GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium

Naresh Babu Bynagari

Director of Sales, Career Soft Solutions Inc, 145 Talmadge rd Edison NJ 08817, Middlesex, **USA**

Corresponding Email: naresh@careersoftusa.com

ABSTRACT

When it comes to the formation of real-looking images using some complex models, Generative Adversarial Networks do not disappoint. The complex models involved are often the types with infeasible maximum likelihoods. Be that as it may, there is not yet any proof for the convergence of GANs training. This paper proposes a TTUR (a two-time scale update rule) for training the Generative Adversarial Networks with a descent of stochastic gradient based on haphazard loss functions. The two time-scale update rule has separate learning rates for the generator and the discriminator. With the aid of the stochastic approximation theory, this paper demonstrates that the TTUR reaches a point of convergence under the influence of mild assumption to a kind of remote and stationary state known as Nash equilibrium. This unification or meeting point principle also applies to the widespread Adam optimization. This is a form or replacement optimization algorithm designed into stochastic gradient descent and used for tutoring the deep learning models in the system. For the Adam optimization theory, this paper evinces that it is in line with the dynamics of a weighty ball in a frictional state. Thus, we prove that it favours flat minina in the objective perspective of things. To carry out an evaluation of how GANs perform during the image creation process, this paper presents what we have termed the 'Fréchet Inception Distance", also known as FID—a concept known to dwell on the resemblance between the images created and the real ones in a way that is more improved compared to the Inception Score. Experimentally, the TTUR helps in the bettering of DCGANs and Improved Wasserstein GANs (WGAN-GP). This makes it perform better than the traditional CelebA GAN training, LSUN Bedrooms, CIFAR-10, SVHN and the One Billion Word Benchmark.

Key Words: GANs Trained, Time-Scale Update, Local Nash Equilibrium

Source of Support: None, **No Conflict of Interest:** Declared

INTRODUCTION

In the making of realistic images and producing text copies that meet the standard benchmark, Generative Adversarial Networks (GANs) are known for remarkable outcomes. GANs has the capacity to learn generative models of overly complex natures, despite the maximum likelihood or approximations for variation being impracticable. Rather than the said likelihood, a network of discriminators fill in as the controlling objective for the model of generation—which is also the generator.

GAN learning is a play involving the generator—the constructor of the man-made data from a randomness of variables—as well as the discriminator. The discriminator separates the synthetic information from realistic data. What is the function or job description of the generator? It is meant to fashion the data in such a way that it would not be able to easily tell the difference between this information and the real-world facts. As such, the discriminator will attempt to reduce the connecting errors between synthetic and real data, after which the generator will increase the discrepancies (Vadlamudi, 2016).

Because GANs training is a sport and since the solution to the puzzle-like system is in a state of Nash equilibrium, the convergence of gradient descent may not be successful. Local Nash equilibria alone will be discovered, due to the reality that the gradient descent is a local method of optimization. Should there be some sort of domestic environment surrounding a spot in the space of parameters where neither generator nor discriminator attempt to dial their various individual losses in a way that is signed by one of two factions, then we choose to call it a point of local Nash equilibrium.

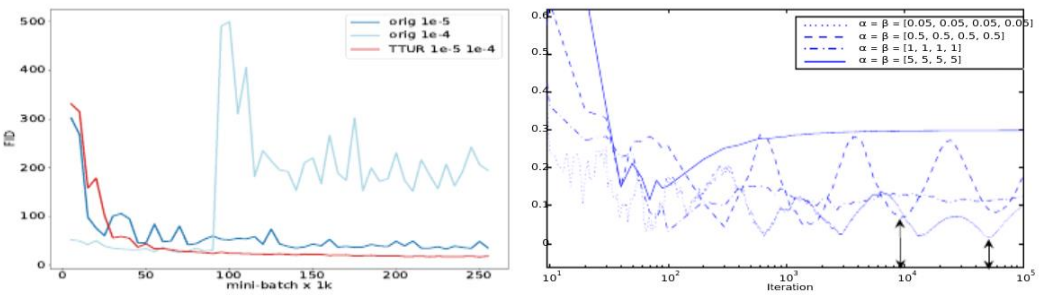

Figure 1: A demonstration of the average distance between the parameter and the maximum position for a one time-scale update process, the one of a 4 node network flow problem.

