Supplementary Material: Leverage Score Sampling for Complete Mode Coverage in Generative Adversarial Networks

Joachim Schreurs[1 [0000−0001−8670−2553]], Hannes De Meulemeester[1 [0000−0002−5938−2387]], Michaël Fanuel[2⋆ [0000−0002−7438−0005]], Bart De Moor[1 [0000−0002−1154−5028]], and Johan A.K. Suykens[1 [0000−0002−8846−6352]]

[1] ESAT-STADIUS, KU Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium
{joachim.schreurs,hannes.demeulemeester, bart.demoor, johan.suykens}@kuleuven.be
[2] Univ. Lille, CNRS, Centrale Lille, UMR 9189 – CRIStAL, F-59000 Lille, France
michael.fanuel@univ-lille.fr

1 Organization

In Section 2, the remaining results from the UNBALANCED MNIST experiment are given. In Section 3, the effect of the ridge parameter γ and dimensionality reduction size k is discussed. Then, in Section 4, the training times of our simulations are reported. Section 5 provides further information about the distribution of generated samples in the artificial datasets. Finally, Section 6 describes the experimental setting, the proposed algorithm, and architectures used in the paper.

2 Extra table unbalanced MNIST dataset

For the UNBALANCED MNIST experiment, only the number of samples in the minority modes are given in the main part. The number of samples for the remaining modes is given in Table 1.

⋆ Most of this work was done when MF was at KU Leuven.

Table 1. Experiments on the UNBALANCED MNIST dataset. Two variants of RLS BuresGAN are trained. First, RLSs are calculated with an explicit feature map obtained from a pre-trained classifier (Class.). The second version uses the next-to-last layer of the discriminator (Discr.) to compute the RLSs. The RLS MwuGAN is initialized using the RLSs with the explicit feature maps obtained from a pre-trained classifier.

	Mode 6	Mode 7	Mode 8	Mode 9	Mode 10
GAN	1831(135)	1835(32)	2009(69)	1747(46)	1915(143)
PacGAN2	1767(116)	1864(16)	1983(105)	1847(25)	1900(110)
BuresGAN	1762(96)	1870(38)	1979(110)	1782(29)	1916(68)
IwGAN	1833(74)	1837(97)	1902(79)	1783(64)	1928(74)
IwMmdGAN	67(45)	0(0)	4(3)	8741(430)	22(10)
MwuGAN (15)	1638(137)	1986(114)	1807(96)	2026(22)	1801(123)
RLS MwuGAN (15) (ours)	1616(17)	1513(144)	2018(78)	1576(30)	1811(34)
RLS BuresGAN Class. (ours)	1311(178)	1137(116)	1743(79)	1069(67)	1396(100)
RLS BuresGAN Discr. (ours)	2012(97)	1839(53)	1608(155)	1841(55)	1545(115)

3 Effect of the regularization parameter in RLS

3.1 Synthetic data

We display in Table 2 an ablation study for different ridge regularization parameters $\gamma > 0$ on the synthetic RING and GRID. A lesson from Table 2 is that the performance on the synthetic datasets does not vary much if different values of γ are chosen in the case of the discriminator feature map. Notice that, for $\gamma = 10^{-3}$ and $\gamma = 10^{-4}$, there is a small improvement in sample quality for RING and in terms of mode coverage for GRID with the Gaussian kernel feature map.

Table 2. Ablation study over the parameter γ on the synthetic datasets for RLS BuresGAN with a discriminator feature map (Discr.) and Gaussian feature map (Gauss.) with bandwidth $\sigma = 0.15$. No significant difference can be seen from the reported performance as the regularization parameter varies.

Feature map	γ	Ring with 8 modes		Grid with 25 modes	
		Nb modes	% in 3σ	Nb modes	% in 3σ
Discr.	10^{-2}	8(0)	0.92(0.02)	24.0(1.3)	0.79(0.03)
	10^{-3}	8(0)	0.90(0.02)	24.4(0.9)	0.78(0.06)
	10^{-4}	8(0)	0.90(0.01)	24.5(0.8)	0.77(0.07)
Gauss.	10^{-2}	8(0)	0.83(0.10)	19.9(1.9)	0.76(0.09)
	10^{-3}	8(0)	0.90(0.02)	24.0(1.5)	0.76(0.11)
	10^{-4}	8(0)	0.87(0.04)	24.8(0.6)	0.78(0.03)

Impact of the number of generators in MwuGAN. RLSs can also be used as initial starting weights in MwuGAN. The RLSs are constructed with a Gaussian kernel with bandwidth $\sigma = 0.15$ and regularization $\gamma = 10^{-3}$. Both methods contain a mixture of 15 GANs. This approach is compared with the classical initialization with uniform weights. In Figure 1, a clear improvement over a uniform initialization is visible when RLSs are used as initial weights.

