Foundations of Deep Learning

ALF

Alfredo Canziani

@alfcnz

Inference for latent variable Energy Based Models (EBMs)

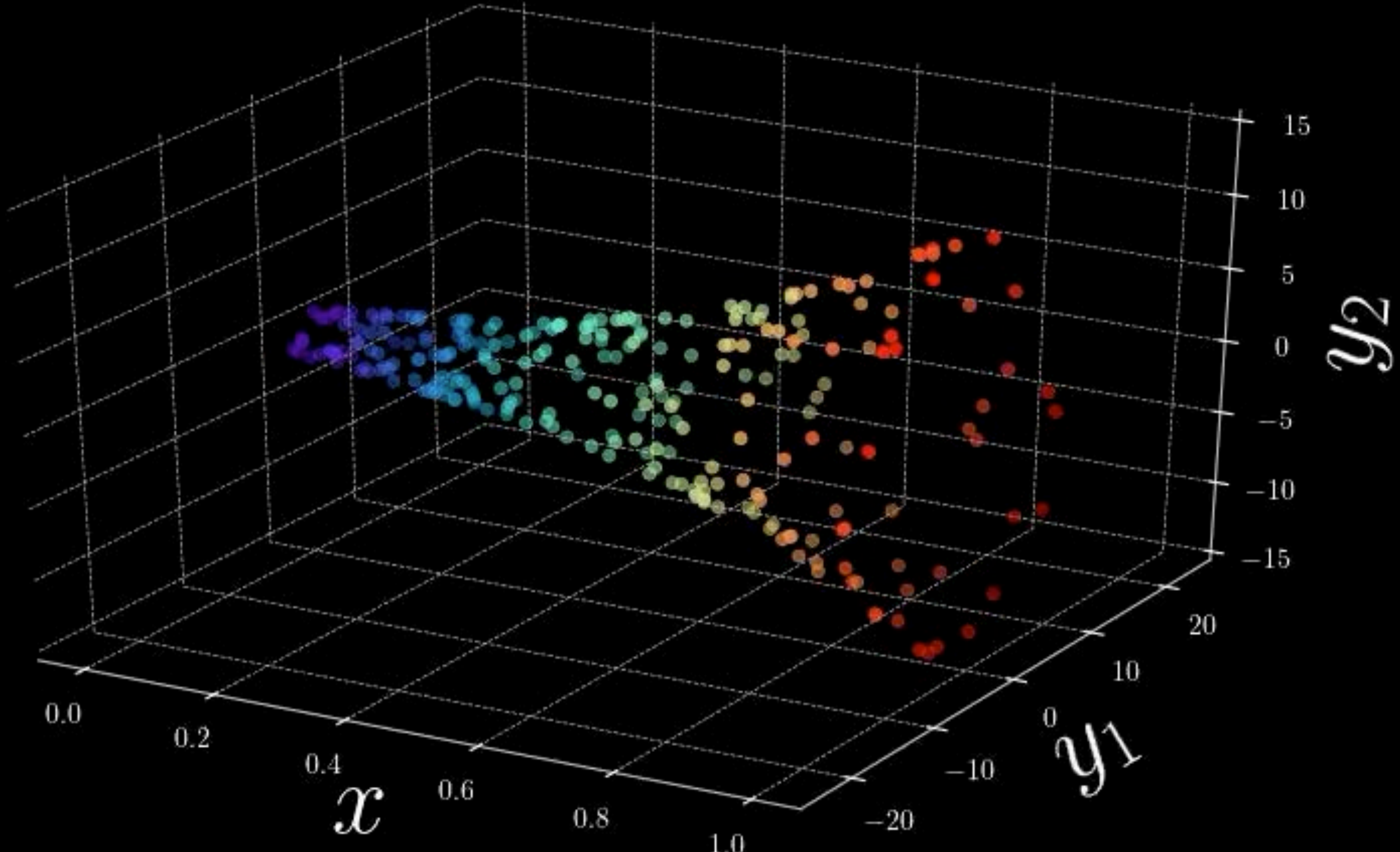
Toy example, the ellipse

INFERENCE
Un-supervised learning

Un-conditional case

Training samples

$$\alpha = 1.5$$
$$\beta = 2$$



$$\boldsymbol{y} = \begin{bmatrix} \rho_1(x)\cos(\theta) + \varepsilon \\ \rho_2(x)\sin(\theta) + \varepsilon \end{bmatrix}$$

$$\rho : \mathbb{R} \to \mathbb{R}^2$$

$$x \mapsto \begin{bmatrix} \alpha x + \beta(1-x) \\ \beta x + \alpha(1-x) \end{bmatrix} \cdot \exp(2x)$$

$$x \sim \mathcal{U}(0,1)$$

$$\theta \sim \mathcal{U}(0,2\pi)$$

$$\varepsilon \sim \mathcal{N}\left[0, \left(\tfrac{1}{20}\right)^2\right]$$

Training samples

$$\alpha = 1.5$$
$$\beta = 2$$

$$\boldsymbol{y} = \begin{bmatrix} \cancel{\rho_1(x)}^{\,2} \cos(\theta) + \varepsilon \\ \cancel{\rho_2(x)}_{\,1.5} \sin(\theta) + \varepsilon \end{bmatrix}$$

$$\boldsymbol{Y} = [\boldsymbol{y}^{(1)}, \ldots, \boldsymbol{y}^{(24)}]$$

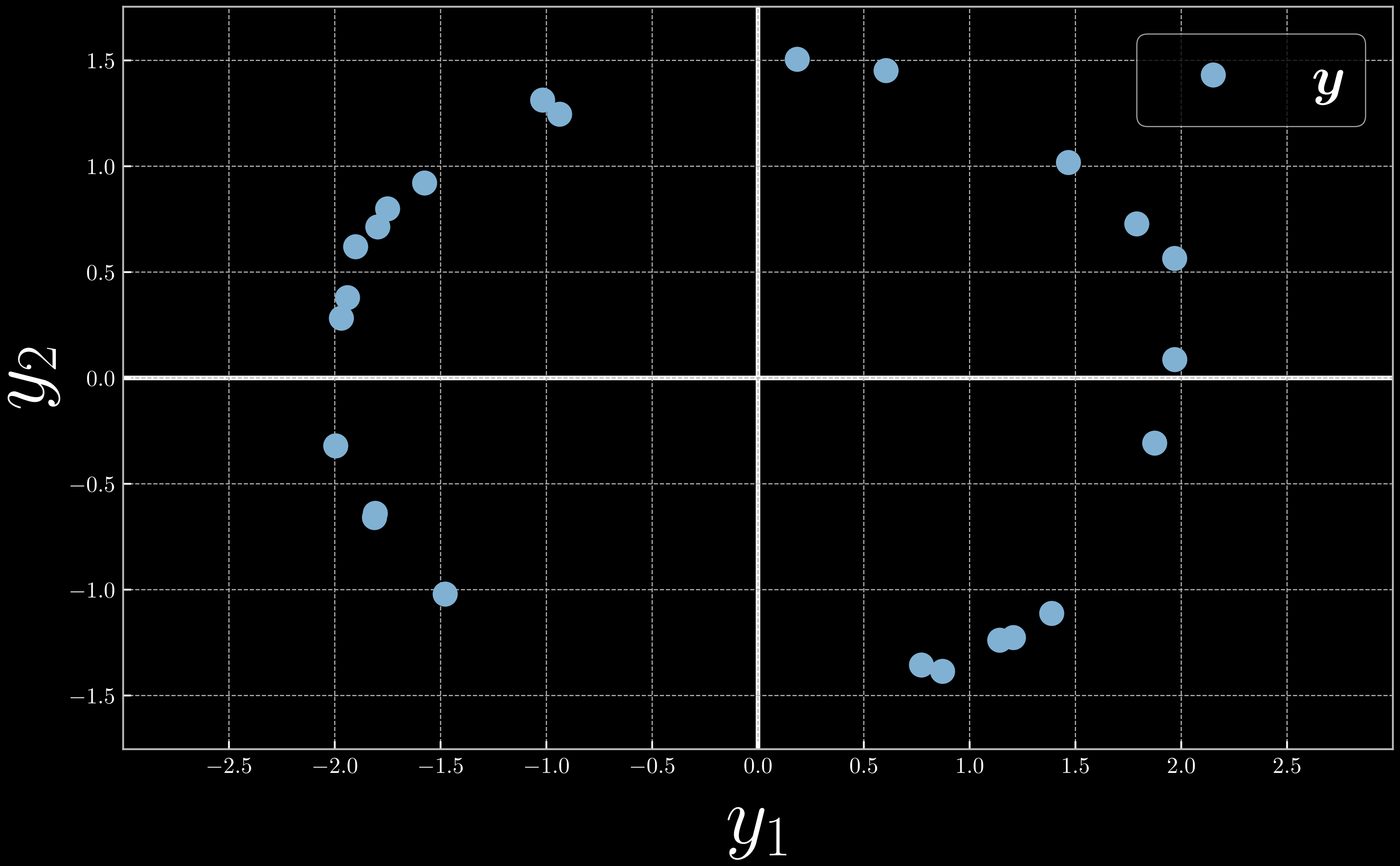
$$\rho : \mathbb{R} \to \mathbb{R}^2$$

$$x \mapsto \begin{bmatrix} \alpha x + \beta(1-x) \\ \beta x + \alpha(1-x) \end{bmatrix} \cdot \exp(2x)$$

