Appendix

In this appendix, we follow the notation introduced before. However, in order to reduce the complexity of the derivation, we simplify all the coefficient with Kronecker product. For example $\mathbf{A}_t := \mathbf{A}_t \otimes \mathbf{I}_d$, $\boldsymbol{\mu}_t := \boldsymbol{\mu}_t \otimes \mathbf{I}_d$.

A Proof of Proposition 3.1

Proof. We first analyze the objective function of momentum diffusion model for $N = 2$ case, and it can be generalize to larger N.

$$\min_{\theta} \mathcal{L}_{\text{MDM}}(\theta) = \mathbb{E}_{x_1,\boldsymbol{\epsilon},t} \|\epsilon_\theta(\mathbf{x}_t, t) - \epsilon^{(1)}\|_2^2 \tag{6}$$

$$= \mathbb{E}_{x_1,\boldsymbol{\epsilon},t} \frac{1}{L_t^{vv2}} \|L_t^{vv}\epsilon_\theta(\mathbf{x}_t, t) - L_t^{vv}\epsilon^{(1)}\|_2^2 \tag{7}$$

$$= \mathbb{E}_{x_1,\boldsymbol{\epsilon},t} \frac{1}{L_t^{vv2}} \|L_t^{vv}\epsilon_\theta(\mathbf{x}_t, t) - \left[x_t^{(1)} - \frac{L_t^{xv}}{L_t^{xx}} x_t^{(0)} - \left(\boldsymbol{\mu}_t{}^{(0)} - \frac{L_t^{xv}}{L_t^{xx}} \boldsymbol{\mu}_t{}^{(1)}\right) x_1\right] \|_2^2 \tag{8}$$

$$= \mathbb{E}_{x_1,\boldsymbol{\epsilon},t} \frac{1}{L_t^{vv2}} \|L_t^{vv}\epsilon_\theta(\mathbf{x}_t, t) - x_t^{(1)} + \frac{L_t^{xv}}{L_t^{xx}} x_t^{(0)} - \left(\frac{L_t^{xv}}{L_t^{xx}} \boldsymbol{\mu}_t{}^{(1)} - \boldsymbol{\mu}_t{}^{(0)}\right) x_1\|_2^2 \tag{9}$$

$$= \mathbb{E}_{x_1,\boldsymbol{\epsilon},t} \frac{\left(\frac{L_t^{xv}}{L_t^{xx}} \boldsymbol{\mu}_t{}^{(1)} - \boldsymbol{\mu}_t{}^{(0)}\right)^2}{L_t^{vv2}} \| \underbrace{\frac{L_t^{vv}\epsilon_\theta(\mathbf{x}_t, t) - x_t^{(1)} + \frac{L_t^{xv}}{L_t^{xx}} x_t^{(0)}}{\frac{L_t^{xv}}{L_t^{xx}} \boldsymbol{\mu}_t{}^{(1)} - \boldsymbol{\mu}_t{}^{(0)}}}_{\text{parameterized Neural Netowrk}} - x_1\|_2^2 \tag{10}$$

Following the same spirit, one can derive the case for N variable. See Appendix I for details.

We know that,

$$\mathbf{x}_t \mid x_1 \sim \mathcal{N}(\boldsymbol{\mu}_t x_1,\ \boldsymbol{\Sigma}_t),$$

Define

$$\mathbf{r}_t := \frac{\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\mu}_t}{\boldsymbol{\mu}_t{}^{\intercal}\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\mu}_t}, \quad y_t := \mathbf{r}_t^{\top}\mathbf{x}_t,$$

and the residual ("noise")

$$\boldsymbol{\epsilon} := \mathbf{x}_t - y_t\, \boldsymbol{\mu}_t.$$

Since the operation is linear, one can transform the $\mathbf{x}_t$ by

$$T:\ \mathbf{x}_t\ \mapsto\ (y_t, \boldsymbol{\epsilon}) = \left(\mathbf{r}_t^{\top}\mathbf{x}_t,\ \mathbf{x}_t - \tfrac{(\mathbf{r}_t^{\top}\mathbf{x}_t)}{a_t}\boldsymbol{\mu}_t\right)$$

and it is *invertible* with linear inverse $\mathbf{x}_t = y_t\boldsymbol{\mu}_t + \boldsymbol{\epsilon}$. Hence the σ-algebras coincide:

$$\sigma(\mathbf{x}_t) = \sigma(y_t, \boldsymbol{\epsilon}).$$

For any integrable random variable Z, equal σ-fields imply $\mathbb{E}[Z \mid \mathbf{x}_t] = \mathbb{E}[Z \mid y_t, \boldsymbol{\epsilon}]$. Taking $Z = x_1$ yields

$$\mathbb{E}[x_1 \mid \mathbf{x}_t] = \mathbb{E}[x_1 \mid y_t, \boldsymbol{\epsilon}].$$

due to the fact that $\boldsymbol{\epsilon}$ is the independent gaussian, thus

$$\mathbb{E}[x_1 \mid \mathbf{x}_t] = \mathbb{E}[x_1 \mid y_t, \boldsymbol{\epsilon}] = \mathbb{E}[x_1 \mid y_t, \boldsymbol{\epsilon}] = \mathbb{E}[x_1 \mid y_t].$$

□

B Proof of Proposition 3.3

Proof. The dynamics we considered reads

$$\frac{\mathrm{d}\mathbf{x}_t}{dt} = \mathbf{A}_t\, \mathbf{x}_t + \mathbf{b}_t F_t, \quad x_0 \sim \mathcal{N}(0, I). \tag{11}$$

and again, in this paper, we only consider,

$$\mathbf{A}_t = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}_{N\times N}, \text{and} \quad \mathbf{b}_t := \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \tag{12}$$

If we expand the system, it basically represent:

$$\mathrm{d}x_t^{(0)} = x_t^{(1)}\mathrm{d}t \tag{13}$$
$$\mathrm{d}x_t^{(1)} = x_t^{(2)}\mathrm{d}t \tag{14}$$
$$\cdots \tag{15}$$
$$\mathrm{d}x_t^{(N-1)} = F_t\mathrm{d}t \tag{16}$$
$$\tag{17}$$

In our case, the F_t function is:

$$F_t := N!\frac{x_1 - \sum_{i=0}^{N-1}\frac{x_t^{(i)}}{i!}(1-t)^i}{(1-t)^N} \tag{18}$$

One can easily verify that, when $N = 1$, it actually degenerate to flow matching:

$$F_t := \frac{x_1 - \mathbf{x}_t}{1-t}, \text{and} \tag{19}$$
$$\mathrm{d}x_t^{(0)} = F_t\mathrm{d}t \tag{20}$$

One can consider it as the higher augment dimension extension of flow matching model. And the magical part is that, we do not need to retrain model.

