JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Lookahead Diffusion Probabilistic Models for Refining Mean Estimation

Guoqiang Zhang[1], Kenta Niwa[2], W. Bastiaan Kleijn[3]

[1]University of Technology Sydney, Australia

[2]NTT Communication Science Laboratories, Japan

[3]Victoria University of Wellington, New Zealand

TUE-AM-135

UTS
UNIVERSITY OF TECHNOLOGY, SYDNEY
NTT
VICTORIA UNIVERSITY WELLINGTON
TE WHARE WĀNANGA O TE ŪPOKO O TE IKA A MĀUI

Background

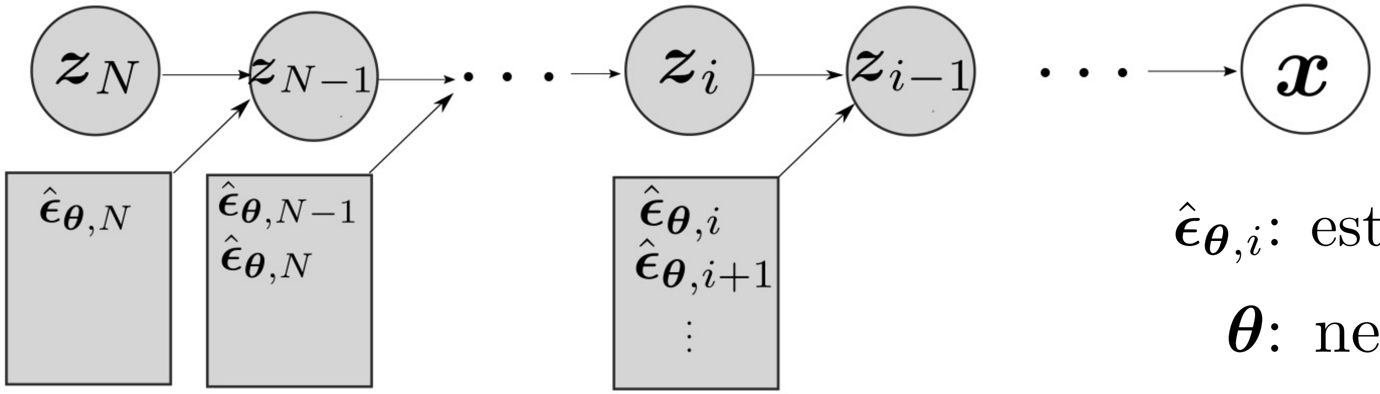
- A forward diffusion process

$$\boldsymbol{z}_t = \alpha_t \boldsymbol{x} + \sigma_t \boldsymbol{\epsilon}_t \quad \boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \boldsymbol{I}), \boldsymbol{x} \sim p_{data}(\boldsymbol{x})$$

- Reverse *probability-flow ODE for sampling*

$$\frac{d\boldsymbol{z}_t}{dt} = \frac{d \log \alpha_t}{dt} \boldsymbol{z}_t + \left(\frac{1}{2\sigma_t} \frac{d\sigma_t^2}{dt} - \frac{d \log \alpha_t}{dt} \sigma_t \right) \hat{\boldsymbol{\epsilon}}_{\boldsymbol{\theta},t} \quad \boldsymbol{z}_{T=1} \sim \mathcal{N}(0, \tilde{\sigma} \boldsymbol{I})$$