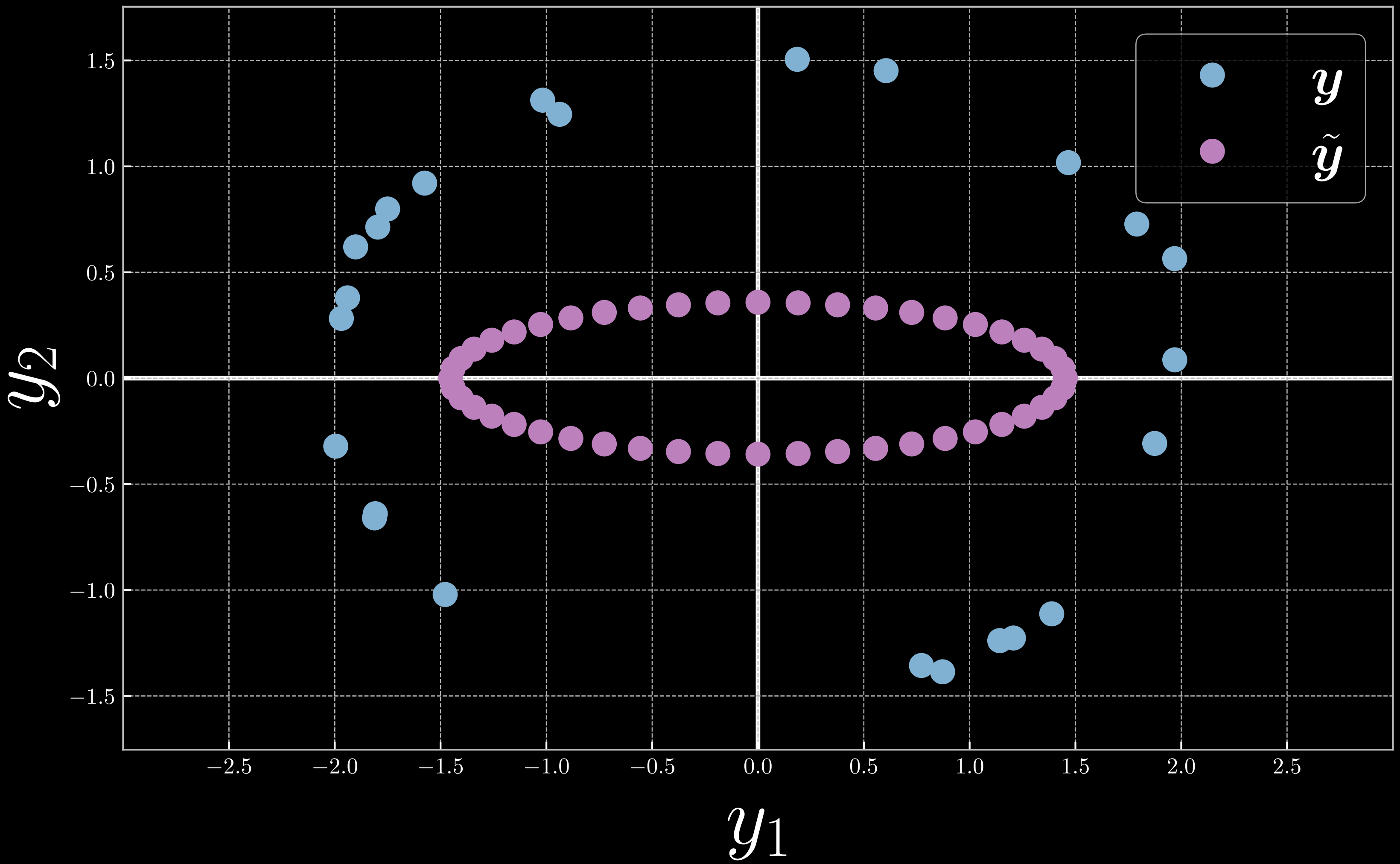
$$x \sim \cancel{\mathcal{U}(0,1)}\ 0$$

$$\theta \sim \mathcal{U}(0, 2\pi)$$

$$\varepsilon \sim \mathcal{N}\left[0, \left(\tfrac{1}{20}\right)^2\right]$$

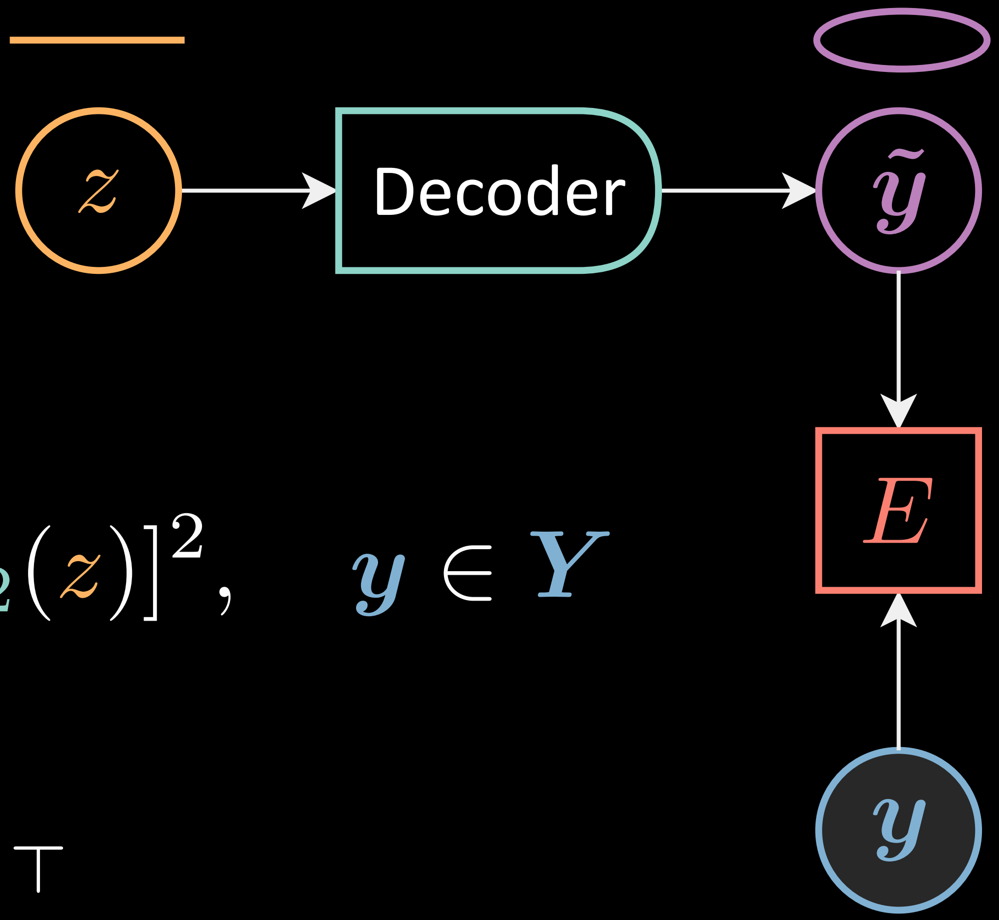


Untrained model manifold



$$z = \left[0 : \frac{\pi}{24} : 2\pi\right[$$