For better analysis, we rearange the system:

$$\frac{\mathrm{d}\mathbf{x}_t}{dt} = \mathbf{A}_t\,\mathbf{x}_t + \mathbf{b}_t F_t \tag{21}$$
$$= \mathbf{A}_t\,\mathbf{x}_t + \mathbf{b}_t N!\frac{x_1 - \sum_{i=0}^{N-1}\frac{x_t^{(i)}}{i!}(1-t)^i}{(1-t)^N} \tag{22}$$
$$= \mathbf{A}_t\,\mathbf{x}_t + \frac{\mathbf{b}_t N!}{(1-t)^N}x_1 - \frac{\mathbf{b}_t N!\sum_{i=0}^{N-1}\frac{x_t^{(i)}}{i!}(1-t)^i}{(1-t)^N} \tag{23}$$
$$= \mathbf{A}_t\,\mathbf{x}_t + \underbrace{\begin{bmatrix} 0 \\ 0 \\ \vdots \\ \frac{N!}{(1-t)^N} \end{bmatrix}}_{\hat{\mathbf{b}}_t} x_1 - \underbrace{\begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \frac{(1-t)^0 N!}{0!(1-t)^N} & \frac{(1-t)^1 N!}{1!(1-t)^N} & \cdots & \frac{(1-t)^N N!}{(N-1)!(1-t)^N} \end{bmatrix}}_{\tilde{\mathbf{A}}_t} \mathbf{x}_t \tag{24}$$
$$:= \hat{\mathbf{A}}_t\mathbf{x}_t + \hat{\mathbf{b}}_t x_1, \quad (\hat{\mathbf{A}}_t := \mathbf{A}_t - \tilde{\mathbf{A}}_t) \tag{25}$$

Consider the linear time-varying system:

$$\frac{d\mathbf{x}_t}{dt} = \hat{\mathbf{A}}_t\,\mathbf{x}_t + \hat{\mathbf{b}}_t\,x_1, \quad \mathbf{x}_0 \sim \mathcal{N}(0, I). \tag{26}$$

Since (26) is linear and deterministic (apart from the random initial condition), the state remains Gaussian. Its mean and covariance evolve as follows [26].

Mean Dynamics. Let
$$\mathbf{m}_t = \mathbb{E}[\mathbf{x}_t].$$
Then
$$\dot{\mathbf{m}}_t = \hat{\mathbf{A}}_t\,\mathbf{m}_t + \hat{\mathbf{b}}_t\,x_1, \quad m_0 = 0. \tag{27}$$
We can write the mean in a factorized form as
$$\mathbf{m}_t = \boldsymbol{\mu}_t\,x_1,$$
so that by dividing by x_1, we obtain
$$\dot{\boldsymbol{\mu}}_t = \hat{\mathbf{A}}_t\,\boldsymbol{\mu}_t + \hat{\mathbf{b}}_t.$$

Covariance Dynamics. Similarly, the convariance follows dynamics
$$\dot{\boldsymbol{\Sigma}}_t = \hat{\mathbf{A}}_t\,\boldsymbol{\Sigma}_t + \boldsymbol{\Sigma}_t\,\hat{\mathbf{A}}_t^T \tag{28}$$
Recall that, Then we define the scalar quantity of interest as
$$y_t := \frac{\left(\boldsymbol{\Sigma}_t{}^{-1}\mathbf{m}_t\right)^T\mathbf{x}_t}{\left(\boldsymbol{\Sigma}_t{}^{-1}\mathbf{m}_t\right)^T\boldsymbol{\mu}_t} = \frac{\left(\boldsymbol{\Sigma}_t{}^{-1}\boldsymbol{\mu}_t\right)^T\mathbf{x}_t}{\boldsymbol{\mu}_t{}^T\boldsymbol{\Sigma}_t{}^{-1}\boldsymbol{\mu}_t} = \frac{\mathbf{r}_t^T\mathbf{x}_t}{\gamma_t}, \quad \text{with } \gamma_t := \boldsymbol{\mu}_t{}^T\boldsymbol{\Sigma}_t{}^{-1}\boldsymbol{\mu}_t. \tag{29}$$

We wish to compute the derivative of
$$y_t = \frac{\mathbf{r}_t^T\mathbf{x}_t}{\gamma_t},$$
Using the quotient rule,
$$\dot{y}_t = \frac{\frac{d}{dt}\left(\mathbf{r}_t^T\mathbf{x}_t\right)\,\gamma_t - \left(\mathbf{r}_t^T\mathbf{x}_t\right)\,\dot{\gamma}_t}{\gamma_t^2}.$$
Since $\mathbf{r}_t^T\mathbf{x}_t = y_t\gamma_t$, this becomes
$$\dot{y}_t = \frac{\dot{\mathbf{r}}_t^T\mathbf{x}_t + \mathbf{r}_t^T\dot{x}_t}{\gamma_t} - y_t\,\frac{\dot{\gamma}_t}{\gamma_t}.$$

Derivative of $\mathbf{r}_t$

Recall that
$$\mathbf{r}_t = \boldsymbol{\Sigma}_t{}^{-1}\boldsymbol{\mu}_t.$$
Differentiating gives
$$\dot{\mathbf{r}}_t = -\boldsymbol{\Sigma}_t{}^{-1}\dot{\boldsymbol{\Sigma}}_t\,\boldsymbol{\Sigma}_t{}^{-1}\boldsymbol{\mu}_t + \boldsymbol{\Sigma}_t{}^{-1}\dot{\boldsymbol{\mu}}_t.$$
Substitute the known dynamics:
$$\dot{\boldsymbol{\Sigma}}_t = \hat{\mathbf{A}}_t\,\boldsymbol{\Sigma}_t + \boldsymbol{\Sigma}_t\,\hat{\mathbf{A}}_t^T,$$
$$\dot{\boldsymbol{\mu}}_t = \hat{\mathbf{A}}_t\,\boldsymbol{\mu}_t + \hat{\mathbf{b}}_t.$$
It follows that
$$\begin{aligned}
\dot{\mathbf{r}}_t &= -\boldsymbol{\Sigma}_t{}^{-1}\dot{\boldsymbol{\Sigma}}_t\,\boldsymbol{\Sigma}_t{}^{-1}\boldsymbol{\mu}_t + \boldsymbol{\Sigma}_t{}^{-1}\dot{\boldsymbol{\mu}}_t && (30)\\
&= -\boldsymbol{\Sigma}_t{}^{-1}\left(\hat{\mathbf{A}}_t\,\boldsymbol{\Sigma}_t + \boldsymbol{\Sigma}_t\,\hat{\mathbf{A}}_t^T\right)\boldsymbol{\Sigma}_t{}^{-1}\boldsymbol{\mu}_t + \boldsymbol{\Sigma}_t{}^{-1}\left(\hat{\mathbf{A}}_t\,\boldsymbol{\mu}_t + \hat{\mathbf{b}}_t\right) && (31)\\
&= -\boldsymbol{\Sigma}_t{}^{-1}\hat{\mathbf{A}}_t\boldsymbol{\mu}_t - \hat{\mathbf{A}}_t^{\mathsf{T}}\boldsymbol{\Sigma}_t{}^{-1}\boldsymbol{\mu}_t + \boldsymbol{\Sigma}_t{}^{-1}\hat{\mathbf{A}}_t\boldsymbol{\mu}_t + \boldsymbol{\Sigma}_t{}^{-1}\hat{\mathbf{b}}_t && (32)\\
&= -\,\hat{\mathbf{A}}_t^{\mathsf{T}}\mathbf{r}_t + \boldsymbol{\Sigma}_t{}^{-1}\,\hat{\mathbf{b}}_t. && (33)
\end{aligned}$$