REVIEW OF RELATED LITERATURE

An increasing count of models in the machine learning realm need to be optimized based on a multiplicity of interacting objectives. With the aid of GANs, imaginative agents, the multi-agent type of reinforcement learning as well as hierarchical reinforcement learning, this is exactly the case (Bynagari, 2015). When it comes to solving saddle-point problems, it is the key (Kim et al., 2006) to achieve encompasses learning and the recreation of images (Chambolle & Pock, 2011). These instances can be planted as games, in which players modules that have been parameterized will compete or align to reduce their respective functions objective-wise. In order to define a flawless solution to the issue of optimizing a multi-objectively, we could bank on the Nashh equilibrium notion (Nash, 1951). At the Nash equilibrium basis, there is no player that's capable of improving its objective by changing strategy unilaterally.

To discover a form of Nash equilibrium in such an environment is analogous to fixing a problem with inequality in variations with the help of a monotone operator (Harker & Pan, 1990; Rosen, 1965). This variational equality problem is solvable with the first-order methods which dominate mono-objective enhancements for machine learning processes. The method best known to achieve convergence to a local minima under relatively simple conditions augmented by ML issues is the stochastic gradient descent (Bottou & Bousquet, 2008). Nonetheless, though gradient descent is completely applicable to a disparity of simultaneous objectives, there is a chance (Gidel et al,. 2019) it would be unable to discover a Nash equilibrium under simplistic settings (Bynagari, 2018).

There are a pair of alternative modifications of the gradient descent necessary to solving variational inequality, and ultimately, the Nash problem. They are: averaging (Magnanti & Perakis, 1997) and extrapolation with averaging (Nedic & Ozdaglar, 2009). The second modification was created as the extra-gradient method (or the EG) by Korpelevish in 1976. The EG method is not only much faster but also has the capacity to work with noisy gradients (Juditsky et al., 2011). Extrapolation is in correspondence to a step that shapes the opponent, as such as every player is in anticipation of its adversaries' next moves in order to update their working strategies (Vadlamudi, 2018).

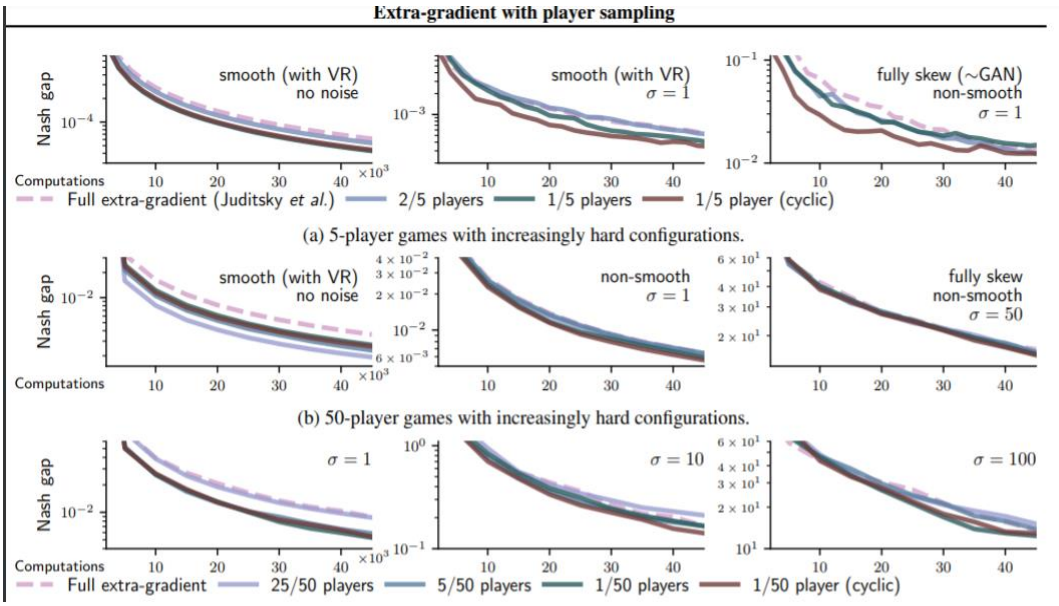

Figure 2: Illustrates a smooth game comprising 50 players with an increasing noise pattern. The sampling is done with reduction in variance.