Energy function $E(\boldsymbol{y}, z)$

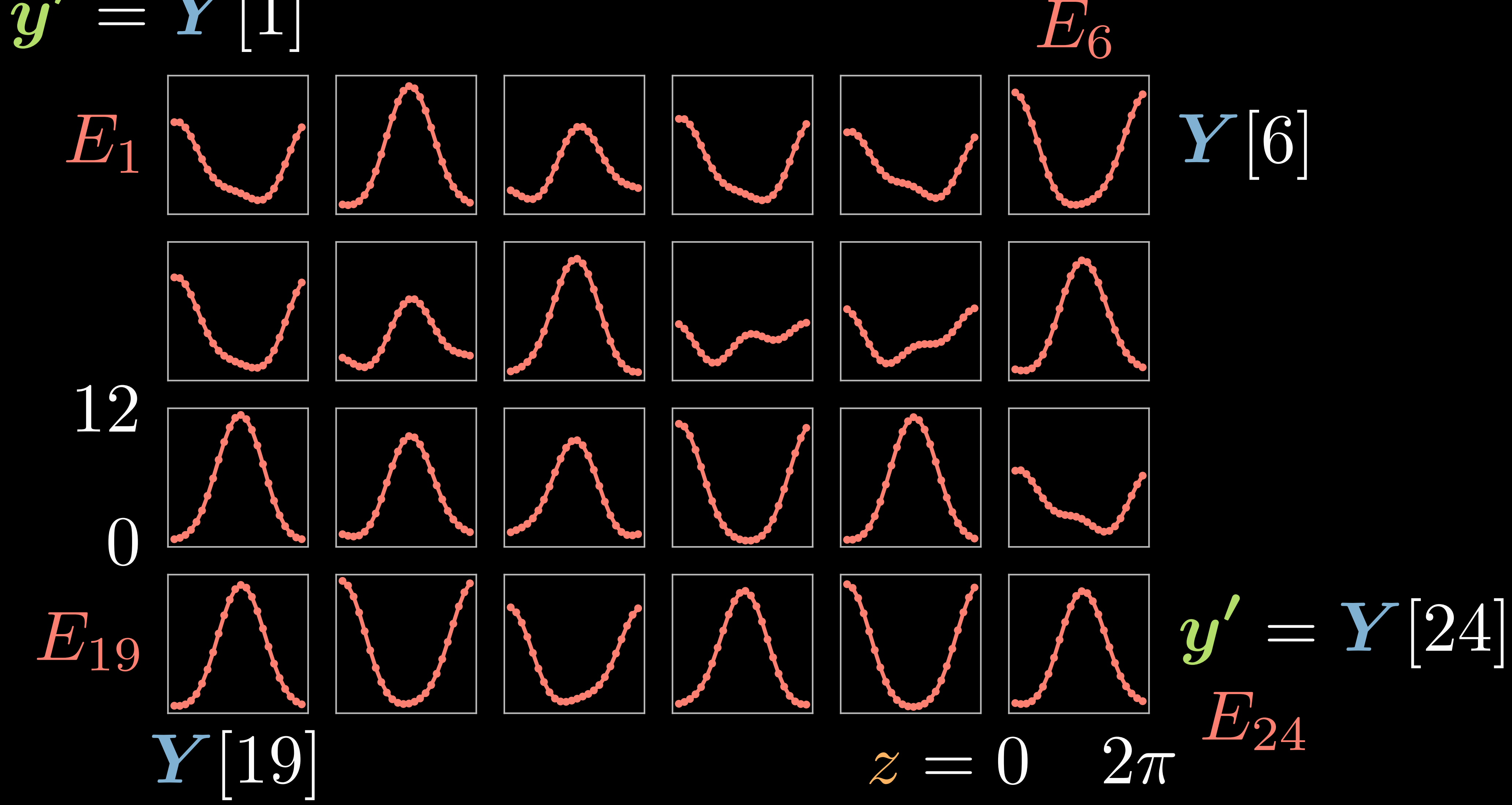


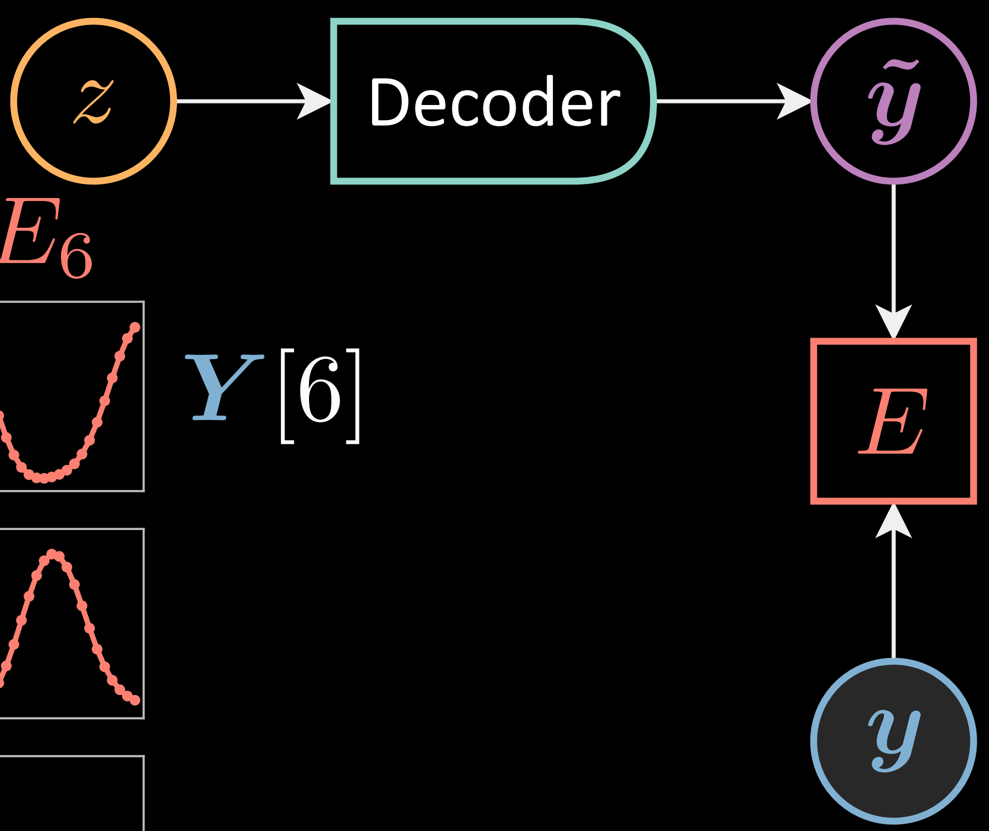
$$E(\boldsymbol{y}, z) = [y_1 - g_1(z)]^2 + [y_2 - g_2(z)]^2, \quad \boldsymbol{y} \in \boldsymbol{Y}$$

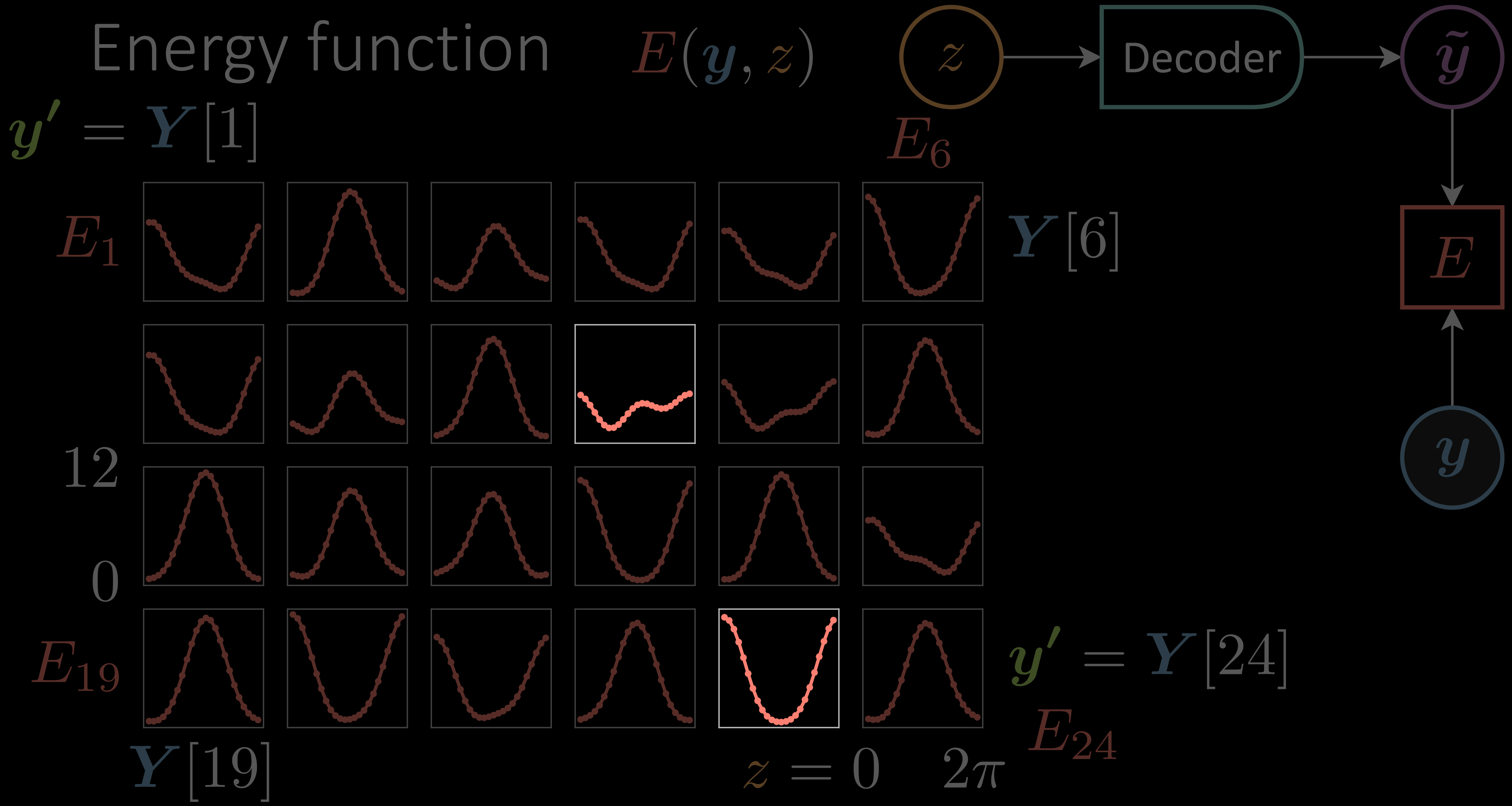
$$g = \begin{bmatrix} g_1 & g_2 \end{bmatrix}^\top : \mathbb{R} \to \mathbb{R}^2$$