- Popular ODE solvers: DDIM, DEIS, SPNDM

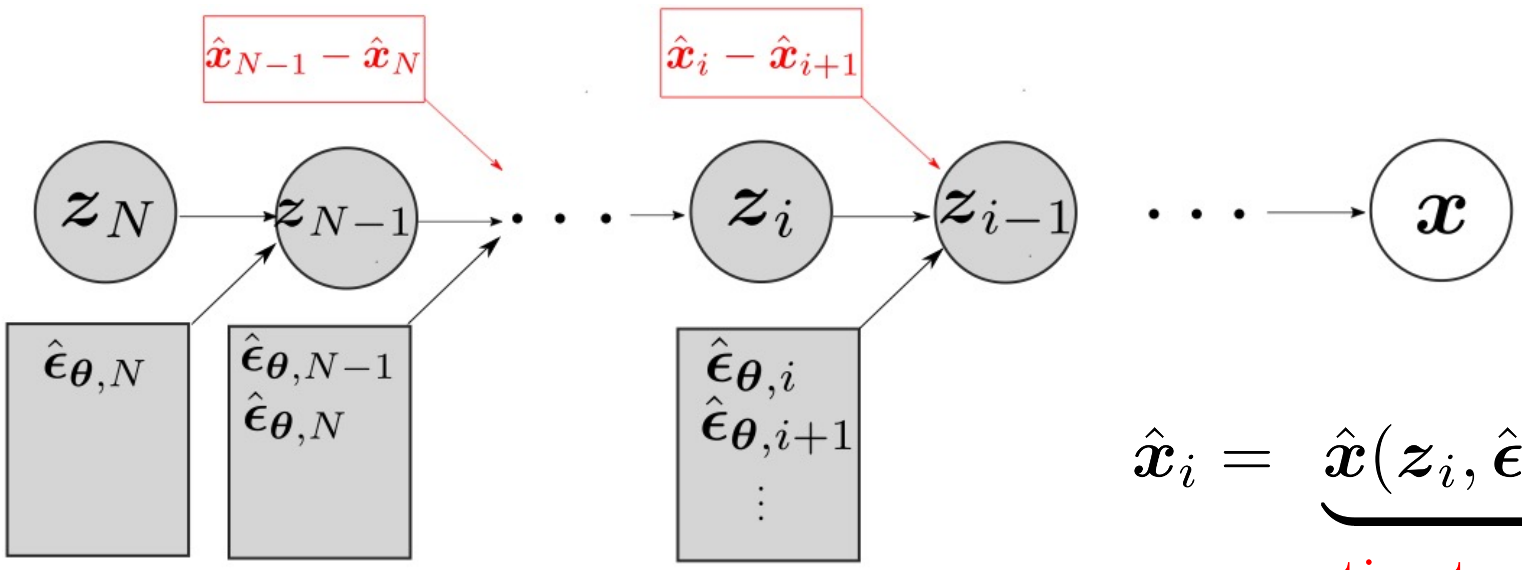


$\hat{\boldsymbol{\epsilon}}_{\boldsymbol{\theta},i}$: estimated Gaussian noise

$\boldsymbol{\theta}$: neural network

Proposed Lookahead Technique (1)

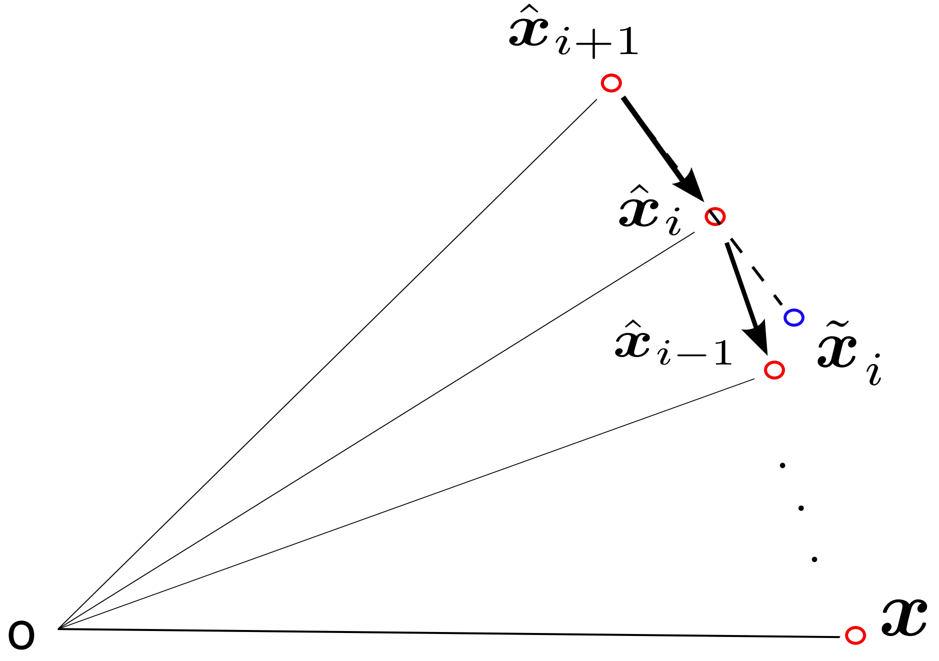
- Objective: To improve performance of existing ODE solvers
- Basic idea: To exploit $\hat{\boldsymbol{x}}_i - \hat{\boldsymbol{x}}_{i+1}$ in computation of $\boldsymbol{z}_{i-1}$



$$\hat{\boldsymbol{x}}_i = \underbrace{\hat{\boldsymbol{x}}(\boldsymbol{z}_i, \hat{\boldsymbol{\epsilon}}_{\boldsymbol{\theta},i}, \hat{\boldsymbol{\epsilon}}_{\boldsymbol{\theta},i+1}, \ldots)}_{\text{estimator of } \boldsymbol{x} \text{ at timestep } i}$$

