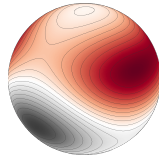




UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386



Organizational Meeting

Seminar: Mathematical Machine Learning – Generative Models and Measure Transport

Winter Semester 2026/27

Introduction



Roland Herzog

roland.herzog@iwr.uni-heidelberg.de

scoop-teaching@uni-heidelberg.de



Georg Müller

georg.mueller@uni-heidelberg.de

- Language: English
- Intended for Master students

Concept of the Seminar

- Read and analyze **scientific papers**.
- Many-to-Many Role-Playing Format from <https://colinraffel.com/blog/role-playing-seminar.html>:
- One paper per meeting.
- Take a specific role.
- Discuss the paper together.

Roles: Author

Assume that you are the author of the paper. Present your work. **Motivate** your research and **present your results**. Keep in mind that you only have 5 minutes, so focus on the main points. Split up the paper with your partner so that you do not repeat each other.

Roles: Reviewer

Assume that the paper has not been published yet and is currently submitted to a top conference where you've been assigned as a peer reviewer. Complete a **full review** of the paper answering all prompts of the official review form of the top venue in this research area, e.g. <https://www.jmlr.org/reviewer-guide.html>. This includes recommending whether to **accept or reject** the paper.

Roles: Archaeologist

Some time in the future: This paper was found buried under ground in the desert. You're an archaeologist who must determine where this paper sits in the **context** of previous and subsequent work. Find and report on one **older** paper cited within the current paper that substantially influenced the current paper and one **newer** paper that cites this current paper.

Roles: Hacker

You're a hacker who needs a demo of this paper ASAP. **Implement** a small part or simplified version of the main algorithm in the paper on a small dataset or toy problem. Prepare to share the core code of the algorithm to the class and demo your implementation. Do **not** simply download and run an existing implementation – though you are welcome to use (and give credit to) an existing implementation for “backbone” code.

Roles: Private Investigator

You are a detective who needs to run a background check on the **paper's authors**. Where have they worked? What did they study? What previous projects might have led them to working on this one? What motivated them to work on this project? Feel free to contact the authors, but remember to be courteous, polite, and on-topic.

- Attending **all** meetings.
- Participation in class
 - In the beginning every person gives a 5 minute report from their role perspective.
 - Afterwards we have a scientific discussion of the paper.
- We grade your performance after each meeting. Your final grade is the weighted sum of those five grades.

Papers

1. Diederik P. Kingma and Max Welling. “Auto-encoding variational Bayes”. In: *2nd International Conference on Learning Representations, ICLR 2014*. Ed. by Yoshua Bengio and Yann LeCun. 2014. arXiv: 1312.6114
2. Ian J. Goodfellow et al. “Generative adversarial nets”. In: *Advances in Neural Information Processing Systems*. Ed. by Z. Ghahramani et al. Vol. 27. NIPS'14. Curran Associates, Inc., 2014. URL: https://papers.nips.cc/paper_files/paper/2014/hash/f033ed80deb0234979a61f95710dbe25-Abstract.html
3. Martin Arjovsky, Soumith Chintala, and Léon Bottou. “Wasserstein generative adversarial networks”. In: *Proceedings of the 34th International Conference on Machine Learning*. Ed. by Doina Precup and Yee Whye Teh. Vol. 70. Proceedings of Machine Learning Research. PMLR, 2017, pp. 214–223. URL: <https://proceedings.mlr.press/v70/arjovsky17a.html>
4. Jonathan Ho, Ajay Jain, and Pieter Abbeel. “Denoising diffusion probabilistic models”. In: *Advances in Neural Information Processing Systems*. Ed. by H. Larochelle et al. Vol. 33. Curran Associates, Inc., 2020, pp. 6840–6851. URL: <https://proceedings.neurips.cc/paper/2020/hash/4c5bcfec8584af0d967f1ab10179ca4b-Abstract.html>
5. Yang Song et al. “Score-based generative modeling through stochastic differential equations”. In: *International Conference on Learning Representations, ICLR 2021*. 2021. URL: <https://openreview.net/forum?id=PxTIG12RRHS>

Timeline

- Today: Distribution of roles
 - Accepting a role is the **binding** registration for the seminar.
- Starting in October: 5 meetings à 120 minutes
- weekday and time:
Tuesday 16:00-18:00 (27.10., 10.11., 24.11., 8.12., 12.1.)
- in person, SR 10 of INF 205

Assigning Roles

	Paper 1	Paper 2	Paper 3	Paper 4	Paper 5
Author	1	9	7	5	3
Author	2	4	6	8	10
Reviewer	3	1	9	7	5
Reviewer	4	6	8	10	2
Archaeologist	5	3	1	9	7
Archaeologist	6	8	10	2	4
Hacker	7	5	3	1	9
Hacker	8	10	2	4	6
Private Investigator	9	7	5	3	1
Private Investigator	10	2	4	6	8

- It is highly recommended to communicate within your team.

What's next?

- Resources are online:
<https://scoop.iwr.uni-heidelberg.de/teaching/2026ws/seminar-mathematical-machine-learning/>
- These slides will be shared on the website.

Questions?