Derivative of $\mathbf{x}_t$

From (26),
$$\dot{\mathbf{x}}_t = \hat{\mathbf{A}}_t\,\mathbf{x}_t + \hat{\mathbf{b}}_t\,\hat{x}_1.$$

Derivative of γ_t

Recall
$$\gamma_t = \boldsymbol{\mu}_t{}^T\,\boldsymbol{\Sigma}_t{}^{-1}\,\boldsymbol{\mu}_t = \mathbf{r}_t^T\,\boldsymbol{\mu}_t.$$
Differentiating,
$$\dot{\gamma}_t = \dot{\mathbf{r}}_t^T\,\boldsymbol{\mu}_t + \mathbf{r}_t^T\,\dot{\boldsymbol{\mu}}_t.$$
Using (30) and $\dot{\boldsymbol{\mu}}_t = \hat{A}_t\boldsymbol{\mu}_t + \hat{\mathbf{b}}_t$, one obtains (after cancellation) the result:
$$\begin{aligned}\dot{\gamma}_t &= \Big(-\,\hat{\mathbf{A}}_t^{\mathsf{T}}\mathbf{r}_t + \boldsymbol{\Sigma}_t{}^{-1}\,\hat{\mathbf{b}}_t\Big)\,\boldsymbol{\mu}_t + \mathbf{r}^{\mathsf{T}}\,\Big(\hat{\mathbf{A}}_t\boldsymbol{\mu}_t + \hat{\mathbf{b}}_t\Big) && (34)\\ &= 2\,\hat{\mathbf{b}}_t^T\,\mathbf{r}_t && (35)\\ &= 2\,\hat{\mathbf{b}}_t^T\,\boldsymbol{\Sigma}_t{}^{-1}\boldsymbol{\mu}_t && (36)\end{aligned}$$

Combining Everything

Substitute the pieces into
$$\dot{y}_t = \frac{\dot{\mathbf{r}}_t^T\mathbf{x}_t + \mathbf{r}_t^T\dot{\mathbf{x}}_t}{\gamma_t} - y_t\,\frac{\dot{\gamma}_t}{\gamma_t}.$$
Using
$$\begin{aligned}\dot{\mathbf{r}}_t^T &= -\,\mathbf{r}_t^T\hat{\mathbf{A}}_t + \hat{\mathbf{b}}_t^T\,\boldsymbol{\Sigma}_t{}^{-1},\\ \mathbf{r}_t^T\dot{\mathbf{x}}_t &= \mathbf{r}_t^T\big(\hat{\mathbf{A}}_t\mathbf{x}_t + \hat{\mathbf{b}}_t\,\hat{x}_1\big),\end{aligned}$$
we have:
$$\begin{aligned}\dot{\mathbf{r}}_t^T\mathbf{x}_t + \mathbf{r}_t^T\dot{\mathbf{x}}_t &= \Big[-\,\mathbf{r}_t^T\,\hat{\mathbf{A}}_t\,\mathbf{x}_t + \hat{\mathbf{b}}_t^T\,\boldsymbol{\Sigma}_t{}^{-1}\,\mathbf{x}_t\Big] + \Big[\mathbf{r}_t^T\,\hat{\mathbf{A}}_t\,\mathbf{x}_t + \mathbf{r}_t^T\,\hat{\mathbf{b}}_t\,\hat{x}_1\Big]\\ &= \hat{\mathbf{b}}_t^T\,\boldsymbol{\Sigma}_t{}^{-1}\,\mathbf{x}_t + \mathbf{r}_t^T\,\hat{\mathbf{b}}_t x_1\end{aligned}$$
Thus,
$$\dot{y}_t = \frac{\hat{\mathbf{b}}_t^T\,\boldsymbol{\Sigma}_t{}^{-1}\,\mathbf{x}_t + x_1\,\mathbf{r}_t^T\,\hat{\mathbf{b}}_t}{\gamma_t} - y_t\,\frac{\dot{\gamma}_t}{\gamma_t}. \tag{37}$$
Recall that $\gamma_t = \boldsymbol{\mu}_t{}^T\,\boldsymbol{\Sigma}_t{}^{-1}\,\boldsymbol{\mu}_t$ and $\dot{\gamma}_t = 2\,\hat{\mathbf{b}}_t^T\,\boldsymbol{\Sigma}_t{}^{-1}\,\boldsymbol{\mu}_t$. Also note that
$$\mathbf{r}_t^T\,\hat{\mathbf{b}}_t = \boldsymbol{\mu}_t{}^T\,\boldsymbol{\Sigma}_t{}^{-1}\,\hat{\mathbf{b}}_t.$$
Thus, the final expression becomes
$$\begin{aligned}\dot{y}_t &= \frac{\hat{\mathbf{b}}_t^T\,\boldsymbol{\Sigma}_t{}^{-1}\,\mathbf{x}_t + \hat{x}_1\,\boldsymbol{\mu}_t{}^T\,\boldsymbol{\Sigma}_t{}^{-1}\,\hat{\mathbf{b}}_t}{\boldsymbol{\mu}_t{}^T\,\boldsymbol{\Sigma}_t{}^{-1}\,\boldsymbol{\mu}_t} - y_t\,\frac{2\,\hat{\mathbf{b}}_t^T\,\boldsymbol{\Sigma}_t{}^{-1}\,\boldsymbol{\mu}_t}{\boldsymbol{\mu}_t{}^T\,\boldsymbol{\Sigma}_t{}^{-1}\,\boldsymbol{\mu}_t} && (38)\\ &= \underbrace{\frac{\hat{\mathbf{b}}_t^T\,\boldsymbol{\Sigma}_t{}^{-1}}{\boldsymbol{\mu}_t{}^T\,\boldsymbol{\Sigma}_t{}^{-1}\,\boldsymbol{\mu}_t}}_{\mathbf{e}_t}\mathbf{x}_t + \frac{\boldsymbol{\mu}_t{}^T\,\boldsymbol{\Sigma}_t{}^{-1}\,\hat{\mathbf{b}}_t}{\boldsymbol{\mu}_t{}^T\,\boldsymbol{\Sigma}_t{}^{-1}\,\boldsymbol{\mu}_t}x_1 - \frac{2\,\hat{\mathbf{b}}_t^T\,\boldsymbol{\Sigma}_t{}^{-1}\,\boldsymbol{\mu}_t}{\boldsymbol{\mu}_t{}^T\,\boldsymbol{\Sigma}_t{}^{-1}\,\boldsymbol{\mu}_t}y_t && (39)\end{aligned}$$