Proposed Lookahead Technique (2)

- Assumption: $\hat{\boldsymbol{x}}_i$ is increasingly accurate as i decreases to 0.
- Perform extrapolation to better estimate $\boldsymbol{x}$



$$\tilde{\boldsymbol{x}}_i = \hat{\boldsymbol{x}}_i + \underbrace{\lambda}_{\text{stepsize}} \underbrace{(\hat{\boldsymbol{x}}_i - \hat{\boldsymbol{x}}_{i+1})}_{\text{gradient}}$$

Existing ODE Solvers

- Update expressions of (DDIM, DEIS, SPNDM)

$$\tilde{\boldsymbol{\epsilon}}_{[i:i+r]} = \sum_{j=0}^{r} c_{ij} \hat{\boldsymbol{\epsilon}}_{\boldsymbol{\theta}, i+j} \qquad \text{[linear combination of order } r\text{]}$$

$$\boldsymbol{z}_{i-1} = \alpha_{i-1} \underbrace{\left(\frac{\boldsymbol{z}_i - \sigma_i \tilde{\boldsymbol{\epsilon}}_{[i:i+r]}}{\alpha_i} \right)}_{\hat{\boldsymbol{x}}_i} + \sigma_{i-1} \tilde{\boldsymbol{\epsilon}}_{[i:i+r]},$$

  - When order $r = 0$, it reduces to DDIM

Lookahead-Based ODE Solvers

- Update expressions

$$\tilde{\boldsymbol{\epsilon}}_{[i:i+r]} = \sum_{j=0}^{r} c_{ij}\hat{\boldsymbol{\epsilon}}_{\boldsymbol{\theta}}(\boldsymbol{z}_{i+j}, i+j) \qquad \text{[linear combination of order } r\text{]}$$

$$\hat{\boldsymbol{x}}_i = \left(\frac{\boldsymbol{z}_i - \sigma_i\tilde{\boldsymbol{\epsilon}}_{[i:i+r]}}{\alpha_i}\right)$$

$$\boldsymbol{z}_{i-1} = \alpha_{i-1}\big[\hat{\boldsymbol{x}}_i + \underbrace{\lambda(\hat{\boldsymbol{x}}_i - \hat{\boldsymbol{x}}_{i+1})}_{\text{extrapolation}}\big] + \sigma_{i-1}\tilde{\boldsymbol{\epsilon}}_{[i:i+r]}$$

- The extrapolation $\lambda(\widetilde{\boldsymbol{x}}_i - \widetilde{\boldsymbol{x}}_{i-1})$ provides additional gradient information to better estimate $\boldsymbol{x}$
- **Theoretical analysis** is provided in the paper, showing that positive λ improves estimation accuracy of $\boldsymbol{x}$ under certain assumptions

Experimental Results (0)

- Summary of Experiments
  - Lookahead technique in DDIM and DDPM
  - Lookahead technique in DEIS
  - Lookahead technique in SPNDM
  - Lookahead technique in consistency models (Song et al. 2023)

- All the above experiments produce positive results, indicating the effectiveness of the lookahead technique.

Experimental Results (1)