The concept of extra-gradient has also been applied to settings that are non-convex. The asymptotic convergence outcomes for extra-gradient have been proven (Bynagari, 2017) without averaging in a case that is slightly non-convex. The effectiveness of this concept has been demonstrated and argued to give allowance for the escape of potentially catastrophic behaviors that comes from concomitant gradient updates (Ganapathy, 2018). Existing research regarding Generative Adversarial Networks propose that simultaneous updates should be replaced with alternated updates imbued with proportional enhancements (Gulrajani et al., 2017).

Characterizing the convergence of elements related to training general GANs remains a challenge open to many. For GAN variants of special nature, convergence can be demonstrated with a selection of assumptions. One prerequisite for many proofs of convergence is local stability, as evidenced by Kolter and Nagarajan (in 2017) for a GAN configuration of minimum-maximum. Nevertheless, these two researchers required either strong, unrealistic presuppositions or a limitation to a discriminator of linear form. Recent proofs of convergence for GANs elicit expectations for training samples or for the count of instances headed for infinity. Due to this, Nagarajan and Kolter are negligent towards the mini-batch learning, which culminates in stochastic gradient descent (Polyak, 1964).

The concept known as extra-gradient can also be comprehended as a method that shapes the opponent. When it comes to the earlier-mentioned extrapolation process, the player takes a look a step into the future and readies for the opponents' next actions. A lot of recent studies proposed algorithms that utilize the moves of opponents' data to converge unto a state of equilibrium (Zhang & Lesser, 2010; Bynagari, 2016; Foerster et al., 2018). Training using an opponent-shaping kind of consciousness is believed to foster cooperative behaviors in games (Foerster et al., 2018).

Likewise, recent research (Ganapathy, 2018) suggests algorithms that can bring about a change in the conventional dynamics of concurrent gradient descent. This they proposed to happen by adding an adjustment to enable the convergence to the Nash equilibrium. A caveat with these works, nonetheless, is that they demand an estimation for the Jacobian simultaneous gradient. That could very well cost significantly more when it comes to systems that are of large scales. It remains an impossibility even when non-smooth losses are involved.

RESEARCH METHODOLOGY

Actor-critic learning has been subject to recent analyses using the stochastic approximation. Research has demonstrated that a TTUR process makes sure that the induced learning gets to a state of native Nash equilibrium should the critic learn faster compared to the actor (Prasad et al, 2001). The concept called convergence was proved with the aid of an ordinary differential equation ODE), which has stable limit spots that position or occur alongside remote Nash equilibrium. This paper follows the suit of this approach to demonstrate that there is a convergence of GANs to a remote Nash equilibrium should they be learned by a TTUR. That happens in a scenario where the generator and discriminator possess different rates of learning, which also culminates in better experimental outcomes.

The major school of thought here is: the discriminator reaches a convergence point with the said local equilibrium in a case where the generator has been fixed. Should the generator undergo modifications gradually, in turn, the discriminator will still converge because its interferences remain minimal. Apart from making the convergence take place, the performance also has the capacity to improve because the discriminator initially needs to initially master the new patterns prior to the transfer onto the generator. Contrastingly, a generator is significantly fast, enough to steadily drive the discriminator into new regions without getting hold of the data it has collected. In GANs implementations, it's often the discriminator learning much faster than the generator.