$$z \mapsto \begin{bmatrix} w_1 \cos(z) & w_2 \sin(z) \end{bmatrix}^\top$$

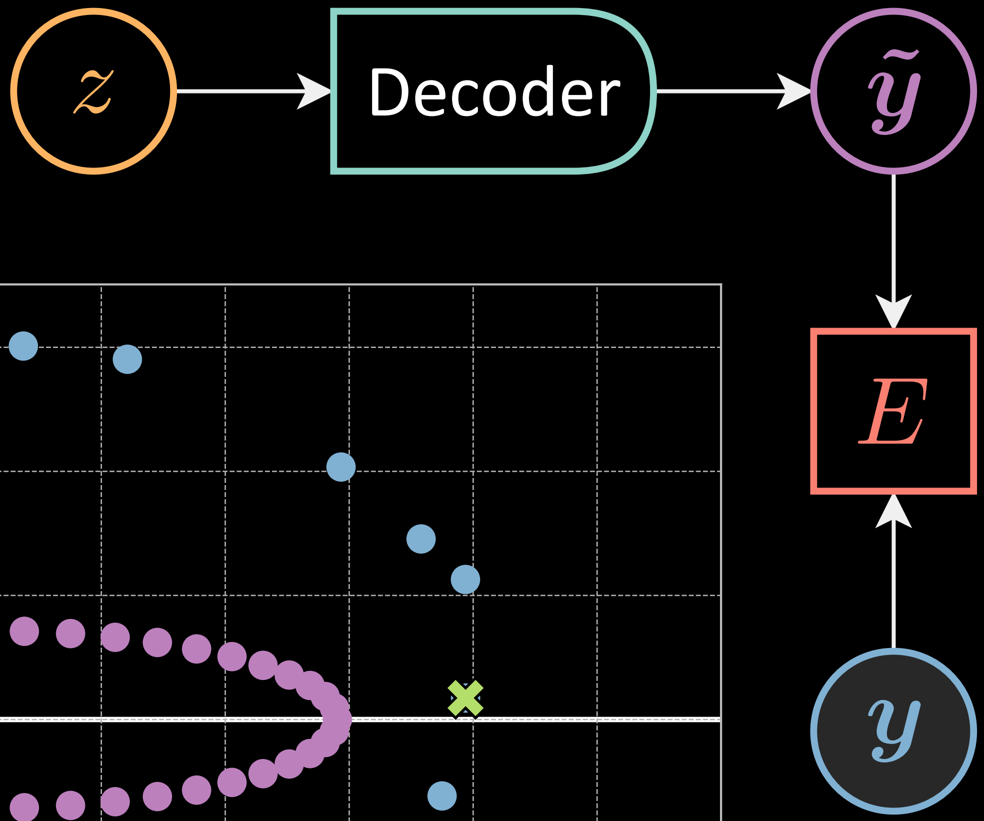
Energy function $E(\boldsymbol{y}, z)$



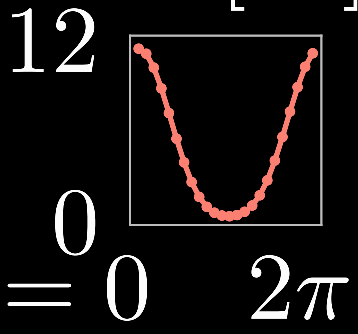


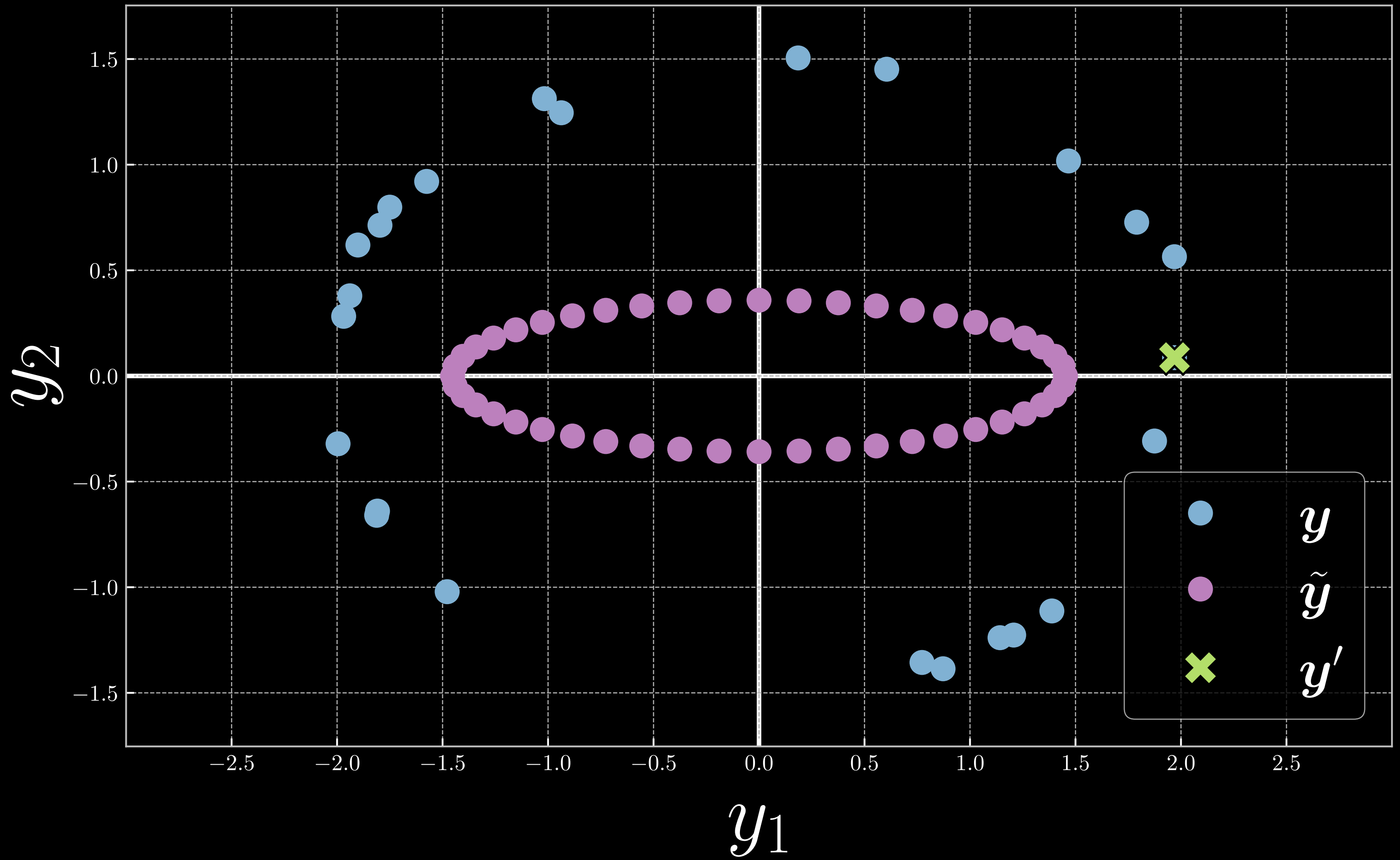

Energy function
$E(\boldsymbol{y}, z)$
z
Decoder
$\tilde{\boldsymbol{y}}$
E
$\boldsymbol{y}$
$\boldsymbol{y}' = \boldsymbol{Y}[1]$
E_6
E_1
$\boldsymbol{Y}[6]$
12
0
E_{19}
$\boldsymbol{y}' = \boldsymbol{Y}[24]$
E_{24}
$\boldsymbol{Y}[19]$
$z = 0$
2π