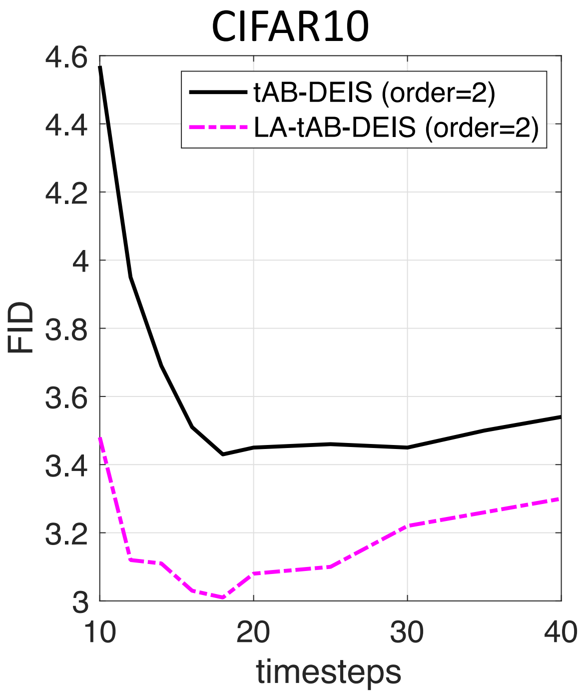
➢ Lookahead technique in DDIM and DDPM

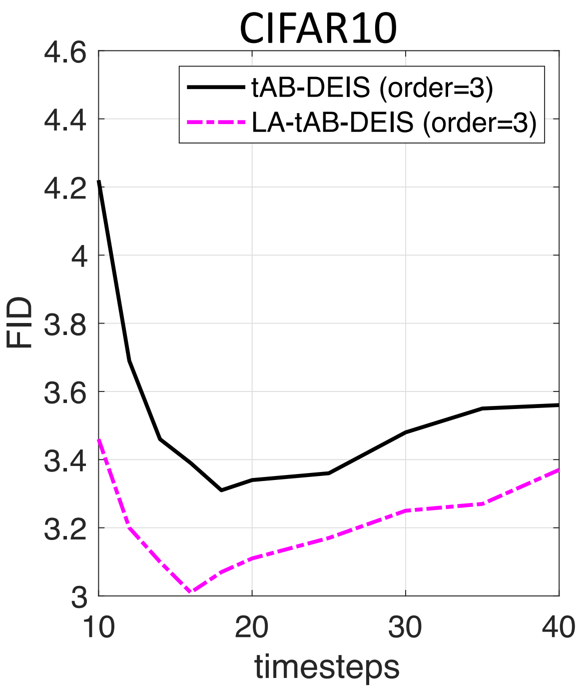
Table 1: Comparison of FID score for CIFAR10, CelebA64, and ImageNet64. **Lower** is better.

Data sets	CIFAR10						CelebA64						ImageNet64					
Timesteps	10	25	50	100	200	1000	10	25	50	100	200	1000	10	25	50	100	200	1000
NPR-DDPM	32.64	10.48	6.18	4.46	3.70	4.04	28.32	15.51	10.70	8.28	7.01	5.26	53.22	28.41	21.05	18.26	**16.75**	16.30
LA-NPR-DDPM	**25.59**	**8.48**	**5.28**	**4.07**	**3.47**	**3.90**	**25.08**	**13.92**	**9.58**	**7.43**	**6.32**	**5.01**	**48.71**	**25.42**	**20.27**	**18.16**	16.83	**16.27**
gain (%)	21.6	19.1	14.6	8.7	6.2	3.5	11.4	10.3	10.4	10.3	9.8	4.75	8.5	10.5	3.7	0.5	-0.5	0.2
SN-DDPM	23.75	6.88	4.58	3.67	3.31	3.65	20.55	11.85	7.58	5.95	4.96	4.44	51.09	27.77	20.65	18.07	**16.70**	16.30
LA-SN-DDPM	**19.75**	**5.92**	**4.31**	**3.55**	**3.24**	**3.55**	**17.43**	**10.08**	**6.41**	**5.12**	**4.41**	**4.21**	**46.13**	**24.67**	**19.83**	**17.95**	16.76	**16.28**
gain (%)	16.8	14.0	5.9	3.3	2.1	2.7	15.2	14.9	15.4	13.9	11.1	5.2	9.7	11.2	4.0	0.7	-0.4	0.1
NPR-DDIM	13.41	5.43	3.99	3.53	3.40	3.67	14.94	9.18	6.17	4.40	3.67	3.12	97.27	28.75	19.79	**17.71**	**17.15**	**17.59**
LA-NPR-DDIM	**10.74**	**4.71**	**3.64**	**3.33**	**3.29**	**3.49**	**14.25**	**8.83**	**5.67**	**3.76**	**2.95**	**2.95**	**71.98**	**25.39**	**19.47**	18.11	17.89	18.41
gain (%)	19.9	13.3	8.8	5.7	3.2	4.9	4.6	3.8	8.1	14.5	19.61	5.4	26.0	11.7	1.6	-2.3	-4.3	-4.7
SN-DDIM	12.19	4.28	3.39	3.22	4.22	3.65	10.17	5.62	3.90	3.21	2.94	2.84	91.29	27.74	19.51	**17.67**	**17.14**	**17.60**
LA-SN-DDIM	**8.48**	**3.15**	**2.93**	**2.92**	**3.08**	**3.47**	**8.05**	**4.56**	**2.93**	**2.39**	**2.19**	**2.48**	**69.11**	**25.07**	**19.32**	18.06	17.89	18.57
gain (%)	30.4	26.4	13.6	9.3	27.0	4.9	20.8	18.9	24.9	25.5	25.5	12.7	24.3	9.6	9.7	-2.2	-4.4	-5.5