A fresh objection made the generator slow down so as to save it from giving the existing denominator too much training. The Wasserstein GAN algorithm makes use of more updated steps for discriminator purposes, much more than it does for the generator. This

paper also compares the TTUR with the conventional GAN training sequence. On the southern panel of a stochastic type of gradient, case in point, there is a CelebA that is designed specifically for authentic GANs training. More often than not, this medium is responsible for oscillations and the TTUR.

The right or northern panel, on the other hand, lies an ideal instance of a flow in network node issue (Zhang et al, 2007). The distance separating the real parameter from its maximum point for a sole time-scale update rule remains visible in several iterations. When the upper bounds on the error are of insignificant sizes of volumes, the iterates will go back to the environment of the optimization. For larger errors, on other hand, the iterations would assume divergence.

We have some new contributions as per this paper, and they include:

- The TTUR mechanism for Generative Adversarial Networks (GANs).
- Evidence that GANs learned using the TTUR do reach convergence with a native Nash equilibrium.
- Describing the Adam optimization concept as a heavy ball with friction and the culminating order of differential equations.
- GANs trained with the Adam optimization and TTUR can also converge to a stationary remote Nash equilibrium.
- This paper also proposes the "Fréchet Inception Distance" (FID) for the evaluation of GANS, which we find to be an approach that betters the Inception Score when it comes to consistency.

EXPLAINING THE TTUR

Because of the purpose of this paper, we took into consideration a discriminator D(/;) and generator with parameter vectors. The learning is heavily built upon the stochastic gradient concept g(θ,w) of the loss function (LG) of the generator. The loss functions, often represented as LD and LG, could be of original nature (Paruchuri, 2018). The loss functions can also be the generator's improved functions or newly proposed losses for GANs (Wasserstein, 2013). This study's setting is not limited to minimum-maximum GANs. Nevertheless, it is relevant to all other GANs in general. For this, the loss function LD of the discriminator is not necessarily in relationship with the generator's own loss functions.

Gradients g8 and h8 are said to be stochastic because they are in the habit of utilizing smaller batches of randomly selected realistic exhibits (i). Should the true gradients be (θ,w) = ∇wLDandh(θ,w) = ∇θLG, then for definition, we say ˜g(θ,w) = g(θ,w)+M(w)and˜h (θ,w) = h(θ,w)+M(θ) with variables in random M(w) and M(θ). As a result, the gradients˜gθ,wand˜hθ, are stochastic approximations to the true gradients. In the same manner, the study analyzes the convergence of GANs by 2 time-scale stochastic approximation algorithms. For a TTUR, we employed the learning rates b(n) and a(n) for the discriminator and the generator update, respectively:

$$wn+1 = wn+b(n)g\theta n, wn+M(w)n, \theta n+1 = \theta n+a(n)h\theta n, wn+M(\theta)n. (1)$$

To demonstrate the convergence of GANs learned by TTUR, this research hinges on the following set of assumptions

- The gradients known as (h and g) are Lipschitz.◭ As a result of this, networks that come with Lipschitz-like smooth activation capacities such as ELUs (α= 1) [13] meet only the earlier assumption. It does not fulfill the ReLU networks as well.