A measure-theoretic refresher: spaces and laws

A short refresher on the measure theory the seminar rests on:

- A **measure space** $(\Omega, \mathcal{F}, \mu)$: a σ -algebra $\mathcal{F}$ and a countably additive $\mu : \mathcal{F} \rightarrow [0, \infty]$ with $\mu(\emptyset) = 0$. A **probability measure** $\mathbb{P}$ adds $\mathbb{P}(\Omega) = 1$, giving a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.
- A **random variable** is a measurable map $X : (\Omega, \mathcal{F}) \rightarrow (\mathcal{X}, \mathcal{B})$ into a **value (data) space** $\mathcal{X}$ with its Borel σ -algebra $\mathcal{B}$.
- Its **law** (distribution) is the **pushforward** of $\mathbb{P}$ under X ,

$$P_X = X_{\#}\mathbb{P}, \quad P_X(\cdot) = \mathbb{P}(X^{-1}(\cdot)) \quad \text{on } \mathcal{B}.$$

- We work with $\mathcal{X} = \mathbb{R}^d$ (Borel $\mathcal{B}$) and forget the abstract Ω .

Integration, expectation, and pushforward

- Expectation is integration against the law (via the Lebesgue integral $\int f \, d\mu$):

$$\mathbb{E}[f(X)] = \int_{\Omega} f(X) \, d\mathbb{P} = \int_{\mathcal{X}} f \, dP_X.$$

- Expectation transports along a **pushforward**: for measurable $T : \mathcal{X} \rightarrow \mathcal{Y}$, the image measure $T_{\#}\mu$, $(T_{\#}\mu)(C) = \mu(T^{-1}(C))$, satisfies the **change-of-variables formula** (for measurable $g \geq 0$),

$$\int_{\mathcal{Y}} g \, d(T_{\#}\mu) = \int_{\mathcal{X}} (g \circ T) \, d\mu.$$

- T need only be *measurable*; e.g. $Z \sim \mathcal{N}(0, I)$, $T(z) = Az + b$ gives $T_{\#}\mathcal{N}(0, I) = \mathcal{N}(b, AA^{\top})$. **Every model in the seminar will be such a pushforward.**

Densities and the Radon–Nikodym theorem

- μ is **absolutely continuous** w.r.t. ν (write $\mu \ll \nu$) if $\nu(A) = 0 \Rightarrow \mu(A) = 0$.
- **Radon–Nikodym**: if $\mu \ll \nu$ (σ -finite) there is a density $p = \frac{d\mu}{d\nu}$ with $\mu(A) = \int_A p \, d\nu$.
- Taking $\nu = \text{Lebesgue}$ recovers the usual probability density — but not every law has one: a law on a lower-dimensional set is **singular** (Lebesgue decomposition $\mu = \mu_{\text{ac}} + \mu_{\text{s}}$).
- Smooth invertible transport changes densities by the Jacobian:

$$p_{T\#\mu}(y) = p_\mu\left(T^{-1}(y)\right) \left| \det DT^{-1}(y) \right|.$$

Joint measures, marginals, and conditioning

- On a product space $\mathcal{X} \times \mathcal{Y}$: the product σ -algebra, product measures, and *Fubini–Tonelli* (iterated integration).
- **Marginals** of a joint law γ are its pushforwards under the coordinate projections, $(\pi_{\mathcal{X}})_{\#}\gamma$ and $(\pi_{\mathcal{Y}})_{\#}\gamma$.
- A **coupling** of P, Q is a joint law with marginals P and Q ; write $\Pi(P, Q)$ for the set of all couplings.
- **Conditional distributions** (Markov kernels) $K(x, \mathrm{d}y)$ disintegrate a joint law,

$$\gamma(\mathrm{d}x, \mathrm{d}y) = P(\mathrm{d}x) K(x, \mathrm{d}y),$$

the rigorous meaning of a “conditional density $p(y \mid x)$ ”.

Comparing measures; convergence

On the space $\mathcal{P}(\mathcal{X})$ of probability measures several notions of “distance” coexist:

- **Total variation** $\|P - Q\|_{\text{TV}} = \sup_A |P(A) - Q(A)|$.
- **Kullback–Leibler** (needs $P \ll Q$) $\text{KL}(P\|Q) = \int \log \frac{dP}{dQ} dP \geq 0$ — a divergence, not a metric.
- **Wasserstein** $W_1(P, Q) = \inf_{\gamma \in \Pi(P, Q)} \int \|x - y\| d\gamma$ (optimal transport).

These induce *different topologies*. The notion matching “samples look right” is **weak convergence**: $P_n \Rightarrow P$ iff $\int f dP_n \rightarrow \int f dP$ for all bounded continuous f .

Outlook: where this returns in the papers

- **Pushforward** — every model is a transported measure $x = T_\theta(z)$, $z \sim \mu$: decoder (VAE), generator (GAN), reverse process (DDPM, Score-SDE).
- **Densities / Radon–Nikodym** — likelihoods $p_\theta(x)$, the ELBO, and Jacobians when T_θ is invertible.
- **Conditional kernels** — the VAE encoder $q_\phi(z | x)$ and decoder; the DDPM forward/reverse Markov chain.
- **Couplings** — the Wasserstein distance behind WGAN.
- **Divergences** — KL in the VAE and DDPM objectives; Jensen–Shannon behind the GAN game.
- **Weak convergence & distances on $\mathcal{P}(\mathcal{X})$** — why Wasserstein is used, and the sense in which the model law approaches the data.