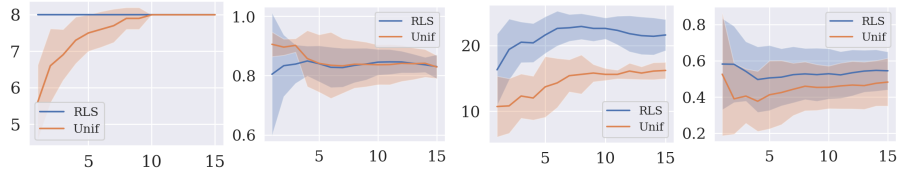

Fig. 1. RING: number of modes (first) and sample quality (second). GRID: number of modes (third) and sample quality (fourth). The x-axis refers to the number of generators of MwuGAN. Sample quality is assessed by counting the number of modes within 3σ of each mode. The blue (resp. red) curve indicates the results obtained for RLS (resp. unif.) sampling with MwuGAN.

3.2 Unbalanced MNIST

We describe here an ablation study for different ridge regularization parameters $\gamma > 0$ and size of dimension reduction k on the UNBALANCED 012-MNIST and UNBALANCED MNIST dataset. Moreover, the use of sketching the fixed explicit feature map is analyzed. From these results is concluded that the non-linear UMAP procedure leads to better mode coverage. The results are given in Tables 3 and 4.

Table 3. Ablation study over the regularization parameter γ and dimension k on the UNBALANCED 012-MNIST dataset for RLS BuresGAN with a discriminator feature map (Discr.) with sketching and an explicit feature map obtained from the next-to-last layer of a pre-trained classifier (Class.) with a UMAP dimensionality reduction and sketched dimensionality reduction. Minority modes are highlighted in black in the first row. The best performance is achieved using the Discr. sketched and Class. UMAP. feature maps with regularization 10^{-4} and dimension $k = 25$.

Feature map	Dim.	γ	Mode 1	Mode 2	**Mode 3**	KL
Discr. Sketched	10	10^{-2}	5561(381)	2756(300)	1409(345)	0.16(0.05)
		10^{-3}	5568(298)	2769(178)	1386(198)	0.15(0.03)
		10^{-4}	5655(279)	2812(135)	1218(253)	0.17(0.04)
	25	10^{-2}	5814(379)	2405(225)	1501(247)	0.17(0.04)
		10^{-3}	5715(312)	2477(139)	1415(310)	0.17(0.04)
		10^{-4}	5748(172)	2416(268)	1566(293)	0.16(0.02)
Class. UMAP	10	10^{-2}	3711(173)	5361(103)	751(161)	0.18(0.03)
		10^{-3}	3813(195)	5425(208)	644(67)	0.19(0.01)
		10^{-4}	3713(221)	5431(226)	712(152)	0.18(0.03)
	25	10^{-2}	4221(104)	4958(137)	672(109)	0.18(0.02)
		10^{-3}	4179(111)	4876(148)	802(105)	0.16(0.01)
		10^{-4}	3414(161)	4862(134)	1461(183)	0.08(0.01)
Class. Sketched	10	10^{-2}	5865(122)	3389(79)	605(111)	0.25(0.03)
		10^{-3}	5715(241)	3456(216)	683(161)	0.31(0.01)
		10^{-4}	5952(309)	3403(206)	513(135)	0.27(0.04)
	25	10^{-2}	5998(193)	3130(105)	672(119)	0.25(0.03)
		10^{-3}	5957(100)	3127(136)	775(95)	0.23(0.01)
		10^{-4}	6009(157)	3138(142)	698(121)	0.32(0.01)

Table 4. Ablation study over the regularization parameter γ and dimension k on the UNBALANCED MNIST dataset for RLS BuresGAN with a discriminator feature map (Discr.) with sketching and an explicit feature map obtained from the next-to-last layer of a pre-trained classifier (Class.) with a UMAP dimensionality reduction and sketched dimensionality reduction. Minority modes are highlighted in black in the first row. The best performance is achieved using the Class. UMAP. feature map.