The frist term is essentially one kind of linear combination of $\mathbf{x}_t$, and Recally that $y_t = \mathbf{r}^{\mathsf{T}}\mathbf{x}_t := \frac{\boldsymbol{\mu}_t^{\mathsf{T}}\boldsymbol{\Sigma}_t^{-1}}{\boldsymbol{\mu}_t^{\mathsf{T}}\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\mu}_t}$ which is another linear combination of $\mathbf{x}_t$. Assume that $\mathbf{x}_t \sim \mathcal{N}(\boldsymbol{\mu}_t x_1, \boldsymbol{\Sigma}_t)$, thus, one can derive the relationship between the frist term and y_t. Thus, according to Lemma H.1
$$\begin{aligned}\frac{\hat{\mathbf{b}}_t^T\,\boldsymbol{\Sigma}_t{}^{-1}}{\boldsymbol{\mu}_t{}^T\,\boldsymbol{\Sigma}_t{}^{-1}\,\boldsymbol{\mu}_t}\mathbf{x}_t &= \mathbf{e}_t^{\mathsf{T}}\left[\mathbf{I} - \frac{\boldsymbol{\Sigma}_t\mathbf{r}_t\mathbf{r}_t^{\mathsf{T}}}{\mathbf{r}_t^{\mathsf{T}}\boldsymbol{\Sigma}_t\mathbf{r}_t}\right]\boldsymbol{\mu}_t x_1 + \frac{\mathbf{e}_t^{\mathsf{T}}\boldsymbol{\Sigma}_t\mathbf{r}_t}{\mathbf{r}_t^{\mathsf{T}}\boldsymbol{\Sigma}_t\mathbf{r}_t}y_t + \mathbf{e}_t^{\mathsf{T}}\mathbf{L}_t\boldsymbol{\epsilon}_\perp, && (40)\\ \boldsymbol{\epsilon}_\perp &\sim \mathcal{N}\big(\mathbf{0},\ I_d - \mathbf{L}_t^{\mathsf{T}}\mathbf{r}_t\mathbf{r}_t^{\mathsf{T}}\mathbf{L}_t/\mathbf{r}_t^{\mathsf{T}}\boldsymbol{\Sigma}_t\mathbf{r}_t\big). && (41)\end{aligned}$$
Thus, by plugging in the expression, one can get:
$$\begin{aligned}\dot{y}_t &= \alpha_t y_t + \beta x_1 + \mathbf{e}_t^{\mathsf{T}}\mathbf{L}_t\boldsymbol{\epsilon}_\perp && (42)\\ &\approx \alpha_t y_t + \beta x_\theta + \mathbf{e}_t^{\mathsf{T}}\mathbf{L}_t\boldsymbol{\epsilon}_\perp && (43)\end{aligned}$$

Where,

$$\alpha_t = \frac{\mathbf{e}_t^\mathsf{T} \mathbf{\Sigma}_t \mathbf{r}_t}{\mathbf{r}_t^\mathsf{T} \mathbf{\Sigma}_t \mathbf{r}_t} - \frac{2\, \hat{\mathbf{b}}_t^T\, {\mathbf{\Sigma}_t}^{-1}\, \boldsymbol{\mu}_t}{{\boldsymbol{\mu}_t}^T\, {\mathbf{\Sigma}_t}^{-1}\, \boldsymbol{\mu}_t} \tag{44}$$

$$\beta_t = \frac{{\boldsymbol{\mu}_t}^T\, {\mathbf{\Sigma}_t}^{-1}\, \hat{\mathbf{b}}_t}{{\boldsymbol{\mu}_t}^T\, {\mathbf{\Sigma}_t}^{-1}\, \boldsymbol{\mu}_t} + \mathbf{e}_t^\mathsf{T} \left[\mathbf{I} - \frac{\mathbf{\Sigma}_t \mathbf{r}_t \mathbf{r}_t^\mathsf{T}}{\mathbf{r}_t^\mathsf{T} \mathbf{\Sigma}_t \mathbf{r}_t} \right] \boldsymbol{\mu}_t \tag{45}$$

$$w_t^{(i)} = (\mathbf{e}_t^\mathsf{T} \mathbf{L}_t)^{(i)} \tag{46}$$

□

C Experiment Details

Here we elaborate more on experiment details.

C.1 EDM and EDM2

For the baselines on EDM and EDM2 codebase, we directly use the code provide in DPM-Solver-v3[36]. For the fair comparision, for all baselines, we controlled $\sigma_{\text{min}} = 0.002$ and $\sigma_{\text{max}} = 80$ as suggested in the original EDM and EDM2 paper.

For DPM-Solver++, we did abalation search over order $\in [1, 2, 3]$, discretization $\in$ $[\text{logSNR},\ \text{time uniform}, , \text{edm}, \text{time quadratic}]$.

For UniPC, we did abalation search over order $\in [1, 2, 3]$, discretization $\in$ $[\text{logSNR},\ \text{time uniform}, , \text{edm}, \text{time quadratic}]$,variant $\in [\text{bh1}, \text{bh2}]$.

For all the ablation results, please see the supplementary material.

C.2 Stable Diffusion 3

For stable diffusion 3, we simply plug in the implementation of all the baselines provided in the Diffuser. We use latest HpsV2.1 to evaluate generate dresults.

D Additional Plots

D.1 General N variable dynamics

This section is not referenced in the main paper and will be removed soon; it is retained only for now to keep the appendix numbering aligned with the main paper.

E Detailed Explanations

E.1 Explicit form of $\mathbf{A}_t$ and $\mathbf{b}_t$

Here we demonstrate the $\mathbf{A}_t$ and $\mathbf{b}_t$ used in AGM[6] and CLD[9]. Here we abuse the notation and inherent the notation from CLD.

