Optimal Flow-Matching - Appendix

Raymond Fan

June 10, 2025

1 Introduction

Let $P_i(\vec{x_i}), P_f(\vec{x_f})$ be N dimensional input and output probability distributions (typically noise and training data respectively). In flow-matching, we randomly sample points $\vec{x_i}, \vec{x_f}$ from the input and output distributions, and calculate a sample velocity

$$\vec{v}_s = \vec{x_i}\alpha'(t) + \vec{x_f}\beta'(t) \tag{1}$$

where t is a randomly chosen time $t\varepsilon[0,1]$, and $\alpha(t), \beta(t)$ are scheduling functions with $\alpha(0) = \beta(1) = 1, \alpha(1) = \beta(0) = 0$. Common choices are $\alpha(t) = 1 - t, \beta(t) = t$, a linear schedule.

We train a model $\vec{v}_m(\vec{x}, t)$ to predict the sample velocity $\vec{v}_s$, where $\vec{x} = \vec{x_i}\alpha(t) + \vec{x_f}\beta(t)$. Our model minimizes the MSEloss

$$L = E_{\vec{x_i}, \vec{x_f}, t}\left[\vec{v}_m(\vec{x}, t) - \vec{v}_s)^2\right] \tag{2}$$

where the expected value is taken over the distribution of inputs and outputs $\vec{x_i}, \vec{x_f}$.

The MSEloss is minimized if our model predicts the average sample velocity for a given $\vec{x}, t$,

$$\implies \vec{v}_m(\vec{x}, t) = \langle \vec{v_s} \rangle_{\vec{x}, t} \tag{3}$$

which can be calculated by integrating over the distribution of inputs and outputs $P(\vec{x_i}, \vec{x_f})$ conditioned on points $\vec{x_i}, \vec{x_f}$ contributing to the velocity at point $\vec{x}, t$. This is given by the following integral

$$\langle x_i \rangle_{\vec{x}, t} = \frac{\int_{\vec{x_i}, \vec{x_f}} x_i \, P(\vec{x_i}, \vec{x_f} \mid \vec{x} = \vec{x_i}\alpha(t) + \vec{x_f}\beta(t)) \, dx_i dx_f}{\int_{\vec{x_i}, \vec{x_f}} \, P(\vec{x_i}, \vec{x_f} \mid \vec{x} = \vec{x_i}\alpha(t) + \vec{x_f}\beta(t)) \, dx_i dx_f} \tag{4}$$

where we normalize by the probability that the chosen points actually lie on the curve to obtain the expected value. Letting $C = \vec{x_i}\alpha(t) + \vec{x_f}\beta(t))$, the condition that our points lie on the curve results in a Dirac delta function

$$\langle \vec{x}_i \rangle_{\vec{x}, t} = \frac{\int_{\vec{x_i}, \vec{x_f}} x_i \, P(\vec{x_i}, \vec{x_f})\delta(C - x) \, dx_i dx_f}{\int_{\vec{x_i}, \vec{x_f}} P(\vec{x_i}, \vec{x_f})\delta(C - x) \, dx_i dx_f} \tag{5}$$

2 General Solution

2.1 T = 0

At $t = 0$ the integrals simplify and we can explicitly compute the expectation values

$$\langle x_i \rangle_{\vec{x}, t} = \frac{\int_{\vec{x_i}, \vec{x_f}} \vec{x}_i \, P(\vec{x_i}, \vec{x_f} \mid \vec{x} = \vec{x_i}), dx_i dx_f}{\int_{\vec{x_i}, \vec{x_f}} \, P(\vec{x_i}, \vec{x_f} \mid \vec{x} = \vec{x_i}), dx_i dx_f} = \vec{x} \tag{6}$$

which is true because only $x_i = \vec{x}$ can satisfy $\vec{x}, t = 0$. Similarly

$$\langle \vec{x}_f \rangle_{\vec{x}, t} = \frac{\int_{\vec{x_i}, \vec{x_f}} \vec{x}_f \, P(\vec{x_i}, \vec{x_f} \mid \vec{x} = \vec{x_i}) \, dx_i dx_f}{\int_{\vec{x_i}, \vec{x_f}} \, P(\vec{x_i}, \vec{x_f} \mid \vec{x} = \vec{x_i}) \, dx_i dx_f} = \int_{\vec{x_f}} x_{fk} P(\vec{x_f}) dx_f = \langle x_f \rangle_{P_f} \tag{7}$$

since the constraint on the curve does not depend on x_f, allowing us the factor it out of the equation. Therefore

$$\vec{v}(t=0) = \vec{x}\alpha'(t) + \langle x_f \rangle_{P_f} \beta'(t) \tag{8}$$

For a linear schedule, this has a simple interpretation: at $t = 0$, the velocity field always forgets the initial condition and moves directly towards the average output. As a result, the single step integration of the velocity flow always results in the same output image, given by the average output seen in training.