Energy function $E(\boldsymbol{y}, z)$



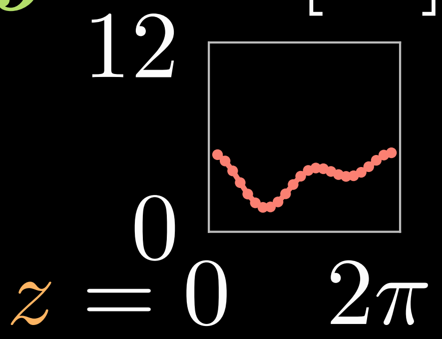
$\boldsymbol{y}' = \boldsymbol{Y}[23]$

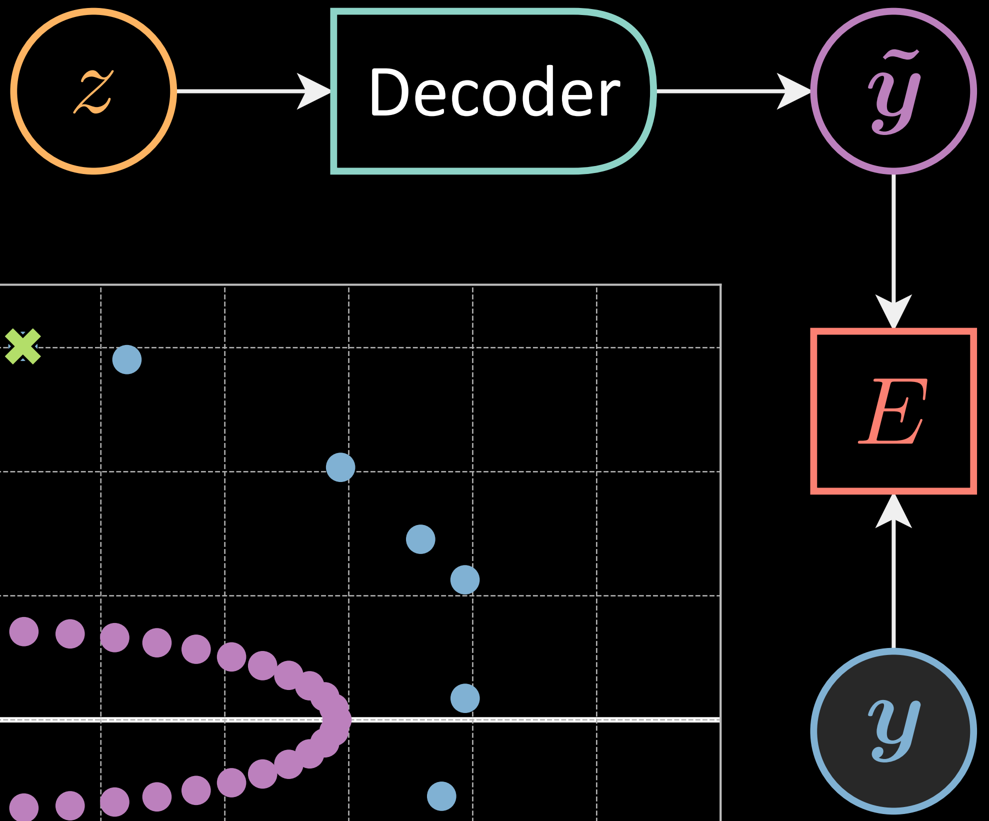


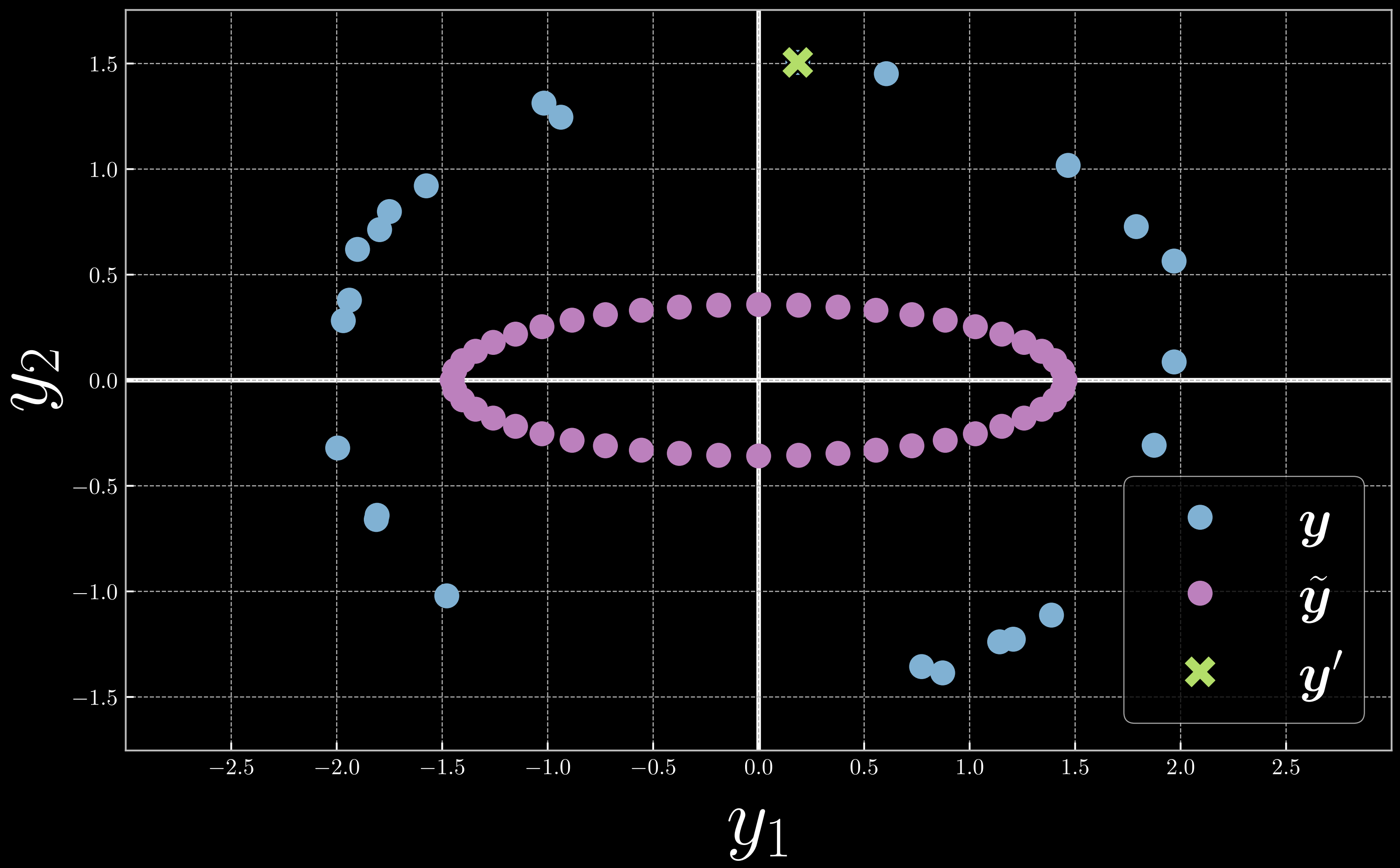


Energy function $E(\boldsymbol{y}, z)$

$\boldsymbol{y}' = \boldsymbol{Y}[10]$



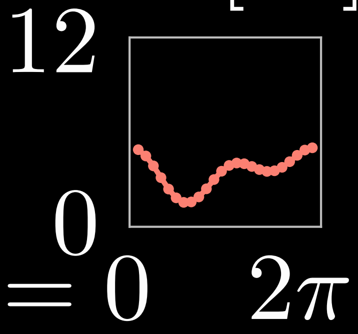




Free energy

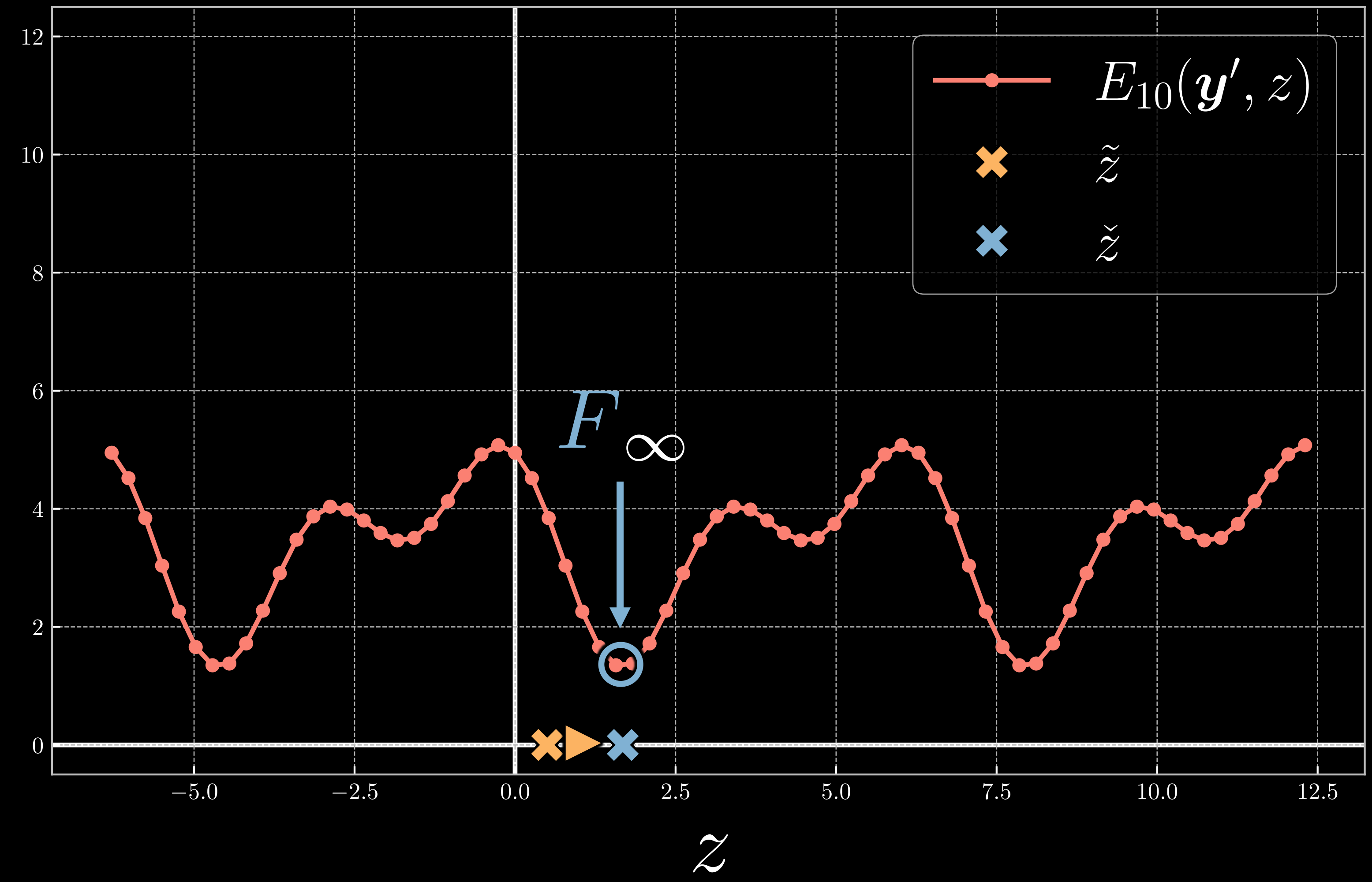
$$F_\infty(\boldsymbol{y}) = \min_z E(\boldsymbol{y}, z) = E(\boldsymbol{y}, \check{z})$$

$$\boldsymbol{y}' = \boldsymbol{Y}[10]$$



$$\check{z} = \arg\min_z E(\boldsymbol{y}, z)$$

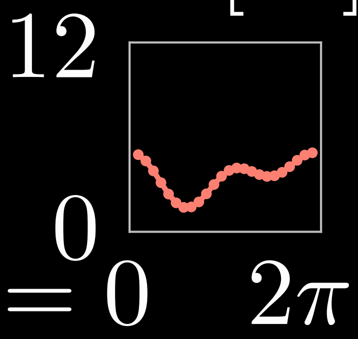
exhaustive search, conjugate gradient,
line search, LBFGS…

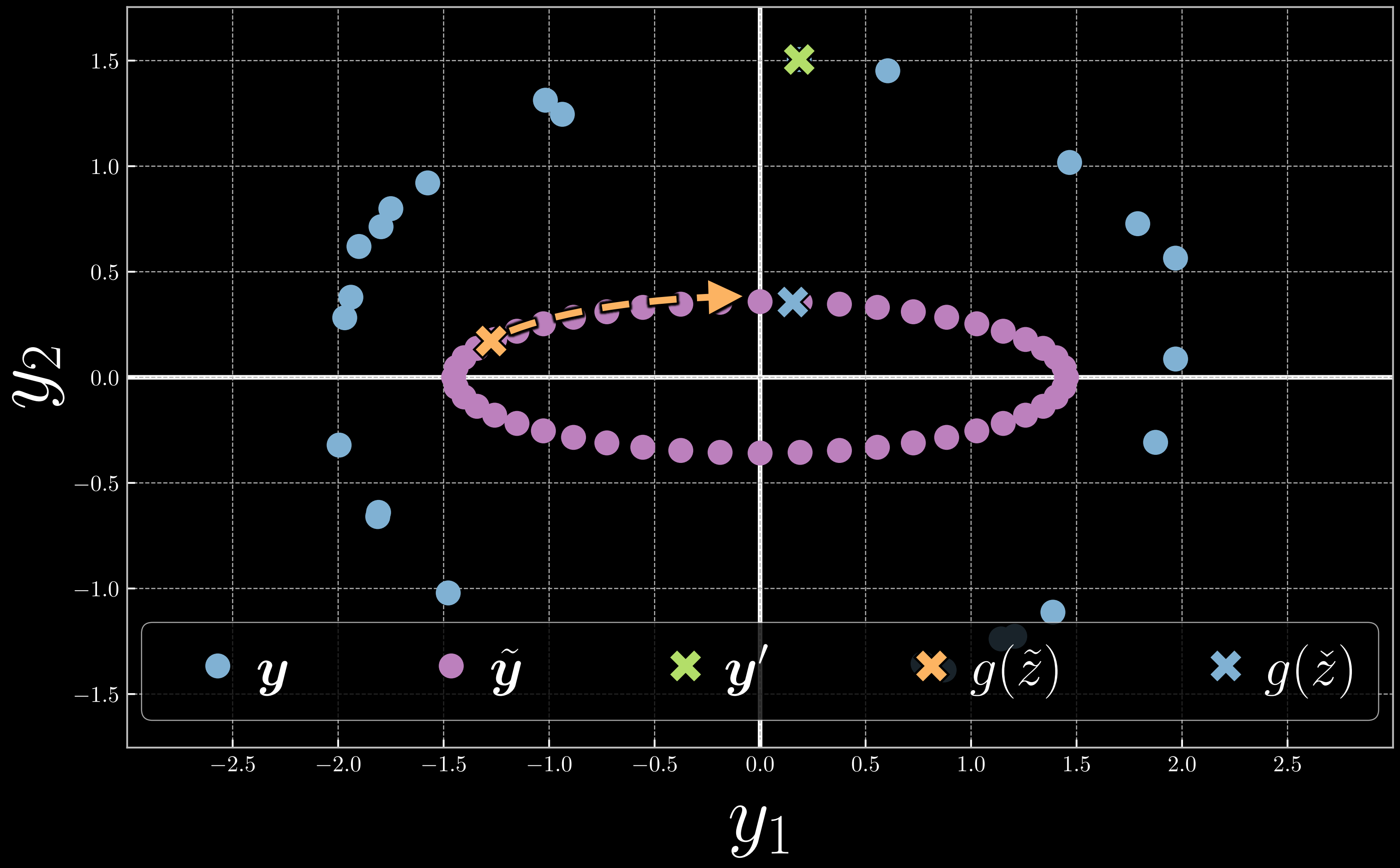


Free energy

$$F_\infty(\boldsymbol{y}) = \min_{z} E(\boldsymbol{y}, z) = E(\boldsymbol{y}, \check{z})$$

$$\boldsymbol{y}' = \boldsymbol{Y}[10]$$





$$\check{z} = \arg\min_{z} E(\boldsymbol{y}, z)$$

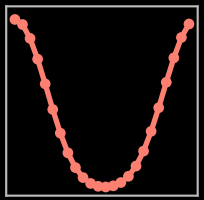
exhaustive search, conjugate gradient, line search, LBFGS…

Free energy

$$F_\infty(\boldsymbol{y}) = \min_z E(\boldsymbol{y}, z) = E(\boldsymbol{y}, \check{z})$$