- Pna(n) = ∞,Pna2(n)<∞,Pnb(n) = ∞,Pnb2(n)<∞,a(n) = o(b(n))◭(A3)The stochastic gradient errors{M(θ)n}and{M(w)n}are martingale difference sequences art the increasingσ-fieldFn=σ(θl,wl,M(θ)l,M(w)l, l 6n), n >0withEhkM(θ)nk2| F(θ)ni6B1andEhkM(w)nk2| F(w)ni6B2, where B1 and B2 are considered to be contents that determine positivity.
- Borkar 1997 maintains the assertion from Lemma 2. The assumption is satisfied inside the Robbins-Monro setting, In this arrangement, smaller batches are sampled in a random manner and the gradients are brought to bounds with one another.
- For eachθ, the ODE˙w(t) = gθ,w(t) possesses a native asymptotically stable attractor λ(θ) within a domain of attraction Gθ such that λis Lipschitz. The ODE˙θ(t) =hθ(t),λ(θ(t))has a locally asymptotically stable attractor θ∗ inside a domain of attraction. ◭The discriminator needs to converge to a minimum for fixed generator parameters. Then, the generator, in turn, needs to as well converge to a minimum for this fixed discriminator minimum. (Borkar, 1997) demands for a unique global asymptotically stable equilibria for this.
- The Assumption of global attractors was relaxed to native attractors via Assumption (A6) and Theorem 2.7 (Karmakar & Bhatnagar). When it comes to this, the GAN objectives can act in the place of Lyapunov functions. These assumptions of locally stable ODEs can be ensured by adding extra weight decay termin the loss function which increases the eigenvalues of the Hessian.
- Ultimately, one can prevent a regional constant discriminator with a zero-second order of derivatives. In a typical sense, it is ensured by either an objective or a term relating to weight decay. The following proposition has been proven in a seminal research as Theorem 1 (Borkar, 1997). Should the assumptions be met, all updates will reach a convergence point.

The solution to the problem is a Nash equilibrium in a local stationary position. Alternatively, conferences can be proven with the use of the Poisson equation. For this approach, the solution takes on a linear update function assumption in the threshold of the fast-update rule, which, nevertheless, is susceptible to being linearly approximated to non-linear gradients. The chief ideas behind Borkar's proof is the use of perturbed ODEs (Hirsch, 1989). The proof depends on one reality: which is that there is a point in time when the perturbation on the slow update rule will be eventually and perhaps relevantly small enough to allow for the unification of the fast update rule, also known as convergence.

For TTUR-related experiments, we wanted to find the learning rates. That way, the slow update will be small enough for a convergence to happen. The slow update is typically the generator, while the fast update is used to refer to the generator. Both learning rates needed to be adjusted in order for the generator to not influence or impair the discriminator's learning curve in an unwanted manner that thwarts the process. Be as it may, when the generator has a learning rate more significant than that of the discriminator, it could ensure that the latter is perturbed only minimally. Learning rates cannot be directly interpreted as perturbations because how it affects the generator is quite different from the way it impacts the discriminator.

Adam's Role in HBF ODE And TTUR Convergence

For experimental purposes, this paper looks to leveraging the Adam stochastic approximation to prevent it from entering the collapsing mode. GANs are subject to the

mode, which manifests when there are humongous sets of probability that have been painstakingly mapped onto a less than a handful of modes covering just little areas. These environments are, in fact, a representation of significant samples. Be as it may, the variety of the real information is eventually lost. At the end, only a few samples of the said prototypes will be created by the system. There are different ways to prevent the collapsing mode (d (Korpelevich, 1976). For this research, obviate the collapsing with the Adam stochastic approximation. Because it can average over past gradients, Adam can be explained in the likeness of a heavy ball with friction (HBF). This averaging is in correspondence with a velocity which enables the generator's resistance to the push effect into smaller regions. Typically, as an HBF method, Adam tends to have an overshoot for small native minima that are in tally with the mode collapse. This stochastic approximation can also find flat minima in good generalization conditions. (Krizhevsky et al., 2009).

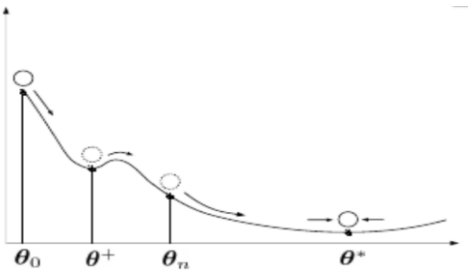

Figure 3: Heavy Ball with Friction, where the ball with mass overshoots the local minimumθ+and settles at the flat minimum θ∗

The next thing we attempt to conduct analysis on is whether or not TTUR-trained GANs can reach a convergence point when Adam is put into active use. The paper tried carries out a recapitulation of the Adam-type update rule at step *n*, with learning rate *a*, exponential averaging factors β1 for the first and β 2for the second moment of the gradient ∇f(θn−1):gn←− ∇f(θn−1)(2)mn←−− (β1/(1 −βn1)) mn−1+ ((1 −β1)/(1 −βn1)) gnvn←−− (β2/(1 −βn2)) vn−1+ ((1 −β2)/(1 −βn2)) gngnθn←−− θn−1−amn/(√vn+), where following operations are met according to their components: the product, the square root√., and the division / in the last line.