For $t = 1$, the same procedure results in

$$\vec{v}(t=1) = \langle x_i \rangle_{P_i} \alpha'(t) + \beta'(t)\vec{x} \tag{9}$$

which tells us that at the end, the velocity flow always moves away from the initial distribution average. This is not so useful, but can be tested by evaluating the velocity field at $t = 1$ and reverse integrating to get the mean of the input distribution.

2.2 T != 0

For $t \neq 0, 1$, we can use the relation

$$\vec{x} = \vec{x_i}\alpha(t) + \vec{x_f}\beta(t) \implies \vec{x_f} = \frac{\vec{x} - \vec{x_i}\alpha(t)}{\beta(t)} \tag{10}$$

allowing us to simplify the Dirac delta

$$\langle \vec{x_i} \rangle = \frac{\int_{\vec{x_i}} x_{ik}\, P_i(\vec{x_i}) P_f(\vec{x_f} = \frac{\vec{x} - \vec{x_i}\alpha(t)}{\beta(t)}) dx_i}{\int_{\vec{x_i}} P_i(\vec{x_i}) P_f(\vec{x_f} = \frac{\vec{x} - \vec{x_i}\alpha(t)}{\beta(t)}) dx_i} \tag{11}$$

A similar integral can be obtained for $\langle \vec{x_f} \rangle$, but it is more convenient to use the linearity of the average

$$\langle \vec{x_f} \rangle = \frac{\vec{x}}{\beta(t)} - \frac{\langle x_i \rangle \alpha(t)}{\beta(t)} = \frac{\vec{x}}{\beta(t)} - \frac{\alpha(t)}{\beta(t)} \frac{\int_{\vec{x_i}} \vec{x_i} P_i(\vec{x_i}) P_f(\vec{x_f} = \frac{\vec{x} - \vec{x_i}\alpha(t)}{\beta(t)}) dx_i}{\int_{\vec{x_i}} P_i(\vec{x_i}) P_f(\vec{x_f} = \frac{\vec{x} - \vec{x_i}\alpha(t)}{\beta(t)}) dx_i} \tag{12}$$

Note that we can calculate the velocity with only $\langle \vec{x_i} \rangle$ via

$$\vec{v}(\vec{x}, t) = \alpha'(t) \langle \vec{x_i} \rangle + \frac{\vec{x} - \langle \vec{x_i} \rangle \alpha(t)}{\beta(t)} \beta'(t) = \frac{\vec{x}\beta'(t)}{\beta(t)} + \left(\alpha'(t) - \frac{\alpha(t)}{\beta(t)} \beta'(t) \right) \langle \vec{x_i} \rangle \tag{13}$$

which via direct substitution becomes

$$\vec{v}(\vec{x}, t) = \frac{\vec{x}\beta'(t)}{\beta(t)} + \left(\alpha'(t) - \frac{\alpha(t)}{\beta(t)} \beta'(t) \right) \left(\frac{\int_{\vec{x_i}} \vec{x_i}\, P_i(\vec{x_i}) P_f(\vec{x_f} = \frac{\vec{x} - \vec{x_i}\alpha(t)}{\beta(t)}) dx_i}{\int_{\vec{x_i}} P_i(\vec{x_i}) P_f(\vec{x_f} = \frac{\vec{x} - \vec{x_i}\alpha(t)}{\beta(t)}) dx_i} \right) \tag{14}$$

This form is simplest for explicit calculations due to the singular integral, but the symmetry between $\alpha, \beta, \vec{x_i}, \vec{x_f}$ is more apparent in the form

$$\vec{v}(\vec{x}, t) = \alpha'(t) \frac{\int_{\vec{x_i}} \vec{x_i}\, P_i(\vec{x_i}) P_f\left(\frac{\vec{x} - \vec{x_i}\alpha(t)}{\beta(t)}\right) dx_i}{\int_{\vec{x_i}} P_i(\vec{x_i}) P_f\left(\frac{\vec{x} - \vec{x_i}\alpha(t)}{\beta(t)}\right) dx_i} + \beta'(t) \frac{\int_{\vec{x_f}} \vec{x_f}\, P_f(\vec{x_f}) P_i\left(\frac{\vec{x} - \vec{x_f}\beta(t)}{\alpha(t)}\right) dx_f}{\int_{\vec{x_i}} P_f(\vec{x_f}) P_i\left(\frac{\vec{x} - \vec{x_i}\beta(t)}{\alpha(t)}\right) dx_f} \tag{15}$$

where the left and right integrals correspond to $\langle \vec{x_{ik}} \rangle, \langle \vec{x_{fk}} \rangle$ respectively.