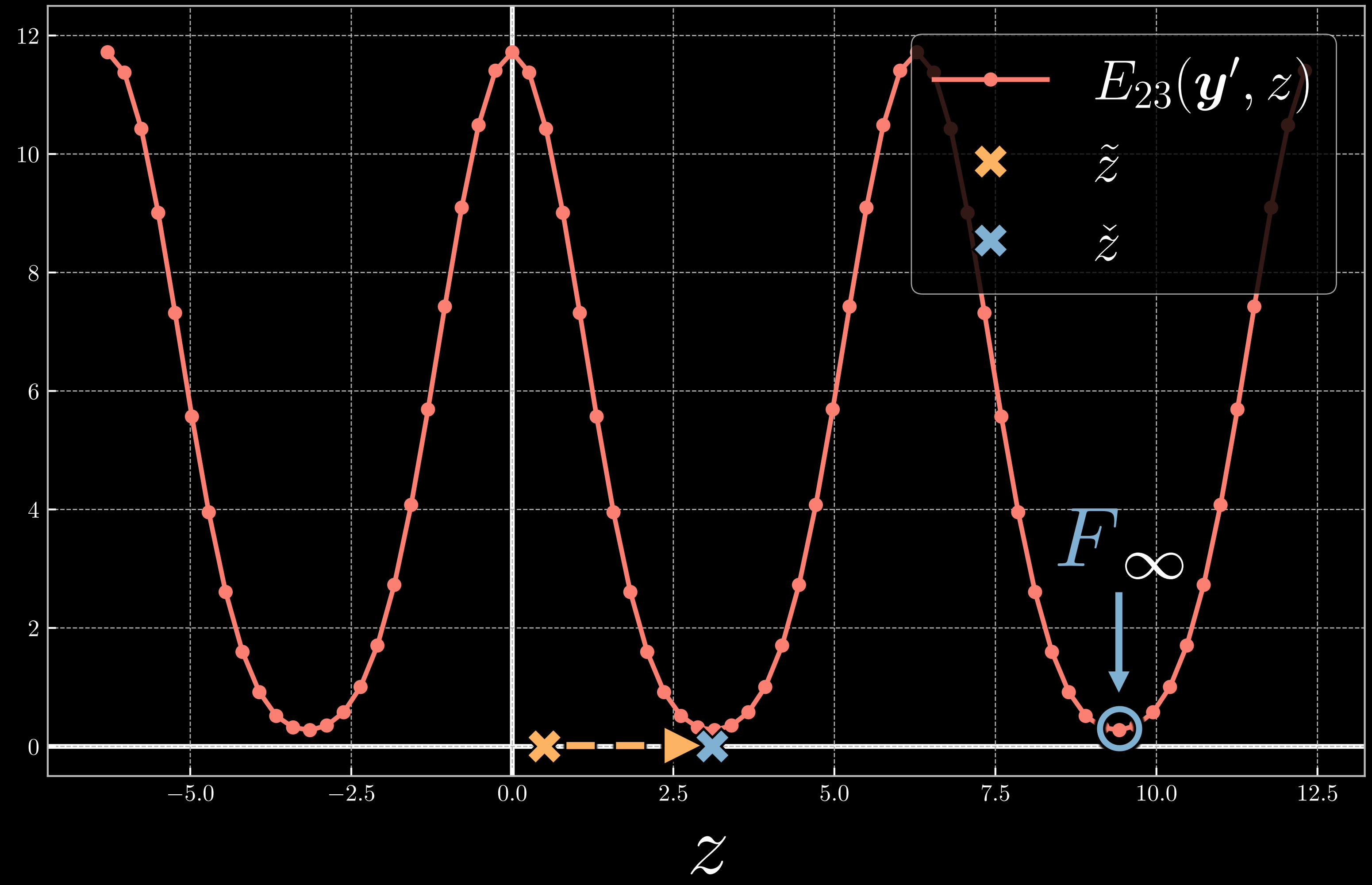
$$\boldsymbol{y}' = \boldsymbol{Y}[23]$$

12



0

$$z = 0 \quad 2\pi$$



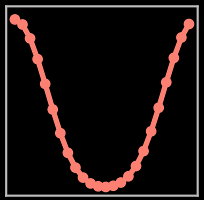
$$\check{z} = \arg\min_z E(\boldsymbol{y}, z)$$

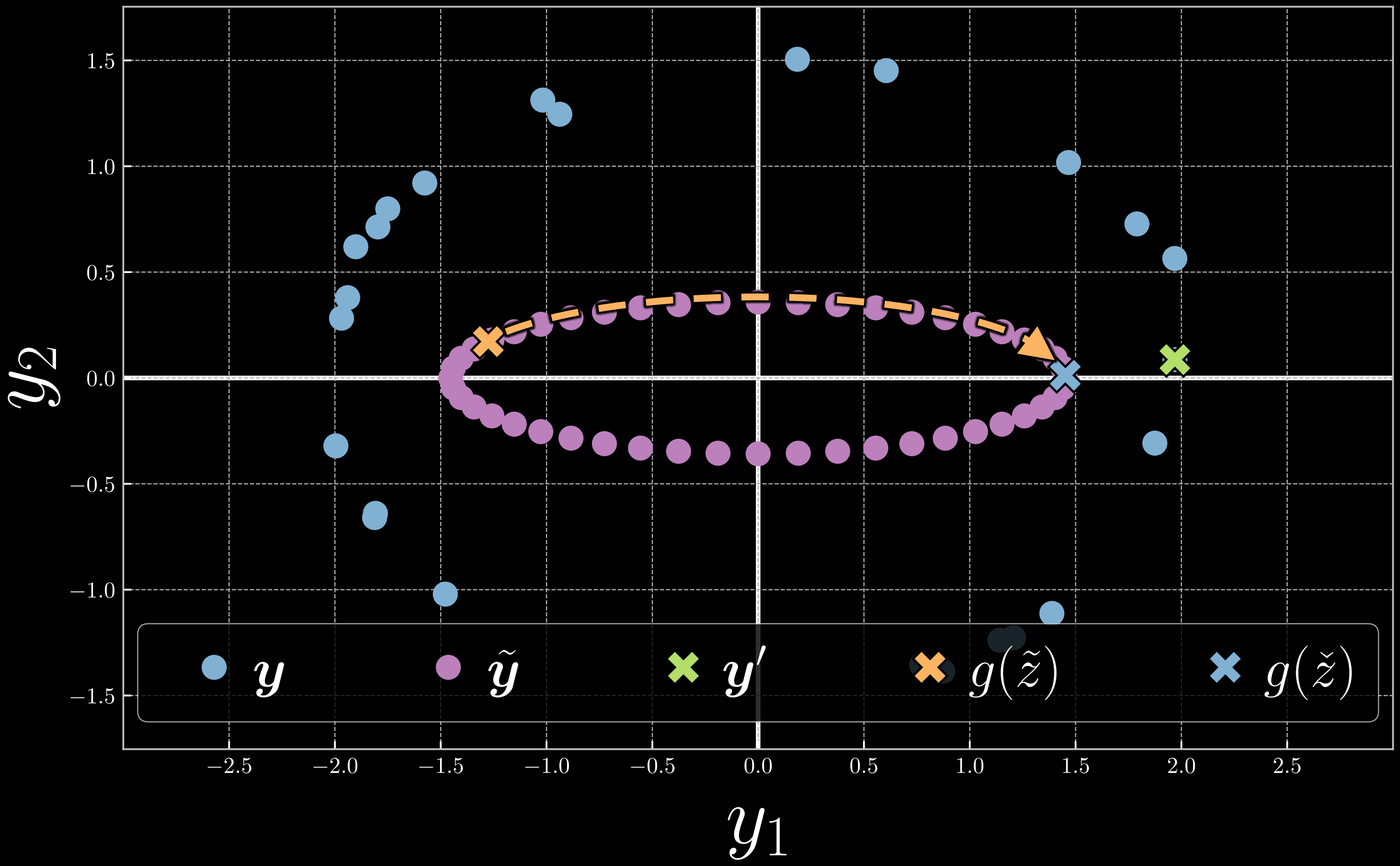
exhaustive search, conjugate gradient,
line search, LBFGS…

Free energy

$$F_\infty(\boldsymbol{y}) = \min_{z} E(\boldsymbol{y}, z) = E(\boldsymbol{y}, \check{z})$$