Rather than learning rate *a*, we proposed a damping coefficienta (n) with a(n) = an−τfortτ∈(0,1]. Adam has parameters β1 to average the gradient and β2 parametrized by a positive α for averaging the gradient that has been squared. These parameters can be described as a defining memory for the Adam concept of optimization. To characterize β1 and β2 in the following, the paper defines the exponential memory r(n) = r as well as the polynomial memory r(n) = r/ Pnl=1 a(l) for some positive constant r.

Looking to get speedier rates for convex games, we introduce the computation of an estimate of the gradient oracle but reduce by a variance. Through this process, we mitigate the noise because we have brought player sampling onto the mix. As a technique, variance reduction is recognized to boost convergence under the assumptions of smoothness in lookalike settings. Variance reduction has been applied on the noise emanating from the gradient estimates (While Palaniappan & Bach, 2016; Iusem et al.,2017; Chavdarova et al.,2019). However, we apply it to the noise from the player sampling. We retain a $f\ \nabla i`i$ estimate of ∇i`i for every player in a table R. Then, we used the said estimate to compute unbiased gradient estimates with lower variance, akin to the approach of SAGA (Defazio et al., 2014). This we did to reduce the variance of data noise.

In order to guarantee convergence, every player in the game needs to have the same amount of opportunity to be extra-gradient with the sampling (Paruchuri & Asadullah, 2018). This is often described as the equiprobable player sampling condition. An effective way to fulfill this is sampling uniformly over b-subsets of [n], as all the players possess a probability of p = b/n of being selected. To accelerate the convergence, we propose the cycling over of the n(n-1) pairs of players (when b = 1). At each point of iteration, the first player of the pair is extrapolated, and the second is updated. The pair orders are shuffled the moment the block is completely visible.

With this scheme, we are able to bridge extrapolation and alternated gradient descent. When it comes to GANs, this is similar to cylically extrapolating the generator before doing the discriminator update, and vice versa. Despite the fact that its convergence is not a surety, the cyclical sampling over players should be enough for convex quadratic games.

LEARNING RATE: DISCRIMINATOR VERSUS GENERATOR

The proof of convergence for training GANs with the TTUR has an assumption that the learning rate of the general will inevitably become the right amount of small to allow for a convergence on the part of the discriminator's training process. At a point in time, the disturbance of the discriminator-related updates will be sufficient to catalyze the convergence of the discriminator. The magnitude of the perturbations is critical to the convergence of this discriminator. The perturbations, which are induced into the updates from the discriminator, are not influenced by only the generator learning rate. They are as well determined by their loss functions, the current worth or natural value of the said function, the method of optimization put into employ, the size nature of the error signals present within the generator, and the architecture of the generator itself. Factors like the complex nature of the learning task for the system's generator and the regularization of the system also weigh in. The size of the generator learning rate doesn't factor into the size of the perturbation on the part of the discriminator updates. Instead, they act as modulators. As such, the learning rate of the generator can be substantially larger than that of the discriminator, doing so without introducing substantial perturbation into the training rate of the discriminator. Likewise, the training dynamics of the generator differ from that of the discriminator despite the fact that they possess similar learning rates.