3 Useful Gaussian Integral Properties

Suppose both the initial and final distributions are Gaussian distributions

$$P_i(\vec{x}) = \frac{1}{(2\pi)^{d/2}|\mathbf{\Sigma}_i|^{1/2}} \exp\left(-\frac{1}{2}(\vec{x}-\boldsymbol{\mu}_i)^\top \mathbf{\Sigma}_i (\vec{x}-\boldsymbol{\mu}_i)\right)$$

where Σ_i is the covariance matrix, and μ_i is the average of the initial Gaussian.

Note that if $P_f(\vec{x})$ is Gaussian, then $P_f(\frac{\vec{x}-\vec{x_i}\alpha(t)}{\beta(t)})$ is also a Gaussian distribution but with modified means and covariance matrices

$$\vec{\mu}_{f,t} = -\frac{\vec{\mu_f}\beta(t)}{\alpha(t)} + \frac{\vec{x}}{\alpha(t)} \tag{16}$$

$$\Sigma_{f,t} = \left(\frac{\beta(t)}{\alpha(t)}\right)^2 \Sigma_f \tag{17}$$

and the multiplication of two Gaussian distributions $P_i \times P_f$ is another Gaussian distribution. Denote the final distribution with a linearly transformed input as

$$P_{f,t,\vec{x}}(\vec{x_i}) = P_f\left(\frac{\vec{x}-\vec{x_i}\alpha(t)}{\beta(t)}\right) \tag{18}$$

Then we are interested in the following integrals

$$\left(\frac{\int_{\vec{x_i}} x_i\, P_i(\vec{x_i})P_{f,t,\vec{x}}(\vec{x_i})dx_i}{\int_{\vec{x_i}} P_i(\vec{x_i})P_{f,t,\vec{x}}(\vec{x_i})dx_i}\right), \int_{\vec{x_i}} P_i(\vec{x_i})P_{f,t,\vec{x}}(\vec{x_i})dx_i$$

Our interest in the later integral will become apparent later. The former integral is given by the mean of the Gaussian that results from the product of two Gaussians. Without loss of generality, let the initial distribution be a Gaussian distribution with average $\vec{\mu_i} = 0$, then

$$\langle x_i \rangle = \Sigma_i \left(\Sigma_i + \Sigma_{f,t}\right)^{-1} \vec{\mu}_{f,t} \tag{19}$$

This requires a matrix inverse, which is very computationally expensive in high dimensions.

The integral for norm of two gaussians is given by

$$\int_{R^d} P_1(\mathbf{x}) \cdot P_2(\mathbf{x})\, d\mathbf{x} = \frac{1}{(2\pi)^{d/2}|\mathbf{\Sigma}_1+\mathbf{\Sigma}_2|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)^\top(\mathbf{\Sigma}_1+\mathbf{\Sigma}_2)^{-1}(\boldsymbol{\mu}_1-\boldsymbol{\mu}_2)\right) \tag{20}$$

which is likewise very expensive, due to the high dimensional inverse and determinant.

3.1 Independent Gaussians

We make a simplifying assumption that the distributions are n-dimensional independent Gaussians with identical variance in each of the k components σ_i, σ_f, meaning the covariance matrices become diagonal matrices

$$\Sigma_i = I\sigma_i$$

where I is the identity matrix in n dimensions. Then the average from **??** simplifies

$$\langle x_i \rangle = \frac{\sigma_i}{\sigma_i + \left(\frac{\beta(t)}{\alpha(t)}\right)^2 \sigma_f}\left(-\frac{\vec{\mu}_j\beta(t)}{\alpha(t)} + \frac{\vec{x}}{\alpha(t)}\right) \tag{21}$$

as well as the norm

$$\int_{\vec{x_i}} P_i(\vec{x_i})P_{f,t,\vec{x}}(\vec{x_i})dx_i = \frac{1}{(2\pi)^{d/2}} \frac{1}{\left(\sigma_i + \left(\frac{\beta(t)}{\alpha(t)}\right)^2\sigma_f\right)^{d/2}} \exp\left(-\frac{1}{2}\frac{\left(\frac{\vec{\mu}_j\beta(t)}{\alpha(t)} - \frac{\vec{x}}{\alpha(t)}\right)^2}{\left(\sigma_i + \left(\frac{\beta(t)}{\alpha(t)}\right)^2\sigma_f\right)}\right)$$

This approximation is valid when considering the typical use cases in flow-matching, where we can treat our training data as a sum of Gaussian distributions with very small variances $\sigma_f \to 0$.