Feature map	Dim.	γ	**Mode 1**	**Mode 2**	**Mode 3**	**Mode 4**	**Mode 5**	Mode 6	Mode 7	Mode 8	Mode 9	Mode 10	KL
Discr. Sketched	10	10^{-2}	254(77)	170(30)	222(40)	227(46)	212(62)	2061(94)	1871(134)	1598(71)	1853(83)	1532(96)	0.38(0.02)
		10^{-3}	214(38)	172(123)	239(35)	223(99)	230(54)	2081(91)	1832(32)	1611(100)	1898(44)	1501(123)	0.39(0.02)
		10^{-4}	235(62)	183(141)	264(44)	255(109)	219(54)	2012(97)	1839(53)	1608(155)	1841(55)	1545(115)	0.37(0.02)
	25	10^{-2}	236(56)	191(145)	286(43)	231(72)	219(60)	2066(113)	1765(48)	1505(81)	1958(37)	1543(56)	0.37(0.02)
		10^{-3}	230(30)	172(129)	267(31)	248(89)	274(68)	2046(100)	1780(42)	1546(122)	1899(73)	1538(135)	0.37(0.02)
		10^{-4}	224(49)	147(179)	255(27)	223(52)	197(53)	2032(116)	1790(49)	1554(119)	1962(55)	1617(143)	0.40(0.02)
Class. UMAP	10	10^{-2}	885(79)	662(143)	359(229)	814(97)	578(66)	1248(152)	1137(152)	1847(154)	1095(98)	1374(112)	0.10(0.01)
		10^{-3}	892(137)	674(90)	370(227)	803(108)	605(75)	1187(189)	1123(103)	1784(111)	1110(95)	1452(189)	0.09(0.02)
		10^{-4}	875(112)	663(122)	360(198)	831(59)	615(82)	1311(178)	1137(116)	1743(79)	1069(67)	1396(100)	0.09(0.01)
	25	10^{-2}	556(60)	525(82)	352(59)	542(71)	473(56)	1471(99)	1449(87)	1915(101)	1327(66)	1392(80)	0.16(0.01)
		10^{-3}	495(58)	473(114)	385(50)	546(83)	463(50)	1450(120)	1452(69)	1964(110)	1290(75)	1483(81)	0.17(0.02)
		10^{-4}	499(73)	443(103)	358(48)	513(104)	450(45)	1491(148)	1468(51)	1957(119)	1329(43)	1493(145)	0.18(0.01)
Class. Sketched	10	10^{-2}	218(72)	188(171)	165(31)	261(114)	246(41)	2019(166)	1726(48)	1859(204)	1683(63)	1637(178)	0.38(0.02)
		10^{-3}	189(45)	194(238)	184(88)	245(232)	221(44)	2147(176)	1712(38)	1824(182)	1666(61)	1618(177)	0.4(0.03)
		10^{-4}	239(45)	127(200)	211(27)	232(133)	214(65)	2127(113)	1783(48)	1804(122)	1642(29)	1622(142)	0.48(0.02)
	25	10^{-2}	238(30)	192(205)	189(35)	308(84)	254(39)	2011(105)	1699(74)	1710(127)	1785(54)	1615(143)	0.36(0.02)
		10^{-3}	262(45)	204(116)	251(45)	258(83)	266(62)	2041(152)	1621(40)	1774(119)	1751(41)	1573(134)	0.35(0.02)
		10^{-4}	246(50)	220(131)	199(41)	312(70)	306(41)	2033(127)	1689(38)	1712(115)	1724(67)	1560(100)	0.34(0.02)

4 Timings

Training times with a single NVIDIA Tesla P100-SXM2-16GB @1.3 GHz GPU. Temporary evaluations and plotting during training are included in the timings, so only relative comparisons are instructive.

Table 5. Total training time in seconds, averaged over 10 runs.

	Ring	Grid	012-MNIST	MNIST
GAN	45(1)	48(1)	1063(4)	1074(3)
PacGAN2	65(1)	71(1)	1205(4)	1211(8)
BuresGAN	100(1)	105(2)	1357(3)	1373(3)
IwGAN	65(1)	65(1)	1066(5)	1080(3.3)
IwMmdGAN	45(1)	56(1)	519(4)	5126(1)
MwuGAN (15)	1195(9)	1225(8)	20730(314)	36074(292)
RLS MwuGAN (15)	1198(9)	1217(9)	20823(233)	35400(566)
RLS BuresGAN Gauss.	130(3)	136(2)	/	/
RLS BuresGAN Class.	/	/	1615(11)	2494(8)
RLS BuresGAN Discr.	176(4)	187(1)	2006(3)	2004(2)

5 Distribution of generated samples for the artificial datasets

The number of samples per mode for all GANs in the Ring experiments are given in Figure 2. The number of samples per mode for all GANs in the Grid experiments are given in Figure 3.