Table 1: Comparison of different solvers

Algorithm	$\mathbf{A}_t$	$\mathbf{b}$	F_t
AGM[9]	$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$	$[0, 1]^\mathsf{T}$	$\frac{-4}{t-1} \left(\frac{x1 - \mathbf{x}_t^{(0)}}{1-t} - \mathbf{x}_t^{(1)} \right)$
CLD[9]	$\begin{bmatrix} 0 & -M^{-1} \\ 1 & \Gamma M^{-1} \end{bmatrix} \beta$	$[0, \Gamma\beta]^\mathsf{T}$	$\nabla_{\mathbf{x}_t^{(1)}} \log p(\mathbf{x}, t)$

E.2 Explicit form of mean and convariance matrix

Here we first quickly derive how the F derived which is straight-forward. We know $x_t^{(0)} := x_t$ be the position and define higher derivatives recursively

$$x_t^{(k)} = \frac{d^k x_t^{(0)}}{dt^k}, \quad k = 1, \dots, N-1.$$

The system dynamics form an *Nth–order chain of integrators* driven by a scalar input $F(t, \mathbf{x}_t)$:

$$\begin{aligned} \dot{x}_t^{(0)} &= x_t^{(1)}, \\ \dot{x}_t^{(1)} &= x_t^{(2)}, \\ &\vdots \\ \dot{x}_t^{(N-2)} &= x_t^{(N-1)}, \\ \dot{x}_t^{(N-1)} &= F(t, \mathbf{x}_t). \end{aligned} \tag{47}$$

Equivalently, the position satisfies the scalar ODE

$$\boxed{\frac{d^N x_t^{(0)}}{dt^N} = F(t, \mathbf{x}_t)}.$$

Our goal is starting at some time $t \in [0, 1)$ with known state $\{x_t^{(k)}\}_{k=0}^{N-1}$, choose F so that the position reaches a prescribed value at $t = 1$:

$$x_1^{(0)} = x_1 \quad \text{(“hit the target”)}.$$

Assume F is *held constant* over the *remaining* interval $[t, 1]$. Repeated integration yields the degree-N Taylor polynomial about t:

$$x_1^{(0)} = \sum_{k=0}^{N-1} \frac{(1-t)^k}{k!} x_t^{(k)} + \frac{(1-t)^N}{N!} F. \tag{48}$$

Thus, one can simply solve the F by Rearranging (48) to isolate F:

$$F = \frac{N!}{(1-t)^N} \Big[x_1 - \sum_{k=0}^{N-1} \frac{(1-t)^k}{k!} x_t^{(k)} \Big].$$

Thanks to the simple form of F, one can readily write down the mean and covariance of the system.

Now we know

$$A = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix}, \qquad \mathbf{b}_t = [0, \dots, 0, 1]^\top.$$

By rearraging the dynamics, gives the *linear* time-varying closed loop

$$\dot{\mathbf{x}}_t = \underbrace{\hat{\mathbf{A}}_t}_{A \text{ w/ control}} \mathbf{x}_t + \underbrace{\hat{\mathbf{b}}}_{b \text{ w/ control}} x_1 \tag{49}$$

We need first compute the transition matrix induced by $\hat{\mathbf{A}}_t$ and we call it controlled transition matrix. By Solving $\dot{\Phi} = \hat{\mathbf{A}}_t \Phi$ column-wise gives the polynomial matrix

$$\Phi(t, 0) = \big[T_{k,m}(t)\big]_{k,m=0}^{N-1}, \quad T_{k,m}(t) = \begin{cases} \dfrac{t^{m-k}}{(m-k)!} - \dfrac{N!\, t^{N-k}}{(N-k)!\, m!}, & m \geq k, \\ -\dfrac{N!\, t^{N-k}}{(N-k)!\, m!}, & m < k. \end{cases}$$

Plugging this $\Phi(t,0)$ into the boxed formulas above supplies $\mu(t)$ and $\Sigma(t)$ explicitly for *every* order N.

Let $\mu(t) = \mathbb{E}[\mathbf{x}_t]$. Because $\hat{\mathbf{b}}_t x_1$ is deterministic,

$$\dot{\mu}(t) = \hat{\mathbf{A}}_t\, \mu(t) + \hat{\mathbf{b}}_t\, x_1, \qquad \mu(0) = \mu_0. \tag{50}$$

Define the state–transition matrix $\Phi(t,\tau)$ of $\hat{\mathbf{A}}_t$:

$$\dot{\Phi}(t,\tau) = \hat{\mathbf{A}}_t\, \Phi(t,\tau),\ \Phi(\tau,\tau) = I.$$

Then the standard variation-of-constants formula gives

$$\boldsymbol{\mu}_t = \Phi(t,0)\, \boldsymbol{\mu}_0 + \int_0^t \Phi(t,\tau)\, \hat{\mathbf{b}}_\tau\, x_1\, d\tau.$$

If the initial derivatives are i.i.d. $\mathcal{N}(0,1)$, then $\mu_0 = \mathbf{0}$ and only the integral term remains. Carrying out the integral (polynomials of τ) yields

$$\boxed{\boldsymbol{\mu}_t^{(k)} = \frac{N!\, t^{\,N-k}}{(N-k)!}\, x_1, \qquad k = 0, \ldots, N-1.}$$

Meanwhile, the propagation of covariance matrix is:

$$\dot{\boldsymbol{\Sigma}}_t = \hat{\mathbf{A}}_t \boldsymbol{\Sigma}_t + \boldsymbol{\Sigma}_t \hat{\mathbf{A}}_t^\top, \quad \Sigma(0) = \Sigma_0. \tag{51}$$

Eq.51 is a homogeneous Lyapunov ODE whose unique solution is exactly (see Appendix.G for more details):

$$\boxed{\boldsymbol{\Sigma}_t = \Phi(t,0)\, \boldsymbol{\Sigma}_0\, \Phi(t,0)^\top.}$$

E.3 Previous Fast Solver

Table 2: Comparison of different solvers

	Order type	Order	Multistep type	Expansion term	Discretize space
Heun[17]	Single Step	2	N/A	N/A	σ_t
DEIS[33]	Multi-step	2/3/4	Adams–Bashforth	ϵ_θ	σ_t
DPM-Solver[22]	Multi/Single-step	2/3/4	Adams–Bashforth	ϵ_θ	Optional
DPM-Solver++[23]	Multi/Single-step	2/3/4	Adams–Bashforth	x_θ	Optional
UniPC[34]	Multi-step	3/4/5	Adams–Moulton	x_θ	Optional
TADA(ours)	Multi-step	2/3	Adams–Bashforth	F_θ	t

E.4 Extended Flow Matching

In the framework of flow mathcing, one obtain the velocity by $v_t = \frac{x_1 - x_t}{1-t}$ because it is the linear interpolation between data x_1 and prior x_0. And meanwhile, it happens to be the solution of optimal control problem:

$$\min_{v_t} \int_t^1 \|v_t\|_2^2 \mathrm{d}t, \quad s.t \quad \mathrm{d}x_t = v_t \mathrm{d}t \tag{52}$$

For the detailed derivation, please see Sec.C.1 in [6].