3.2 Additivity

Suppose we've solved the vector flow for two different output distributions P_{f1}, P_{f2}, obtaining in the process $\langle \vec{x}_i \rangle_1, \langle \vec{x}_i \rangle_2$, and now want to obtain the vector flow corresponding to $P_f = P_{f1}C_1 + P_{f2}(1-C_1)$. This amounts to computing the new $\langle \vec{x}_i \rangle$ for the new probability distribution, given by

$$\begin{aligned}\langle \vec{x}_i \rangle &= \frac{\int_{\vec{x_i}} x_i \, P_i(\vec{x_i}) P_f(\vec{x_f} = \frac{\vec{x} - \vec{x_i}\alpha(t)}{\beta(t)}) dx_i}{\int_{\vec{x_i}} P_i(\vec{x_i}) P_f(\vec{x_f} = \frac{\vec{x} - \vec{x_i}\alpha(t)}{\beta(t)}) dx_i} \\ &= \frac{\int_{\vec{x_i}} x_i \, P_i(\vec{x_i}) \left[C_1 P_{f1}(\vec{x_f} = \frac{\vec{x} - \vec{x_i}\alpha(t)}{\beta(t)}) + (1 - C_1) P_{f2}(\vec{x_f} = \frac{\vec{x} - \vec{x_i}\alpha(t)}{\beta(t)}) \right] dx_i}{\int_{\vec{x_i}} P_i(\vec{x_i}) \left[C_1 P_{f1}(\vec{x_f} = \frac{\vec{x} - \vec{x_i}\alpha(t)}{\beta(t)}) + (1 - C_1) P_{f2}(\vec{x_f} = \frac{\vec{x} - \vec{x_i}\alpha(t)}{\beta(t)}) \right] dx_i}\end{aligned} \tag{22}$$

Denoting the old normalization factors $M_j = \int_{\vec{x_i}} P_i(\vec{x_i}) P_{fj}\big(\frac{\vec{x} - \vec{x_i}\alpha(t)}{\beta(t)}\big)$ we can multiply and divide by the old normalization factors to obtain

$$\langle \vec{x}_i \rangle = \sum_j \left(\frac{\int_{\vec{x_i}} x_i \, P_i(\vec{x_i}) P_{f,j}\big(\frac{\vec{x} - \vec{x_i}\alpha(t)}{\beta(t)}\big) dx_i}{\int_{\vec{x_i}} P_i(\vec{x_i}) P_{f,j}\big(\frac{\vec{x} - \vec{x_i}\alpha(t)}{\beta(t)}\big) dx_i} \times \frac{C_j M_j}{\sum_r C_r M_r} \right) \tag{23}$$

where we see that our integral consists of a sum of the old velocity integrals weighted by new weights $W_l = C_l M_l / \sum_r^J C_r M_r$ (which appropriately add up to 1). This thus implies that the new average x_i is a weighted sum of old x_i

$$\langle \vec{x}_i \rangle = \sum_j \langle \vec{x}_i \rangle_j W_j \tag{24}$$

and similarly the new velocity field is a weighted sum of old velocity fields

$$\vec{v}(\vec{x}, t) = \sum_j \vec{v}_j(\vec{x}, t) W_j \tag{25}$$

which, since the prefactors are independent of specific distribution, becomes

$$\vec{v}(\vec{x}, t) = \frac{\vec{x}\beta'(t)}{\beta(t)} + \left(\alpha'(t) - \frac{\alpha(t)}{\beta(t)} \beta'(t) \right) \left(\sum_j \langle \vec{x}_i \rangle_j W_j \right) \tag{26}$$

