilon/2) - x_2 + \cdots + x_d &\leq b - \epsilon/2 \\ &\vdots \\ -(x_1 - \epsilon/2) - x_2 - \cdots - x_d &\leq b - \epsilon/2 \end{aligned}$$

which define an ℓ_1 ball of radius $b - \epsilon/2$ centered at $(\epsilon/2, 0, \ldots, 0)$, that is, $\epsilon/2$ in the first coordinate and zero everywhere else. □

C. Additional Plots of Certificate Upper Bounds

See Figure 1.

D. Experimental Details

Our experiments are adapted from the released code for ℓ_2 smoothing from (Cohen et al., 2019). In particular, for each Generalized Gaussian distribution with varying parameter q and standard deviation σ, we trained a ResNet-110 classifier on CIFAR-10 for 90 epochs, with the training under the same noise distribution as used for certification. All training and certification parameters are identical to those used in (Cohen et al., 2019) unless otherwise specified. In particular, all certificates are reported to 99.9% confidence, and we tested using a 500-image subset of the CIFAR-10 test set. For lower-resolution versions of CIFAR-10, we again trained separate models for each resolution used, with the resolution at training time matching the resolution at test time. We first reduced the image resolutions before adding noise, then, once the noise was added, scaled the images back to the original 32×32 resolution (by repeating pixel values) before classifying with ResNet-110: this ensured that the number of parameters did not vary between classifiers.

We trained with $\sigma = 0.12, 0.25, 0.50, 1.00$ for resolutions 32×32, 16×16 and 8×8. At higher levels of noise for each scale ($\sigma = 0.25$ for 8×8, $\sigma = 0.5$ for 8×8 and 16×16, $\sigma = 1.00$ on all scales) the resulting classifiers could not correctly certify the median image ($\underline{p_1(x)} < .5$), so we do not report any certificates.

Values for ImageNet for the median certificate under Gaussian noise are adapted from the released certificate data from (Cohen et al., 2019).

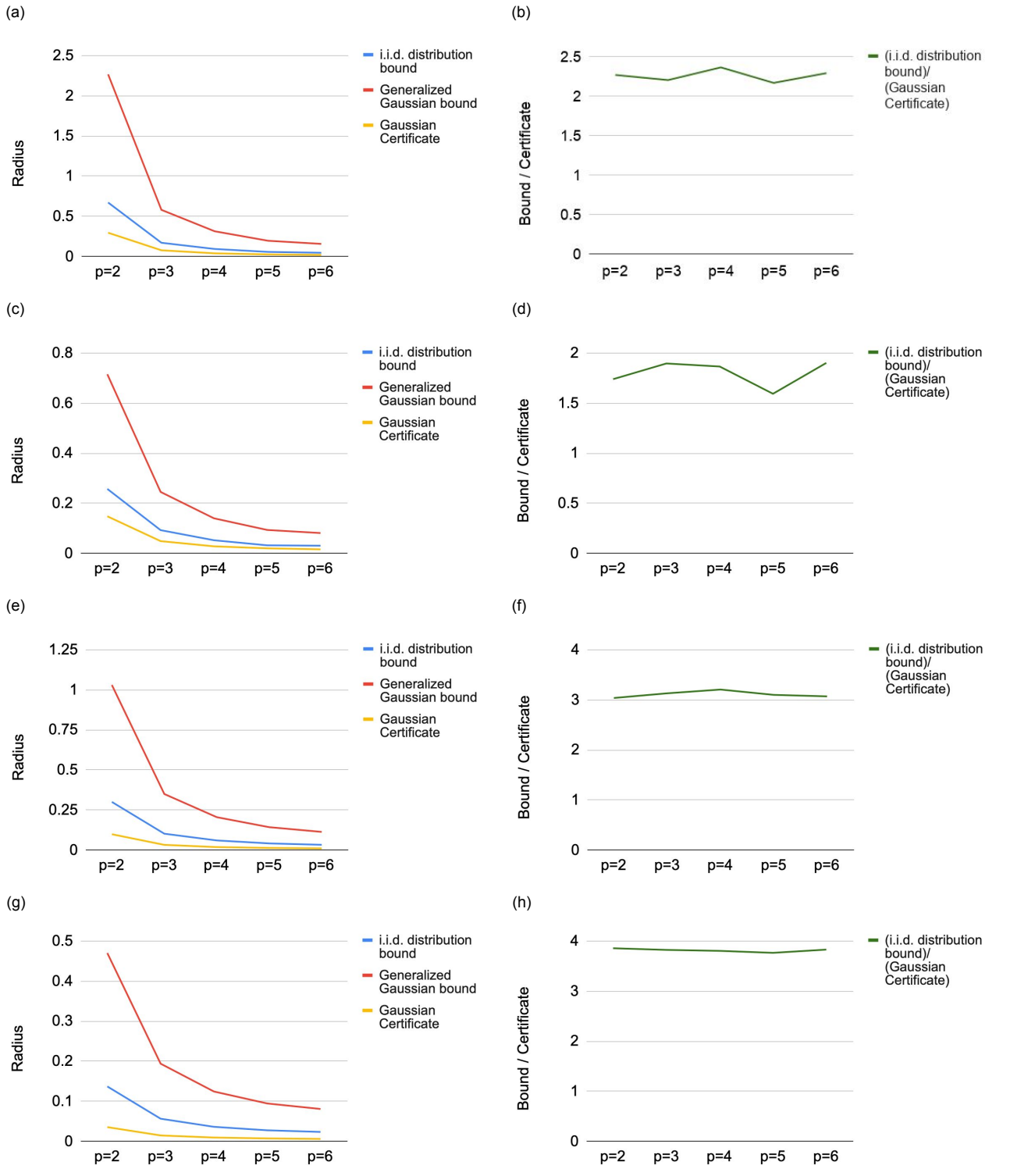

Figure 1. Upper bounds for certifying with Generalized Gaussian noise on CIFAR-10 images, with $q = p$, compared with certificates using Gaussian noise directly.Left panels show the certificates and the bounds directly, while right panels show the ratios between the i.i.d. distribution bounds (tighter in each case) and the certificates. Panels (a,b) use unaltered CIFAR-10 images with $\sigma = 0.5$ noise. Panels (c,d) and (e,f) use CIFAR-10 images at 16×16 scale with $\sigma = 0.12$ and $\sigma = 0.25$ respectively. Panels (g,h) use CIFAR-10 images at 8×8 scale with $\sigma = 0.12$.