For AGM, they consider a momentum system, which reads

$$\min_{a_t} \int_t^1 \|a_t\|_2^2 \mathrm{d}t, \quad s.t \quad \mathrm{d}x_t = v_t \mathrm{d}t, \quad \mathrm{d}v_t = a_t \mathrm{d}t + \mathrm{d}w_t \tag{53}$$

The differences is that, AGM consider the injection of stochasticity in the velocity channel. For our case, the F_θ derived in Sec.3.2, is the solution for

$$
\begin{aligned}
\min_{F_t} \int_t^1 & \|F_t\|_2^2 \mathrm{d}t \\
\mathrm{d}x_t^{(0)} &= x_t^{(1)}\mathrm{d}t \\
\mathrm{d}x_t^{(1)} &= x_t^{(2)}\mathrm{d}t \\
&\cdots \\
\mathrm{d}x_t^{(N-1)} &= F_t\mathrm{d}t
\end{aligned}
$$

and its spirit keeps same as previous formulation, move $x_t^{(0)}$ to $x_1^0 \sim p_{\text{data}}$ from $t=0$ to $t=1$.

E.5 Degenerate Case of TADA

Here we discuss about the degenerated case of TADA. The reasoning behind it is rather simple. We show the dynamics of y_t (eq.39) again here,

$$
\dot{y}_t = \underbrace{\frac{\hat{\mathbf{b}}_t^T\, \mathbf{\Sigma}_t{}^{-1}}{\boldsymbol{\mu}_t{}^T\, \mathbf{\Sigma}_t{}^{-1}\, \boldsymbol{\mu}_t}}_{\mathbf{e}_t} \mathbf{x}_t + \frac{\boldsymbol{\mu}_t{}^T\, \mathbf{\Sigma}_t{}^{-1}\, \hat{\mathbf{b}}_t}{\boldsymbol{\mu}_t{}^T\, \mathbf{\Sigma}_t{}^{-1}\, \boldsymbol{\mu}_t} x_1 - \frac{2\,\hat{\mathbf{b}}_t^T\, \mathbf{\Sigma}_t{}^{-1}\, \boldsymbol{\mu}_t}{\boldsymbol{\mu}_t{}^T\, \mathbf{\Sigma}_t{}^{-1}\, \boldsymbol{\mu}_t} y_t \tag{54}
$$

and recall that

$$
y_t = \frac{\boldsymbol{\mu}_t{}^\mathsf{T} \mathbf{\Sigma}_t^{-1}}{\boldsymbol{\mu}_t{}^\mathsf{T} \mathbf{\Sigma}_t^{-1} \boldsymbol{\mu}_t} \tag{55}
$$

Thus, if the first term depends exclusively on y_t, the system reduces to the scalar ODE for y_t and becomes formally identical to other diffusion-model parameterizations such as VP, VE, or FM. More precisely, in order to degenerate TADA, one only requires

$$
\boldsymbol{\mu}_t \;\propto\; \hat{\mathbf{b}}_t \tag{56}
$$

where $\hat{\mathbf{b}}_t$ is defined in eq.24. This proportionality holds in two scenarios:

1. When $N=1$, so that $\boldsymbol{\mu}_t$ and $\hat{\mathbf{b}}_t$ are scalars. In that case, the framework collapses to flow matching—a mere reparameterization of the diffusion model.

2. When $\mathbf{A}_t$ is diagonal and its components evolve independently. Then every dimension of $\boldsymbol{\mu}_t$ and $\hat{\mathbf{b}}_t$ shares the same mean and variance, and proportionality follows directly.

A simple empirical check is to propagate the model from different random initializations using our formulation: it yields identical FID scores after generation, confirming the degeneracy.

F Additional Qualitative Comparision

Please see fig.10 and fig.9.

G Solution of the homogeneous Lyapunov ODE

Let $\Phi(t,\tau)$ be the *state–transition matrix* of the (possibly time–varying) coefficient A_t:

$$
\dot{\Phi}(t,\tau) = A_t\,\Phi(t,\tau), \qquad \Phi(\tau,\tau) = I.
$$

Throughout we abbreviate $\Phi(t,0) \equiv \Phi(t)$.

1. Candidate solution. Consider

$$
\Sigma(t) \;=\; \Phi(t)\,\Sigma_0\,\Phi(t)^\top.
$$

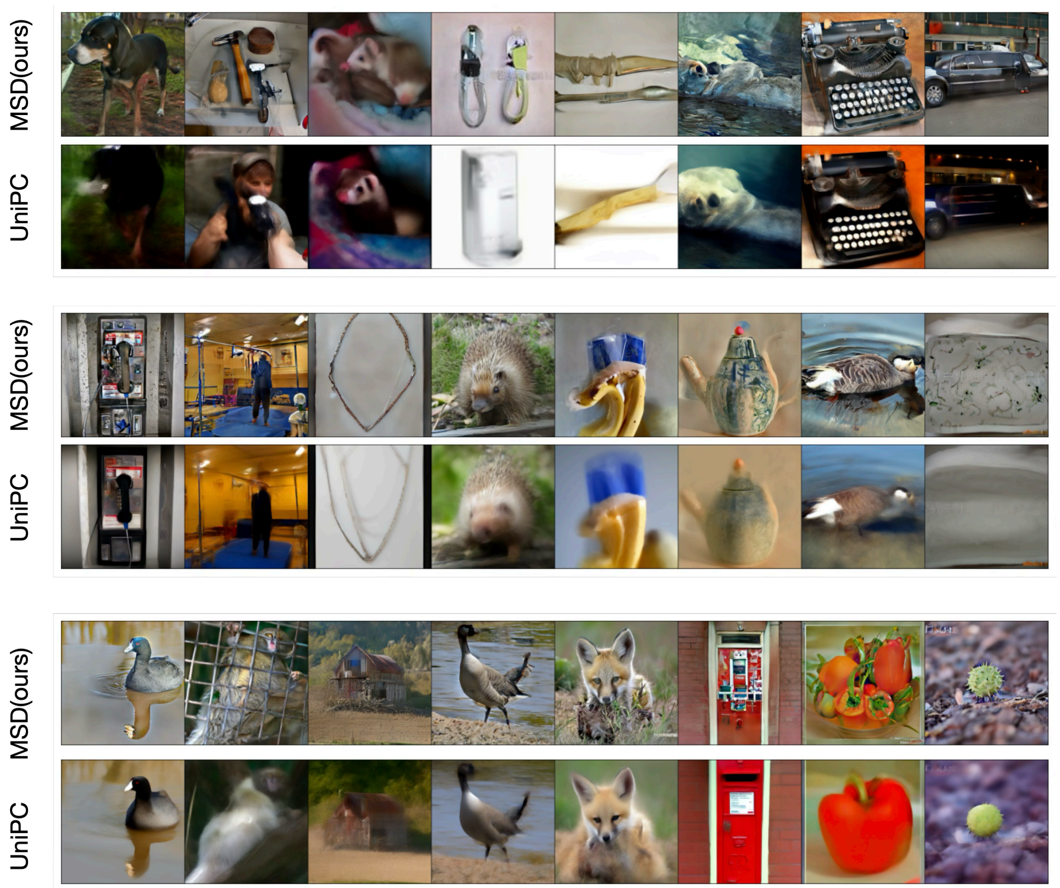

Figure 9: Additional Visual comparision with UniPC using EDM2 w/ 5 NFEs.