4 Independent Gaussian Formula

For flow-matching a single independent Gaussian to a superposition of multiple independent Gaussians, the velocity field depends on

$$\sum_j \langle \vec{x}_i \rangle_j W_j \tag{27}$$

Recall that weights are proportional to the norms computed earlier. Furthermore, our target distribution's Gaussians are all identical in variance, only differing by the specific mean, so the weights can be simplified

$$\begin{aligned} W_j &\propto \frac{C_j}{(2\pi)^{d/2}} \frac{1}{\left(\sigma_i + \left(\frac{\beta(t)}{\alpha(t)}\right)^2 \sigma_f\right)^{d/2}} \exp\left(-\frac{1}{2} \frac{\left(\frac{\vec{\mu}_j \beta(t)}{\alpha(t)} - \frac{\vec{x}}{\alpha(t)}\right)^2}{\left(\sigma_i + \left(\frac{\beta(t)}{\alpha(t)}\right)^2 \sigma_f\right)}\right) \\ &\to C_j \exp\left(-\frac{1}{2} \frac{\left(\frac{\vec{\mu}_j \beta(t)}{\alpha(t)} - \frac{\vec{x}}{\alpha(t)}\right)^2}{\left(\sigma_i + \left(\frac{\beta(t)}{\alpha(t)}\right)^2 \sigma_f\right)}\right) \end{aligned} \tag{28}$$

where the exponent term survives due to the presence of $\vec{\mu}_j$. Our quantity of interest is thus

$$\sum_j \langle \vec{x}_i \rangle_j W_j = \sum_j \frac{C_j \exp\left(-\frac{1}{2} \frac{\left(\frac{\vec{\mu}_j \beta(t)}{\alpha(t)} - \frac{\vec{x}}{\alpha(t)}\right)^2}{\left(\sigma_i + \left(\frac{\beta(t)}{\alpha(t)}\right)^2 \sigma_f\right)}\right)}{\sum_k C_k \exp\left(-\frac{1}{2} \frac{\left(\frac{\vec{\mu}_k \beta(t)}{\alpha(t)} - \frac{\vec{x}}{\alpha(t)}\right)^2}{\left(\sigma_i + \left(\frac{\beta(t)}{\alpha(t)}\right)^2 \sigma_f\right)}\right)} \times \frac{\sigma_i}{\sigma_i + \left(\frac{\alpha(t)}{\beta(t)}\right)^2 \sigma_f} \left(-\frac{\vec{\mu}_j \beta(t)}{\alpha(t)} + \frac{\vec{x}}{\alpha(t)}\right) \tag{29}$$

We drop (t) from $\alpha(t), \beta(t)$ for simplicity. Multiplying by α/α and using $\alpha^2 \left(\sigma_i + \left(\frac{\beta(t)}{\alpha(t)}\right)^2 \sigma_f\right) = \alpha^2 \sigma_i + \beta^2 \sigma_f$ yields

$$\sum_j \langle \vec{x}_i \rangle_j W_j = \sum_j \left(\frac{C_j \exp\left(-\frac{1}{2} \frac{(\vec{x} - \vec{\mu}_j \beta)^2}{(\alpha^2 \sigma_i + \beta^2 \sigma_f)}\right)}{\sum_k C_k \exp\left(-\frac{1}{2} \frac{(\vec{x} - \vec{\mu}_k \beta)^2}{(\alpha^2 \sigma_i + \beta^2 \sigma_f)}\right)} \times \frac{\alpha \sigma_i}{(\alpha^2 \sigma_i + \beta^2 \sigma_f)} (\vec{x} - \vec{\mu}_j \beta) \right) \tag{30}$$

This can be simplified using the following notation where we sum over vectors $\vec{z}_j$

$$\text{Softmax}_{\vec{z}_j}(f(\vec{z}_j)) = \frac{e^{f(z_j)}}{\sum_k e^{f(z_k)}} \tag{31}$$

which results in

$$\frac{C_j \exp\left(-\frac{1}{2} \frac{(\vec{x} - \vec{\mu}_j \beta)^2}{(\alpha^2 \sigma_i + \beta^2 \sigma_f)}\right)}{\sum_k C_k \exp\left(-\frac{1}{2} \frac{(\vec{x} - \vec{\mu}_k \beta)^2}{(\alpha^2 \sigma_i + \beta^2 \sigma_f)}\right)} = \text{Softmax}_{\vec{\mu}_j} \left(-\frac{1}{2} \frac{(\vec{x} - \vec{\mu}_j \beta)^2}{(\alpha^2 \sigma_i + \beta^2 \sigma_f)} + \log C_j\right) \tag{32}$$

This yields the final equation for the velocity flow

$$\vec{v}(\vec{x}, t) = \frac{\vec{x} \beta'}{\beta} + \left(\alpha' - \alpha \frac{\beta'}{\beta}\right) \sum_j \frac{\alpha \sigma_i}{(\alpha^2 \sigma_i + \beta^2 \sigma_f)} (\vec{x} - \vec{\mu}_j \beta) \times \text{Softmax}_{\vec{\mu}_j} \left(-\frac{1}{2} \frac{(\vec{x} - \vec{\mu}_j \beta)^2}{(\alpha^2 \sigma_i + \beta^2 \sigma_f)} + \log C_j\right) \tag{33}$$