$$\boldsymbol{y}' = \boldsymbol{Y}[23]$$



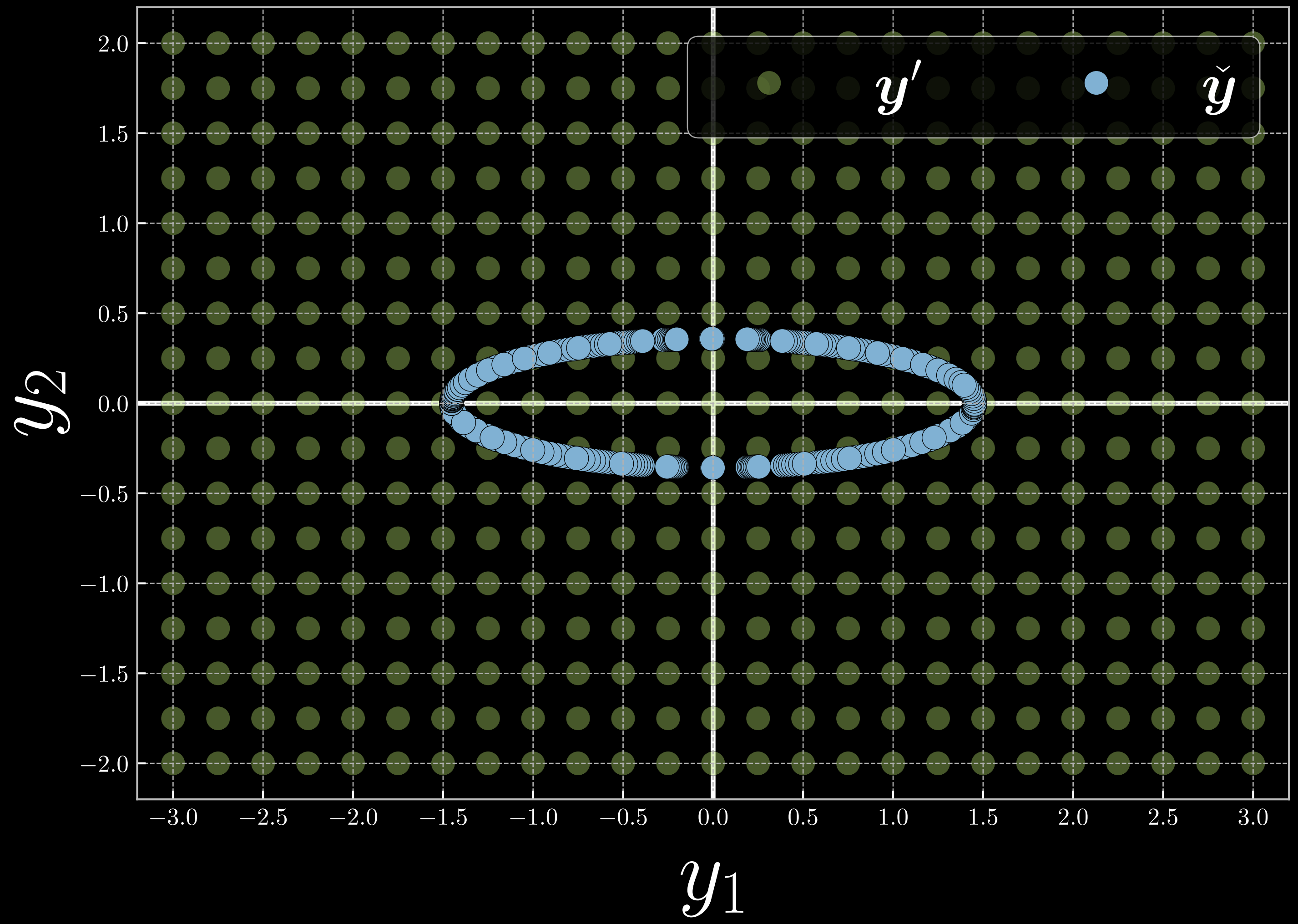


$$\check{z} = \arg\min_{z} E(\boldsymbol{y}, z)$$

exhaustive search, conjugate gradient,
line search, LBFGS…

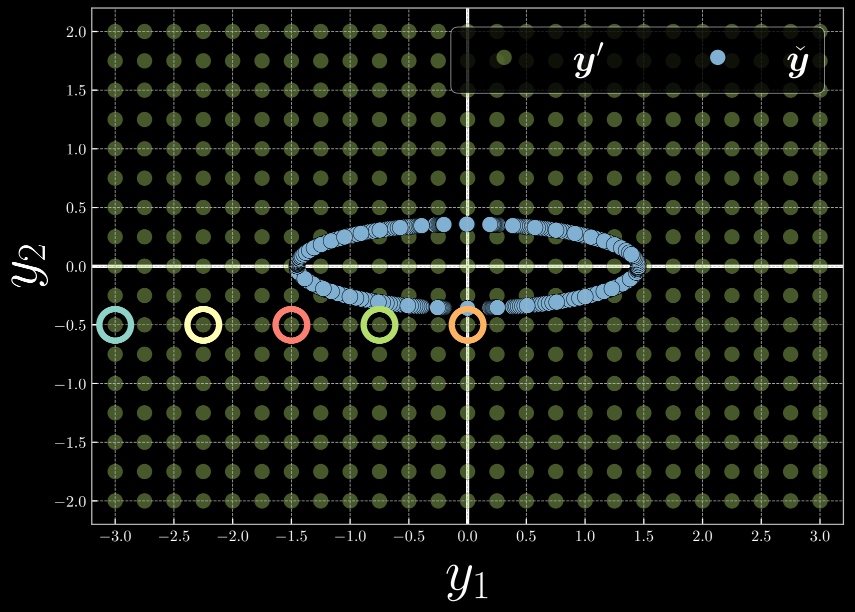
Free energy

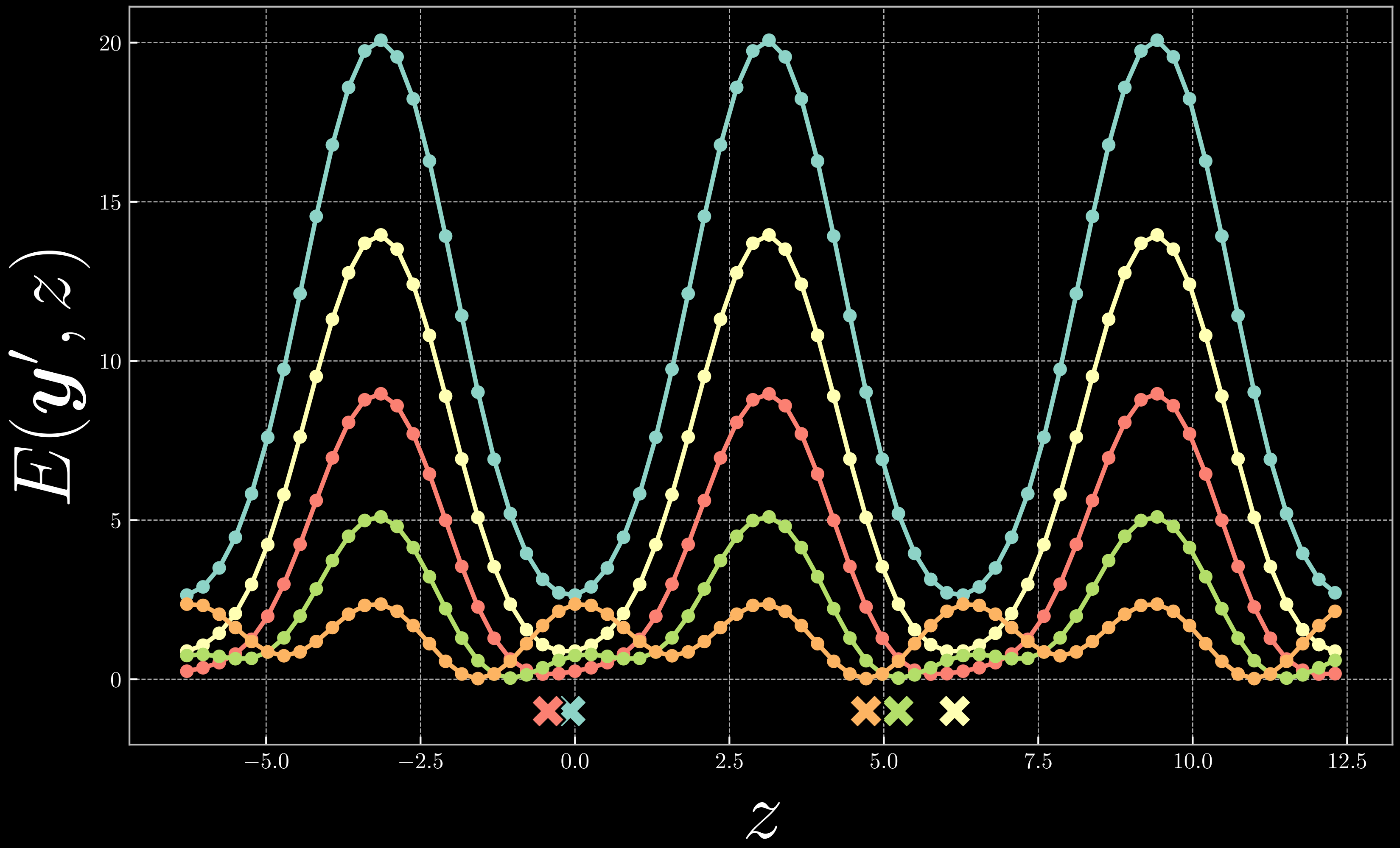
$$F_{\infty}(\boldsymbol{y}) = \min_{z} E(\boldsymbol{y}, z) = E(\boldsymbol{y}, \check{z})$$



Free energy

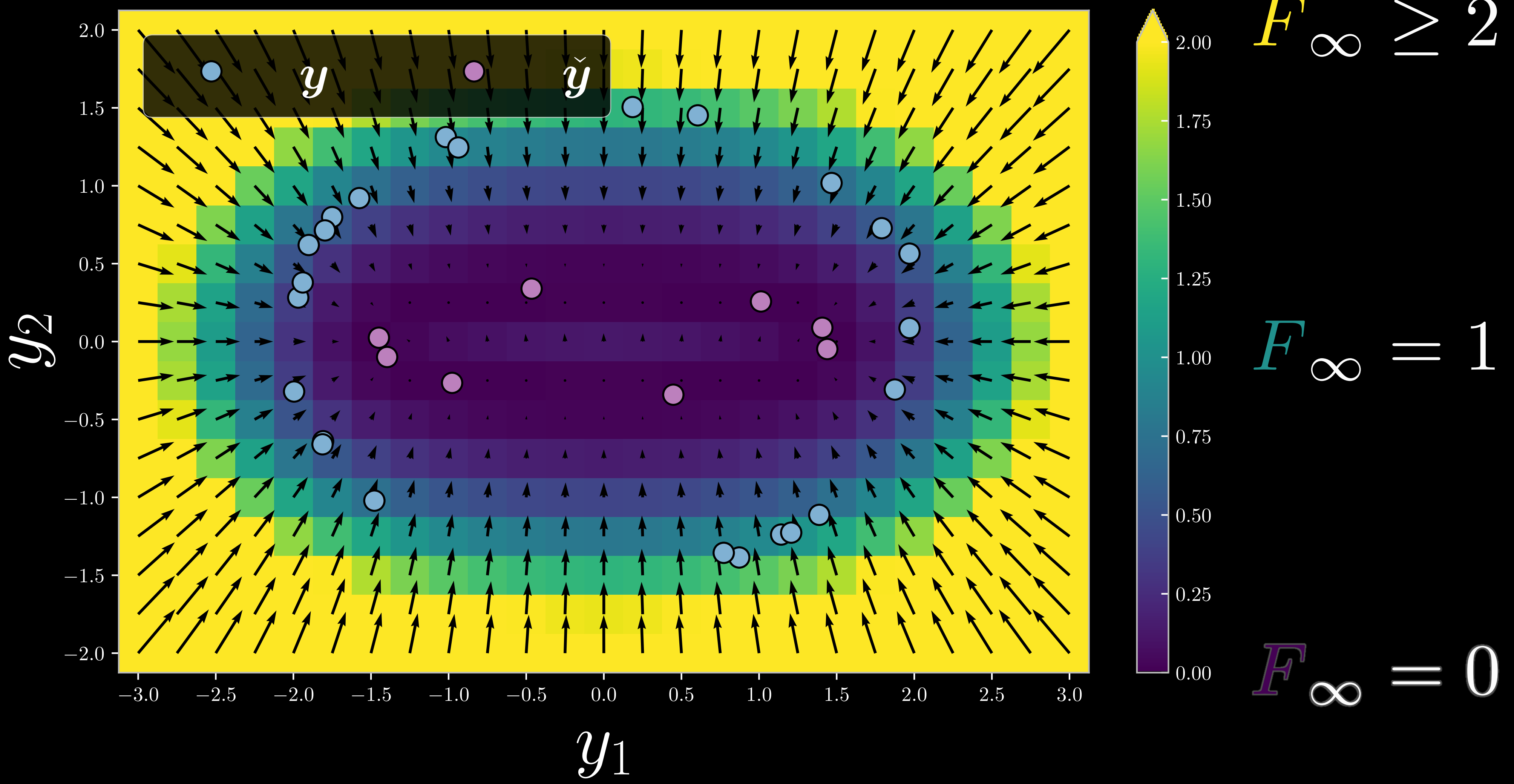
$$F_\infty(\boldsymbol{y}) = \min_{z} E(\boldsymbol{y}, z) = E(\boldsymbol{y}, \check{z})$$