2. Verification. Differentiate previous equation and use the product rule together with $\dot{\Phi}(t) = A_t\Phi(t)$ and $\frac{d}{dt}\Phi(t)^\top = \Phi(t)^\top A_t^\top$:

$$\begin{aligned}\dot{\Sigma}(t) &= \dot{\Phi}\,\Sigma_0\Phi^\top + \Phi\,\Sigma_0\dot{\Phi}^\top \\ &= A_t\Phi\Sigma_0\Phi^\top + \Phi\Sigma_0\Phi^\top A_t^\top \\ &= A_t\Sigma(t) + \Sigma(t)A_t^\top.\end{aligned}$$

Hence $\Sigma(t)$ satisfies the differential equation in (Lyap), and $\Sigma(0) = \Phi(0)\Sigma_0\Phi(0)^\top = \Sigma_0$.

3. Uniqueness. Lyapunov Equation is linear in the matrix variable Σ; by the Picard–Lindelöf theorem its solution is unique. Therefore (S) is *the* solution.

H Correlation between two Guassian Variable

Lemma H.1. *Let the random vector*

$$\mathbf{x}_t \sim \mathcal{N}\big(\boldsymbol{\mu}\,x_1,\ \boldsymbol{\Sigma}_t\big), \qquad \boldsymbol{\Sigma}_t = \mathbf{L}_t\mathbf{L}_t^\top \quad (\textit{Cholesky factorisation}).$$

For two fixed column-vectors $\mathbf{r}, \mathbf{e} \in \mathbb{R}^d$ *set*

$$y_t := \mathbf{r}_t^\top\mathbf{x}_t, \quad z_t := \mathbf{e}_t^\top\mathbf{x}_t.$$

Write the convenient abbreviations

$$c_t := \mathbf{r}_t^\top\boldsymbol{\mu}_t\,x_1, \quad d_t := \mathbf{e}_t^\top\boldsymbol{\mu}_t\,x_1, \quad \mathbf{g}_t := \mathbf{L}_t^\top\mathbf{r}_t, \quad \mathbf{h}_t := \mathbf{L}_t^\top\mathbf{e}_t, \quad \sigma_y^2 := \|\mathbf{g}_t\|^2.$$

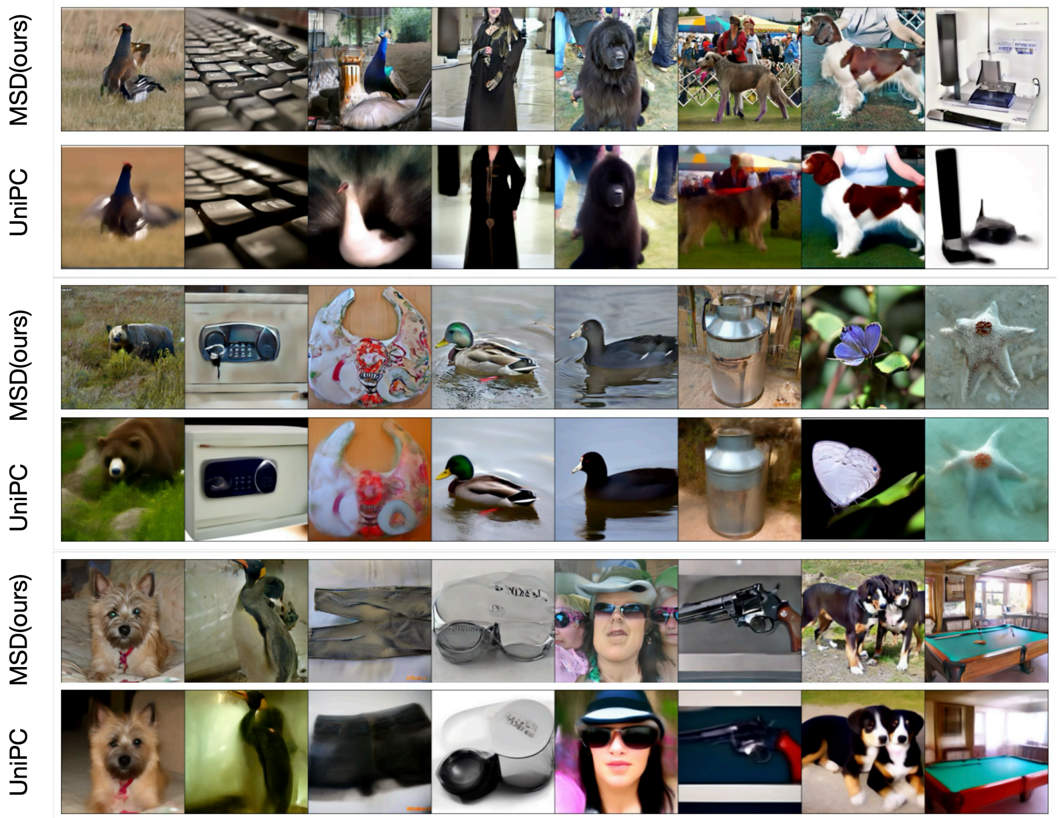

Figure 10: Additional Visual comparision with UniPC using EDM2 w/ 5 NFEs.

Then

$$\boxed{z_t = \mathbf{e}_t^\mathsf{T}\left[\mathbf{I} - \frac{\boldsymbol{\Sigma}_t\mathbf{r}_t\mathbf{r}_t^\mathsf{T}}{\mathbf{r}_t^\mathsf{T}\boldsymbol{\Sigma}_t\mathbf{r}_t}\right]\boldsymbol{\mu}_t x_1 + \frac{\mathbf{e}_t^\mathsf{T}\boldsymbol{\Sigma}_t\mathbf{r}_t}{\mathbf{r}_t^\mathsf{T}\boldsymbol{\Sigma}_t\mathbf{r}_t}y_t + \mathbf{e}_t^\mathsf{T}\mathbf{L}_t\boldsymbol{\epsilon}_\perp, \qquad \boldsymbol{\epsilon}_\perp \sim \mathcal{N}\big(\mathbf{0},\ I_d - \mathbf{L}_t^\mathsf{T}\mathbf{r}_t\mathbf{r}_t^\mathsf{T}\mathbf{L}_t/\mathbf{r}_t^\mathsf{T}\boldsymbol{\Sigma}_t\mathbf{r}_t\big).}$$

Proof.

$$y_t = c_t + \mathbf{g}_t^\mathsf{T}\epsilon, \qquad z_t = d_t + \mathbf{h}_t^\mathsf{T}\epsilon.$$

Any vector can be decomposed into the component along $\mathbf{s}$ and the component orthogonal to $\mathbf{s}$:

$$\boldsymbol{\epsilon} = \frac{\mathbf{g}_t}{\sigma_y^2}\,(\mathbf{g}_t^\mathsf{T}\boldsymbol{\epsilon}) + \boldsymbol{\epsilon}_\perp, \qquad \mathbf{g}_t^\mathsf{T}\boldsymbol{\epsilon}_\perp = 0$$

Because $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, I_d)$ and the projector onto $\mathbf{g}_t$ is orthogonal to the projector onto the complement, $\mathbf{g}_t^\mathsf{T}\boldsymbol{\epsilon}$ and $\boldsymbol{\epsilon}_\perp$ are independent Gaussian variables.