We can move some terms outside of the summation since they're independent of j

$$\vec{v}(\vec{x},t) = \frac{\vec{x}\beta'}{\beta} + \left(\alpha' - \alpha\frac{\beta'}{\beta}\right)\left(\frac{\alpha\sigma_i}{\alpha^2\sigma_i + \beta^2\sigma_f}\right)\left(\vec{x} - \sum_j \vec{\mu}_j\beta \times \text{Softmax}_{\vec{\mu}_j}\left(-\frac{1}{2}\frac{(\vec{x}-\vec{\mu}_j\beta)^2}{(\alpha^2\sigma_i + \beta^2\sigma_f)} + \log C_j\right)\right) \tag{34}$$

Expanding out the factors allows us to simplify to the following

$$\vec{v}(\vec{x},t) = \left(\vec{x}\left(\alpha'\alpha\sigma_i + \beta'\beta\sigma_f\right) - \alpha\sigma_i\left(\alpha'\beta - \alpha\beta'\right)\left(\sum_j \vec{\mu}_j \times \text{Softmax}_{\vec{\mu}_j}\left(-\frac{1}{2}\frac{(\vec{x}-\vec{\mu}_j\beta)^2}{(\alpha^2\sigma_i + \beta^2\sigma_f)} + \log C_j\right)\right)\right)$$
$$\times\left(\frac{1}{\alpha^2\sigma_i + \beta^2\sigma_f}\right) \tag{35}$$

To simplify further, we require specific assumptions on α, β, σ_f.

4.1 Independent Gaussians With Linear Schedule

For a linear schedule $\alpha = 1-t, \beta = t$, we have

$$\begin{aligned}\vec{v}(\vec{x},t) = &\left(\frac{(t-1)\sigma_i + t\sigma_f}{(1-t)^2\sigma_i + t^2\sigma_f}\right)\vec{x} + \left(\frac{(1-t)\sigma_i}{(1-t)^2\sigma_i + t^2\sigma_f}\right)\\ &\times \sum_j \vec{\mu}_j \times \text{Softmax}_{\vec{\mu}_j}\left(-\frac{1}{2}\frac{(\vec{x}-\vec{\mu}_j t)^2}{((1-t)^2\sigma_i + t^2\sigma_f)} + \log C_j\right)\end{aligned} \tag{36}$$

4.2 Independent Gaussians With Negligible Output Variances

In the limit $\sigma_f \to 0$, which corresponds to the typical flow-matching problem where we have a discrete set of data samples, the expression simplifies

$$\vec{v}(\vec{x},t) = \frac{\alpha'}{\alpha}\vec{x} + \frac{\alpha\beta' + \alpha'\beta}{\alpha}\sum_j \vec{\mu}_j \times \text{Softmax}_{\vec{\mu}_j}\left(-\frac{1}{2}\frac{(\vec{x}-\vec{\mu}_j\beta)^2}{\alpha^2\sigma_i} + \log C_j\right) \tag{37}$$

If we additionally have a linear schedule, then we obtain

$$\vec{v}(\vec{x},t) = \frac{1}{1-t}\left(\sum_j \vec{\mu}_j \times \text{Softmax}_{\vec{\mu}_j}\left(-\frac{1}{2}\frac{(\vec{x}-\vec{\mu}_j t)^2}{(1-t)^2\sigma_i} + \log C_j\right) - \vec{x}\right) \tag{38}$$

This final formula has a simple interpretation that the model's predicted velocity takes it straight towards some estimated output position, obtained from the softmax over our data distribution.

4.3 Non-uniform Variances

The previous results assumed flowing towards independent Gaussians with nonzero but equal variances, enabling several terms to simplify.

Here, we consider the most relevant case for using the analytic approach on real data, where real data has been clustered into various (approximately) independent Gaussians of differing variance.