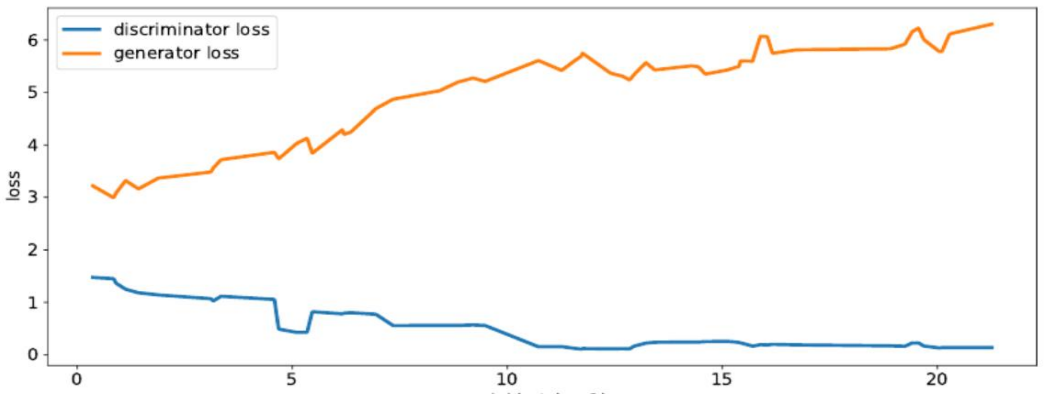

Figure 4: Illustrates the losses of the discriminator as well as the generator for a CelebA-based DCGAN experiment. The training rate here was at 0.0005 for both the generator and discriminator.

Be as it may, the discriminator loss decreases while that of the generator increases. With this instance, it is provable that the learning rate does not determine the perturbations, nor does it factor into the learning progress of two-coupled update rules. The generator's learning rate of choice ought to be uninfluenced by the choice of the discriminant. Plus, the search ranges of both generator and discriminator need to be non-dependent of one another however affixed to the related assignment and architecture, among other considerations.

CONCLUSION

For the purpose of carrying out GANs training, our research has suggested the TTUR and have also demonstrated that it can reach a convergence to a stationary and native Nash equilibrium. The paper also looked into the Adam stochastic optimization in terms of heavy ball with friction (HBF). In one way, that shows Adam also has the ability to converge. It also demonstrates that Adam has the tendency to locate flat minima whilst dodging minute native minima. The second set of differential equations refers to Adam's learning dynamics as an HBF setting.

Through the differential equation, the convergence of trained GANs to a definite natve Nash equilibrium is something that could extend into the realm of the Adam stochastic optimization. In order to carry out an evaluation on GANs, this paper proposed the "Fréchet Inception Distance" (FID), an idea that dwells heavily on the semblance between the formulated images and the real copies, much better than an Inception Score can perform. In these experiments, our research has also made a comparison between TTUR-trained GANs and the traditional process which involves learning with a one time-scale update rule on LSUN Bedrooms, CelebA, SVHN, CIFAR-10 and the One Billion Benchmark. In all of the experiments, two time-scale update rule consistently performed better than normal GANs training.

REFERENCES

Abbeel, P., and Mordatch, I. Learning with an opponent learning awareness. In Proceedings of the International Conference on Autonomous Agents and Multi-Agent Systems, 2018.

Bottou, L. and Bousquet, O. The tradeoffs of large scale learning. In Advances in Neural Information Processing Systems, pp. 161–168, 2008.

Bynagari, N. B. (2015). Machine Learning and Artificial Intelligence in Online Fake Transaction Alerting. *Engineering International*, *3*(2), 115-126. https://doi.org/10.18034/ei.v3i2.566

Bynagari, N. B. (2016). Industrial Application of Internet of Things. *Asia Pacific Journal of Energy and Environment*, *3*(2), 75-82. https://doi.org/10.18034/apjee.v3i2.576

Bynagari, N. B. (2017). Prediction of Human Population Responses to Toxic Compounds by a Collaborative Competition. *Asian Journal of Humanity, Art and Literature*, *4*(2), 147-156. https://doi.org/10.18034/ajhal.v4i2.577

Bynagari, N. B. (2018). On the ChEMBL Platform, a Large-scale Evaluation of Machine Learning Algorithms for Drug Target Prediction. *Asian Journal of Applied Science and Engineering*, *7*, 53–64. Retrieved from https://upright.pub/index.php/ajase/article/view/31

Chambolle, A. and Pock, T. A first-order primal-dual algorithm for convex problems with applications to imaging. Journal of Mathematical Imaging and Vision, 40(1):120–145, 2011.