Insert $\mathbf{g}_t^\mathsf{T}\boldsymbol{\epsilon} = y_t - c_t$ to obtain

$$\boldsymbol{\epsilon} = \frac{\mathbf{g}_t}{\sigma_y^2}\,(y_t - c_t) + \boldsymbol{\epsilon}_\perp, \qquad \boldsymbol{\epsilon}_\perp \sim \mathcal{N}\big(\mathbf{0},\ I_d - \mathbf{g}_t\mathbf{g}_t^\mathsf{T}/\sigma_y^2\big).$$

Then we can plug this into z_t:

$$\begin{aligned}
z_t &= d_t + \mathbf{h}_t^\top \left(\frac{\mathbf{g}_t}{\sigma_y^2} (y_t - c_t) + \boldsymbol{\epsilon}_\perp \right) && (57) \\
&= \mathbf{e}_t^\top \boldsymbol{\mu}_t x_1 + \frac{\mathbf{h}_t^\top \mathbf{g}_t}{\sigma_y^2} (y_t - c_t) + \mathbf{h}_t^\top \boldsymbol{\epsilon}_\perp && (58) \\
&= \mathbf{e}_t^\top \boldsymbol{\mu}_t x_1 + \frac{\mathbf{h}_t^\top \mathbf{g}_t}{\sigma_y^2} y_t - \frac{\mathbf{h}_t^\top \mathbf{g}_t \mathbf{r}_t^\top \boldsymbol{\mu}_t}{\sigma_y^2} x_1 + \mathbf{h}_t^\top \boldsymbol{\epsilon}_\perp && (59) \\
&= \mathbf{e}_t^\top \boldsymbol{\mu}_t x_1 + \frac{\mathbf{e}_t^\top \boldsymbol{\Sigma}_t \mathbf{r}_t}{\mathbf{r}_t^\top \boldsymbol{\Sigma}_t \mathbf{r}_t} y_t - \frac{\mathbf{e}_t^\top \boldsymbol{\Sigma}_t \mathbf{r}_t}{\mathbf{r}_t^\top \boldsymbol{\Sigma}_t \mathbf{r}_t} \left(\mathbf{r}_t^\top \boldsymbol{\mu}_t \right) x_1 + \mathbf{e}_t^\top \mathbf{L}_t \boldsymbol{\epsilon}_\perp && (60) \\
&= \mathbf{e}_t^\top \left[\mathbf{I} - \frac{\boldsymbol{\Sigma}_t \mathbf{r}_t \mathbf{r}_t^\top}{\mathbf{r}_t^\top \boldsymbol{\Sigma}_t \mathbf{r}_t} \right] \boldsymbol{\mu}_t x_1 + \frac{\mathbf{e}_t^\top \boldsymbol{\Sigma}_t \mathbf{r}_t}{\mathbf{r}_t^\top \boldsymbol{\Sigma}_t \mathbf{r}_t} y_t + \mathbf{e}_t^\top \mathbf{L}_t \boldsymbol{\epsilon}_\perp && (61)
\end{aligned}$$

□

I General N variable MDM loss

At time t the N-variables are generated by

$$\boxed{\mathbf{x}_t \ = \ \boldsymbol{\mu}_t \, x_1 \ + \ \mathbf{L}_t \, \boldsymbol{\varepsilon}} \qquad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, I_N),$$

with known $\boldsymbol{\mu}_t \in \mathbb{R}^N$ and *invertible* $\mathbf{L}_t \in \mathbb{R}^{N \times N}$. The goal is to predict $\varepsilon^{(N-1)}$.

By whitening trick, one can isolate $\boldsymbol{\epsilon}^{(N-1)}$. Let $\mathbf{e}_{N-1}^\top = [0, \ldots, 0, 1]$ be the row vector that selects the last coordinate. Left-multiplying by $\mathbf{e}_{N-1}^\top \mathbf{L}_t^{-1}$ gives

$$\mathbf{e}_{N-1}^\top \mathbf{L}_t^{-1} \mathbf{x}_t \ = \ \mathbf{e}_{N-1}^\top \mathbf{L}_t^{-1} \boldsymbol{\mu}_t \, x_1 \ + \ \mathbf{e}_{N-1}^\top \underbrace{\mathbf{L}_t^{-1} L_t}_{I_N} \boldsymbol{\varepsilon}.$$

Define the *time-dependent scalars*

$$\mathbf{a}_t^\top := \mathbf{e}_{N-1}^\top \mathbf{L}_t^{-1} \in \mathbb{R}^N, \quad b_t := \mathbf{a}_t^\top \boldsymbol{\mu}_t \neq 0,$$

then

$$\boldsymbol{\epsilon}^{(N-1)} \ = \ \mathbf{a}_t^\top \mathbf{x}_t \ - \ b_t \, x_1. \tag{62}$$

A neural network $\varepsilon_\theta(\mathbf{x}_t, t)$ is trained to approximate $\varepsilon^{(N-1)}$ with the standard

$$\mathcal{L}_{\text{MDM}}(\theta) := \mathbb{E} \| \varepsilon_\theta(\mathbf{x}_t, t) - \boldsymbol{\epsilon}^{(N-1)} \|_2^2$$

Insert eq.62 and multiply the interior by b_t:

$$\begin{aligned}
\mathcal{L}_{\text{MDM}}(\theta) &= \mathbb{E} \| \varepsilon_\theta(\mathbf{x}_t, t) - \mathbf{a}_t^\top \mathbf{x}_t + b_t \, x_1 \|_2^2 \\
&\propto \mathbb{E} \| g_\theta(\mathbf{x}_t, t) - x_1 \|_2^2
\end{aligned}$$

where

$$g_\theta(\mathbf{x}_t, t) := -\frac{\varepsilon_\theta(\mathbf{x}_t, t) - \mathbf{a}_t^\top \mathbf{x}_t}{b_t}$$