$$W_j \propto \frac{C_j}{\left(\sigma_i + \left(\frac{\beta(t)}{\alpha(t)}\right)^2 \sigma_{f,j}\right)^{d/2}} \exp\left(-\frac{1}{2}\frac{\left(\frac{\vec{\mu}_j \beta(t)}{\alpha(t)} - \frac{\vec{x}}{\alpha(t)}\right)^2}{\left(\sigma_i + \left(\frac{\beta(t)}{\alpha(t)}\right)^2 \sigma_{f,j}\right)}\right) \tag{39}$$

where $\sigma_{f,j}$ denotes the variance of the jth data point, and W_j must be normalized. This equation follows from applying Gaussian integrals to Eq. 23. Using $\alpha^2\left(\sigma_i + \left(\frac{\beta(t)}{\alpha(t)}\right)^2 \sigma_f\right) = \alpha^2\sigma_i + \beta^2\sigma_f$ and leaving t dependence implicit for notational simplicity yields

$$W_j \propto \frac{\alpha C_j}{\left(\alpha^2\sigma_i + \beta^2\sigma_{f,j}\right)^{d/2}} \exp\left(-\frac{1}{2}\frac{\left(\vec{\mu}_j\beta(t) - \vec{x}\right)^2}{\left(\alpha^2\sigma_i + \beta^2\sigma_{f,j}\right)}\right) \tag{40}$$

Our quantity of interest is thus

$$\sum_j \langle \vec{x}_i \rangle_j W_j = \sum_j \frac{\frac{\alpha C_j}{(\alpha^2\sigma_i+\beta^2\sigma_{f,j})^{d/2}} \exp\left(-\frac{1}{2}\frac{(\vec{\mu}_j\beta(t)-\vec{x})^2}{(\alpha^2\sigma_i+\beta^2\sigma_{f,j})}\right)}{\sum_k \frac{\alpha C_k}{(\alpha^2\sigma_i+\beta^2\sigma_{f,k})^{d/2}} \exp\left(-\frac{1}{2}\frac{(\vec{\mu}_k\beta(t)-\vec{x})^2}{(\alpha^2\sigma_i+\beta^2\sigma_{f,k})}\right)} \times \frac{\alpha\sigma_i}{\alpha^2\sigma_i + \beta^2\sigma_{f,j}} \left(\vec{x} - \vec{\mu}_j\beta\right) \tag{41}$$

Letting $C_{j,t} = \frac{\alpha C_j}{(\alpha^2\sigma_i+\beta^2\sigma_{f,j})^{d/2}}$, we use the following identity

$$\frac{\alpha C_j}{\left(\alpha^2\sigma_i + \beta^2\sigma_{f,j}\right)^{d/2}} \exp\left(-\frac{1}{2}\frac{\left(\vec{\mu}_j\beta(t) - \vec{x}\right)^2}{\left(\alpha^2\sigma_i + \beta^2\sigma_{f,j}\right)}\right) = \exp\left(-\frac{1}{2}\frac{\left(\vec{\mu}_j\beta(t) - \vec{x}\right)^2}{\left(\alpha^2\sigma_i + \beta^2\sigma_{f,j}\right)} + \log C_{j,t}\right) \tag{42}$$

giving

$$\sum_j \langle \vec{x}_i \rangle_j W_j = \sum_j \frac{\alpha\sigma_i}{\alpha^2\sigma_i + \beta^2\sigma_{f,j}} \left(\vec{x} - \vec{\mu}_j\beta\right) \times \mathrm{Softmax}_j\left(-\frac{1}{2}\frac{\left(\vec{\mu}_j\beta(t) - \vec{x}\right)^2}{\left(\alpha^2\sigma_i + \beta^2\sigma_{f,j}\right)} + \log C_{j,t}\right) \tag{43}$$

where the softmax no longer only depends on $\vec{\mu}_j$, but instead all quantities with subscript j.

Previously only the softmax and $\vec{\mu}_j$ depended on index j, allowing us to factor $\vec{x}$ and various coefficients out from the summation.

This is no longer the case. Although the summation over j and the softmax add to 1, the variance dependent weights do not allow us to factor. We have

$$\begin{aligned} \vec{v}(\vec{x}, t) = &\left(\sum_j \frac{\alpha\sigma_i}{\alpha^2\sigma_i + \beta^2\sigma_{f,j}} \left(\vec{x} - \vec{\mu}_j\beta\right) \times \mathrm{Softmax}_j\left(-\frac{1}{2}\frac{\left(\vec{\mu}_j\beta(t) - \vec{x}\right)^2}{\left(\alpha^2\sigma_i + \beta^2\sigma_{f,j}\right)} + \log C_{j,t}\right)\right) \\ &\times \left(\alpha' - \frac{\alpha}{\beta}\beta'\right) + \frac{\vec{x}\beta'}{\beta} \end{aligned} \tag{44}$$