Experimental Results (2)

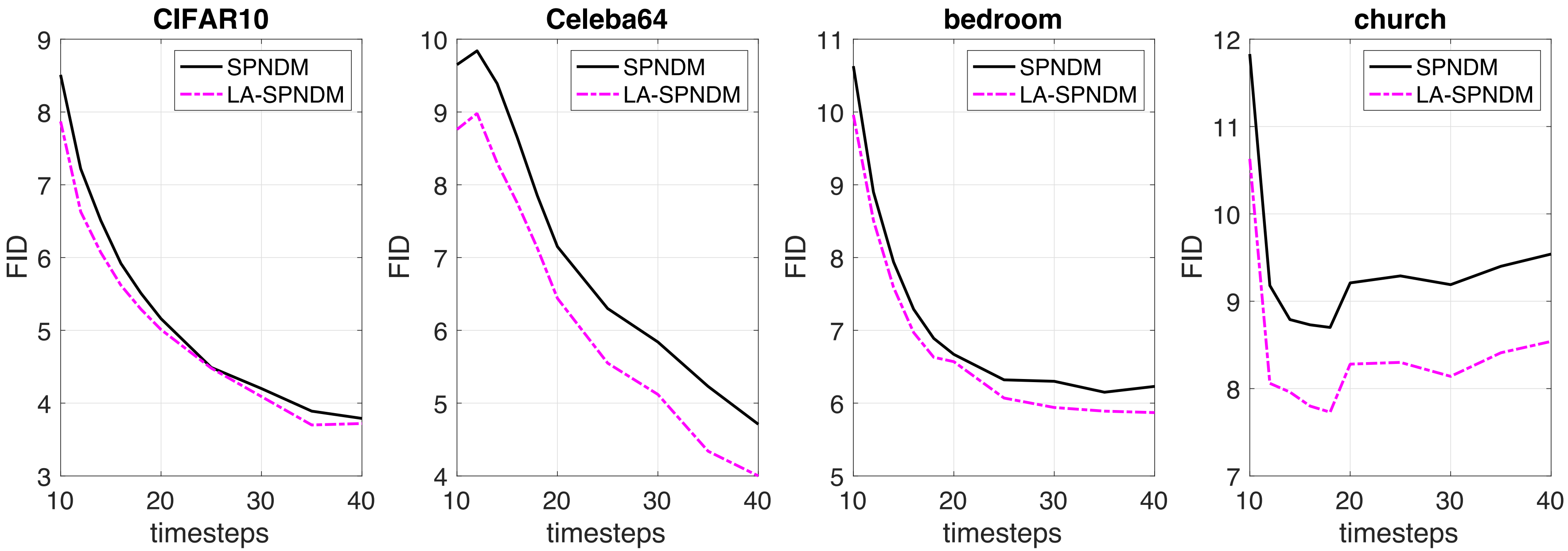
- Lookahead technique in DEIS





Experimental Results (3)

- Lookahead technique in SPNDM



Experimental Results (4)

- Lookahead technique in consistency models (Song et al. 2023)

	FID over ImageNet64	
time-steps	3	4
CD (LPIPS)	4.99	4.75
LA-CD (LPIPS)	**4.78**	**4.65**

CD: consistency distillation
LPIPS: A DNN-based distance criterion

- The above results are obtained recently, and are not included in the paper
- Song et al., “Consistency Models”, arXiv:2303:01469 [cs.LG], 2023.

Conclusions and New Results

- Conclusions
  - Extrapolation on both the estimated clean images and estimated Gaussian noises are compatible.
  - However, manual tuning of λ is needed.

- Our recent progress
  - Developed a new approach that can learn the optimal strengths of the extrapolations (no manual-tuning is required any more).
  - The new paper will be made public on arXiv soon.