Foerster, J., Chen, R. Y., Al-Shedivat, M., Whiteson, S., Abbeel, P., and Mordatch, I. Learning with an opponent learning awareness. In Proceedings of the International Conference on Autonomous Agents and Multi-Agent Systems, 2018.

Ganapathy, A. (2018). Cascading Cache Layer in Content Management System. *Asian Business Review*, *8*(3), 177-182. https://doi.org/10.18034/abr.v8i3.542

Ganapathy, A. (2018). UI/UX Automated Designs in the World of Content Management Systems. *Asian Journal of Applied Science and Engineering*, *7*(1), 43-52.

Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A. C. Improved training of Wasserstein GANs. In Advances in Neural Information Processing Systems, pp. 5767–5777, 2017.

Harker, P. T. and Pang, J.-S. Finite-dimensional variational inequality and nonlinear complementarity problems: a survey of theory, algorithms and applications. Mathematical Programming, 48(1-3):161–220, 1990.

Hirsch, M. W. (1989). Convergent activation dynamics in continuous time networks. Neural Networks, 2(5):331–349.

Juditsky, A., Nemirovski, A., and Tauvel, C. Solving variational inequalities with stochastic mirror-prox algorithm. Stochastic Systems, 1(1):17–58, 2011.

Kim, S.-J., Magnani, A., and Boyd, S. Robust Fisher discriminant analysis. In Advances in Neural Information Processing Systems, 2006.

Korpelevich, G. The extragradient method for finding saddle points and other problems. Matecon, 12:747–756, 1976.

Krizhevsky, A. and Hinton, G. (2009). Learning multiple layers of features from tiny images. Technical report, Citeseer.

Magnanti, T. L. and Perakis, G. Averaging schemes for variational inequalities and systems of equations. Mathematics of Operations Research, 22(3):568–587, 1997.

Nash, J. Non-cooperative games. Annals of Mathematics, pp. 286–295, 1951.

Nedic, A. and Ozdaglar, A. Subgradient methods for saddle-point problems. Journal of Optimization Theory and Applications, 142(1):205–228, 2009.

Paruchuri, H. (2018). AI Health Check Monitoring and Managing Content Up and Data in CMS World. *Malaysian Journal of Medical and Biological Research*, *5*(2), 141-146. https://doi.org/10.18034/mjmbr.v5i2.554

Paruchuri, H., & Asadullah, A. (2018). The Effect of Emotional Intelligence on the Diversity Climate and Innovation Capabilities. *Asia Pacific Journal of Energy and Environment*, *5*(2), 91-96. https://doi.org/10.18034/apjee.v5i2.561

Polyak, B. T. and Juditsky, A. B. (1992) Acceleration of stochastic approximation by averaging. SIAM J. Control and Optimization, 30(4):838–855, doi: 10.1137/0330046.

Rosen, J. B. Existence and uniqueness of equilibrium points for concave n-person games. Econometrica, 33(3):520–534, 1965.

Vadlamudi, S. (2016). What Impact does Internet of Things have on Project Management in Project based Firms?. *Asian Business Review*, *6*(3), 179-186. https://doi.org/10.18034/abr.v6i3.520

Vadlamudi, S. (2018). Agri-Food System and Artificial Intelligence: Reconsidering Imperishability. *Asian Journal of Applied Science and Engineering*, *7*(1), 33-42.

Vaishnavh Nagarajan and J. Zico Kolter. Gradient descent and optimization is locally stable. In Advances in Neural Information Processing Systems, pp. 5585–5595, 2017.

Zhiming Zhou, Han Cai, Shu Rong, Yuxuan Song, Kan Ren, Weinan Zhang, Jun Wang, and Yong Yu. Activation maximizing generative adversarial nets. In Proc. International Conference on Learning Representations (ICLR), 2018.

--0--