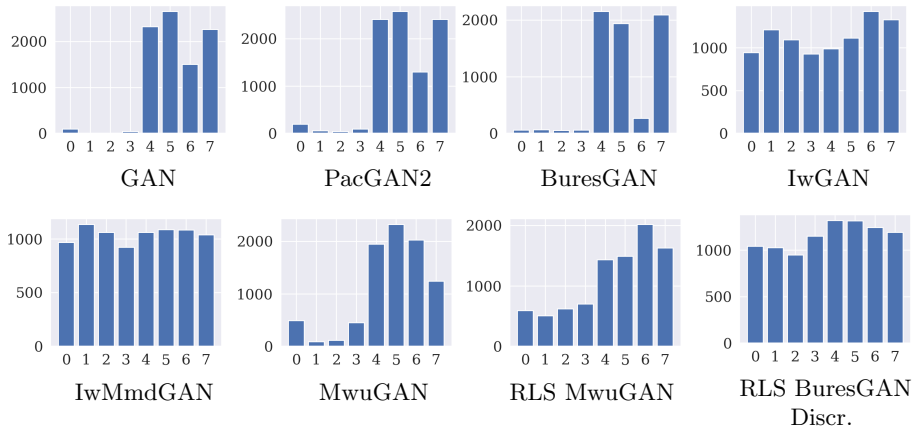

Fig. 2. Samples per mode for a single GAN trained on the RING dataset. Only IwMmdGAN, MwuGAN (15), RLS MwuGAN (15), and RLS BuresGAN with a discriminator feature map (Discr.) are capable of covering all the modes.

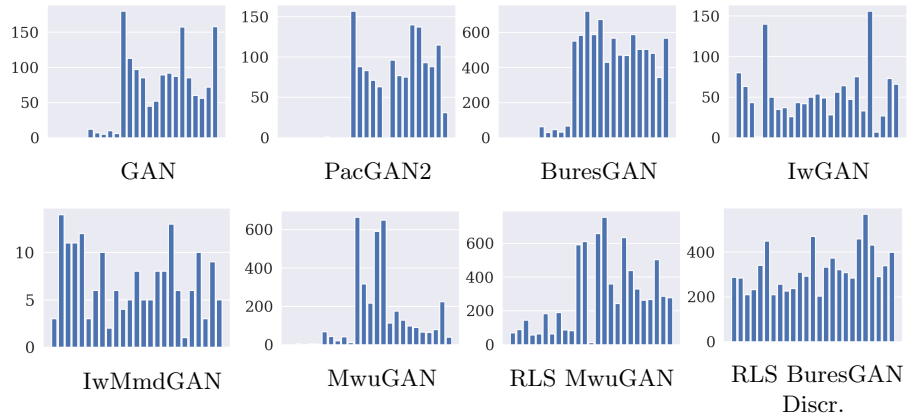

Fig. 3. Samples per mode for a single GAN trained on the GRID dataset. Only the proposed RLS BuresGAN with a discriminator feature map (Discr.) is capable of covering all the modes.

6 Architectures, algorithms, and settings

Settings experimental section. Unless specified otherwise, the models are trained for 30k iterations with a batch size of 64, by using the Adam optimizer with $\beta_1 = 0.5$, $\beta_2 = 0.999$ and learning rate 10^{-3} for both the generator and discriminator. The dimensionality of the generator latent space ℓ is equal to 100. The results of the experimental section always refer to the performance achieved at the end of the training. All the images are scaled in between -1 and 1 before running the algorithms. The hyperparameters of the competing methods are chosen as suggested in the authors' reference implementations. For MwuGAN, we take $\delta = 0.25$ and run for an increasing number of generators (displayed in brackets). In all the simulations, PacGAN always uses 2 as packing number. Note that IwMmdGAN and IwGAN require full knowledge of the Radon-Nikodym derivative M, where we want to generate data from a target distribution p but only have access to representative samples from a modified distribution Mp. The method is thus not an unsupervised method and tackles a slightly different problem. However, the method was included for completeness.

Algorithm 1 RLS BuresGAN with a fixed feature map.