Terms can be pulled inside since the summation over the softmax adds to one,

$$
\begin{aligned}
&\sum_j \left(\frac{\alpha\sigma_i}{\alpha^2\sigma_i + \beta^2\sigma_{f,j}} \left(\alpha' - \frac{\alpha}{\beta}\beta' \right) (\vec{x} - \vec{\mu}_j\beta) + \frac{\vec{x}\beta'}{\beta} \right) \times \text{Softmax}_j\,(...) \\
&= \sum_j \left(\vec{x} \left(\frac{\alpha\alpha'\sigma_i}{\alpha^2\sigma_i + \beta^2\sigma_{f,j}} - \frac{\beta'}{\beta}\frac{\alpha^2\sigma_i}{\alpha^2\sigma_i + \beta^2\sigma_{f,j}} + \frac{\beta'}{\beta} \right) - \frac{\alpha\sigma_i\,(\alpha'\beta - \alpha\beta')}{\alpha^2\sigma_i + \beta^2\sigma_{f,j}}\vec{\mu}_j \right) \times \text{Softmax}_j\,(...) \\
&= \sum_j \left(\vec{x} \left(\frac{\alpha\alpha'\sigma_i + \beta\beta'\sigma_{f,j}}{\alpha^2\sigma_i + \beta^2\sigma_{f,j}} \right) - \frac{\alpha\sigma_i\,(\alpha'\beta - \alpha\beta')}{\alpha^2\sigma_i + \beta^2\sigma_{f,j}}\vec{\mu}_j \right) \times \text{Softmax}_j\,(...)\,,
\end{aligned} \tag{45}
$$

yielding the final result

$$
\vec{v}(\vec{x},t) = \sum_j \left(\vec{x} \left(\frac{\alpha\alpha'\sigma_i + \beta\beta'\sigma_{f,j}}{\alpha^2\sigma_i + \beta^2\sigma_{f,j}} \right) - \frac{\alpha\sigma_i\,(\alpha'\beta - \alpha\beta')}{\alpha^2\sigma_i + \beta^2\sigma_{f,j}}\vec{\mu}_j \right) \times \text{Softmax}_j \left(-\frac{1}{2}\frac{\left(\vec{\mu}_j\beta(t) - \vec{x}\right)^2}{(\alpha^2\sigma_i + \beta^2\sigma_{f,j})} + \log C_{j,t} \right) \tag{46}
$$

4.4 Non-uniform Variances with Linear Schedule

For a linear schedule $\alpha = 1 - t, \beta = t$

$$
\begin{aligned}
\vec{v}(\vec{x},t) = &\sum_j \left(\vec{x} \left(\frac{(t-1)\sigma_i + t\sigma_{f,j}}{(1-t)^2\sigma_i + t^2\sigma_{f,j}} \right) - \frac{(1-t)\sigma_i\,(-t-(1-t))}{(1-t)^2\sigma_i + t^2\sigma_{f,j}}\vec{\mu}_j \right) \\
&\times \text{Softmax}_j \left(-\frac{1}{2}\frac{\left(\vec{\mu}_j t - \vec{x}\right)^2}{((1-t)^2\sigma_i + t^2\sigma_{f,j})} + \log\left(\frac{(1-t)C_j}{\left((1-t)^2\sigma_i + t^2\sigma_{f,j}\right)^{d/2}} \right) \right)
\end{aligned} \tag{47}
$$

Simplifying

$$
\begin{aligned}
\vec{v}(\vec{x},t) = &\sum_j \left(\frac{\vec{x}\,(t\sigma_{f,j} - (1-t)\sigma_i) + (1-t)\sigma_i\vec{\mu}_j}{(1-t)^2\sigma_i + t^2\sigma_{f,j}} \right) \\
&\times \text{Softmax}_j \left(-\frac{1}{2}\frac{\left(\vec{\mu}_j t - \vec{x}\right)^2}{((1-t)^2\sigma_i + t^2\sigma_{f,j})} + \log\left(\frac{(1-t)C_j}{\left((1-t)^2\sigma_i + t^2\sigma_{f,j}\right)^{d/2}} \right) \right)
\end{aligned} \tag{48}
$$

When all $\sigma_{f,j} = 0$, this equation simplifies into Eq. 38. To efficiently compute this quantity, it is helpful to define $D_j = \left((1-t)^2\sigma_i + t^2\sigma_{f,j}\right)$, yielding

$$
\vec{v}(\vec{x},t) = \sum_j \left(\frac{\vec{x}\,(t\sigma_{f,j} - (1-t)\sigma_i) + (1-t)\sigma_i\vec{\mu}_j}{D_j} \right) \times \text{Softmax}_j \left(-\frac{1}{2}\frac{\left(\vec{x} - \vec{\mu}_j t\right)^2}{D_j} + \log\left((1-t)C_j\right) - \frac{d}{2}\log D_j \right) \tag{49}
$$