Input: $\{x_i\}_{i=1}^n$, regularization γ, dimension reduction size k, discriminator D, generator G, feature map $\varphi(\cdot) \in \mathbb{R}^m$ - explicit *or* implicit via kernels $k(\cdot, \cdot)$

Calculate ridge leverage scores.

if $\varphi(\cdot)$ = Implicit **then**
  Calculate the RLSs for all points: $\ell_i = \left(K(K + n\gamma\mathbb{I})^{-1}\right)_{ii}$.
else if $\varphi(\cdot)$ = Explicit **then**
  if dimension reduction **then**
    $\varphi(x_i) \leftarrow$ UMAP of size k calculated on the full projected dataset $\{\varphi(x_i)\}_{i=1}^n$.
  else
    $k \leftarrow m$.
  end if
  if $n > k$ **then**
    Calculate the RLSs for all points: $\ell_i = \varphi(x_i)^\top (C + n\gamma\mathbb{I})^{-1}\varphi(x_i)$.
  else
    Calculate the RLSs for all points: $\ell_i = \left(K(K + n\gamma\mathbb{I})^{-1}\right)_{ii}$.
  end if
end if

Train BuresGAN.

while not converged **do**
  Sample a real and fake batch with probability $x_i \sim \ell_i$.
  Update the generator G by minimizing $V_G + \mathcal{B}(\hat{C}_r, \hat{C}_g)^2$
  Update the discriminator D by maximizing $-V_D$
end while

Algorithm 2 RLS BuresGAN with a discriminator-based explicit feature map.

Input: $\{x_i\}_{i=1}^n$, regularization γ, dimension reduction size k, mini-batch size s, discriminator D with discriminator feature map $\varphi_D(\cdot) \in \mathbb{R}^m$, generator G.

Train BuresGAN with adaptive RLSs.

while not converged **do**
  Sample a mini-batch $\hat{x}$ uniform at random of size $b = 20s$.
  if dimension reduction **then**
    $\varphi(\hat{x}_i) \leftarrow S^\top \varphi_D(\hat{x}_i) \in \mathbb{R}^k$, with sketching matrix $S = A/\sqrt{k} \in \mathbb{R}^{m \times k}$.
  else
    $\varphi(\hat{x}_i) \leftarrow \varphi_D(\hat{x}_i)$, $k \leftarrow m$.
  end if
  if $b > k$ **then**
    Calculate the RLSs for all points: $\ell_i = \varphi(x_i)^\top (C + n\gamma\mathbb{I})^{-1} \varphi(x_i)$.
  else
    Calculate the RLSs for all points: $\ell_i = \left(K(K + n\gamma\mathbb{I})^{-1}\right)_{ii}$.
  end if
  Sample a real and fake batch of size s with probability $\hat{x}_i \sim \ell_i$.
  Update the generator G by minimizing $V_G + \mathcal{B}(\hat{C}_r, \hat{C}_g)^2$
  Update the discriminator D by maximizing $-V_D$
end while

Table 6. The generator (left) and discriminator (right) architectures for the synthetic examples.

Layer	Output	Activation
Input	25	-
Dense	128	tanh
Dense	128	tanh
Dense	2	-

Layer	Output	Activation
Input	2	-
Dense	128	tanh
Dense	128	tanh
Dense	1	-

Table 7. The generator (left) and discriminator (right) architectures for the UNBALANCED 012-MNIST and the UNBALANCED MNIST experiments. BN indicates if batch normalization is used.

Layer	Output	Activation	BN
Input	100	-	-
Dense	12544	ReLU	Yes
Reshape	7, 7, 256	-	-
Conv'	7, 7, 128	ReLU	Yes
Conv'	14, 14, 64	ReLU	Yes
Conv'	28, 28, 1	ReLU	Yes

Layer	Output	Activation	BN
Input	28, 28, 1	-	-
Conv	14, 14, 64	Leaky ReLU	No
Conv	7, 7, 128	Leaky ReLU	Yes
Conv	4, 4, 256	Leaky ReLU	Yes
Conv	2, 2, 512	Leaky ReLU	Yes
Flatten	-	-	-
Dense	1	-	-

Table 8. The CNN architecture of the classifier used during the evaluation of the MNIST experiments. Dropout with a rate of 0.5 is used before the final dense layer.

Layer	Output	Activation
Input	28, 28,1	-
Conv	24, 24, 32	ReLU
MaxPool	12, 12, 32	-
Conv	8, 8, 64	ReLU
MaxPool	4, 4, 64	-
Flatten	-	-
Dense	1024	ReLU
Dense	10	-