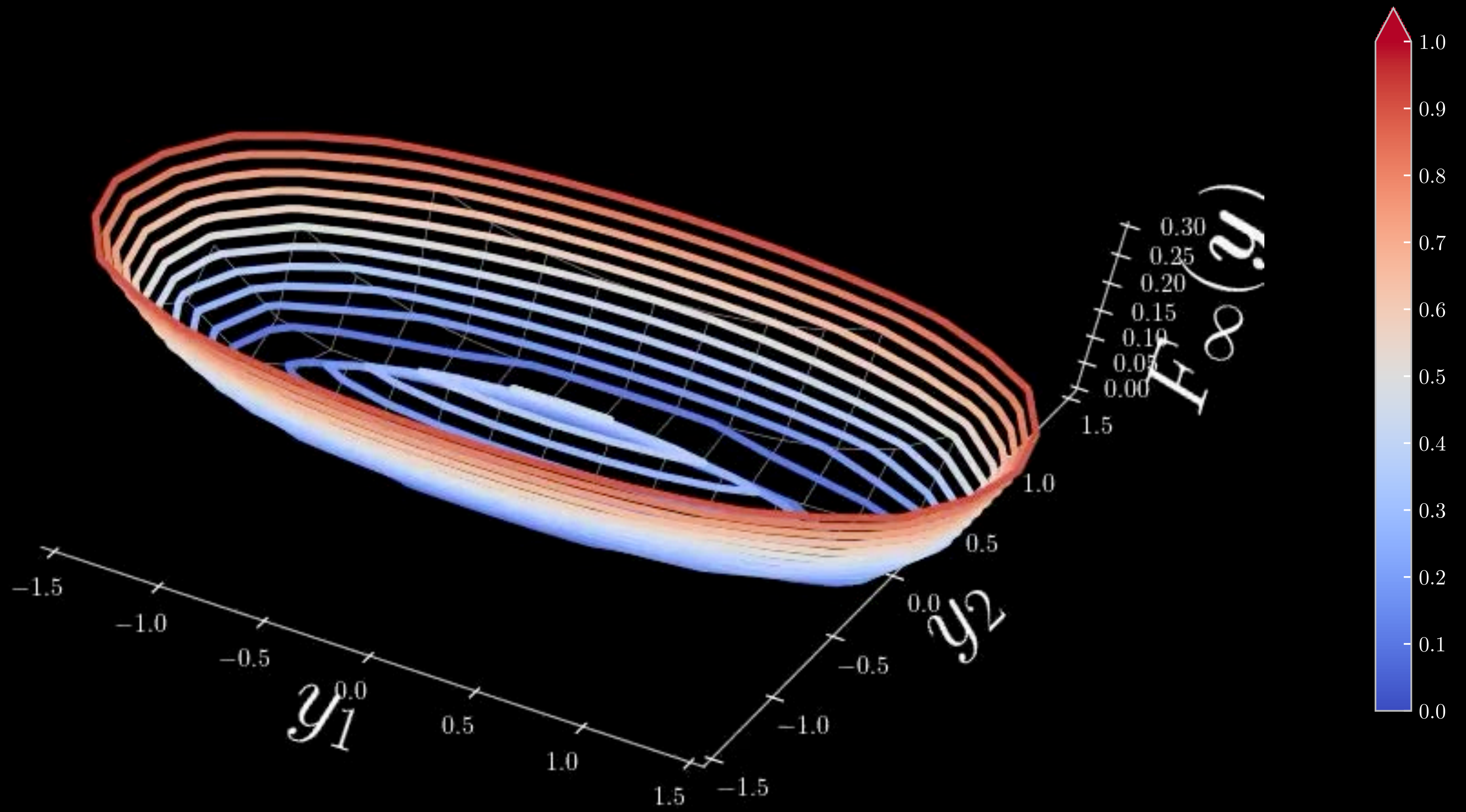
Free energy

$$F_\infty(\boldsymbol{y}) = \min_{z} E(\boldsymbol{y}, z) = E(\boldsymbol{y}, \check{z})$$



Free energy

$$\boldsymbol{F}_\infty(\boldsymbol{y}) = \min_{z} E(\boldsymbol{y}, z) = E(\boldsymbol{y}, \check{z})$$



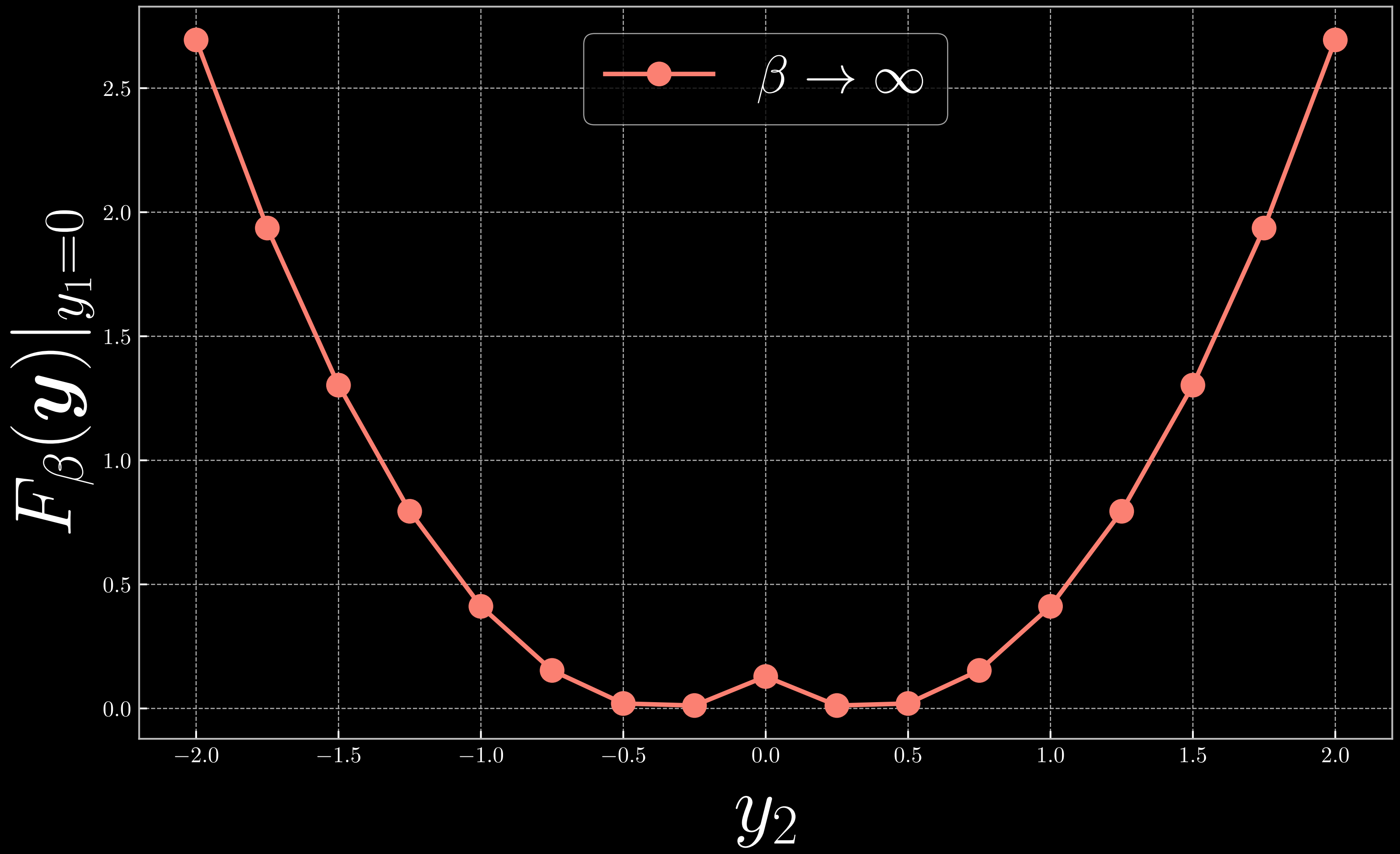
$F_\infty \geq 1$

$F_\infty = 0.5$

$F_\infty = 0$

Free energy

$$F_\infty(\boldsymbol{y}) = \min_{z} E(\boldsymbol{y}, z) = E(\boldsymbol{y}, \check{z})$$



Free energy

$$F_\infty(\boldsymbol{y}) = \min_z E(\boldsymbol{y}, z) = E(\boldsymbol{y}, \check{z})$$

zero temperature limit free energy

$$F_\beta(\boldsymbol{y}) \doteq -\frac{1}{\beta} \log \frac{1}{|\mathcal{Z}|} \int_{\mathcal{Z}} \exp[-\beta E(\boldsymbol{y}, z)] \, \mathrm{d}z$$

Boltzmann constant

average translational kinetic energy

$$\beta = (k_B T)^{-1}, \quad K_{\text{avg}} = \frac{2}{3} k_B T \; [\mathrm{J}]$$

temperature

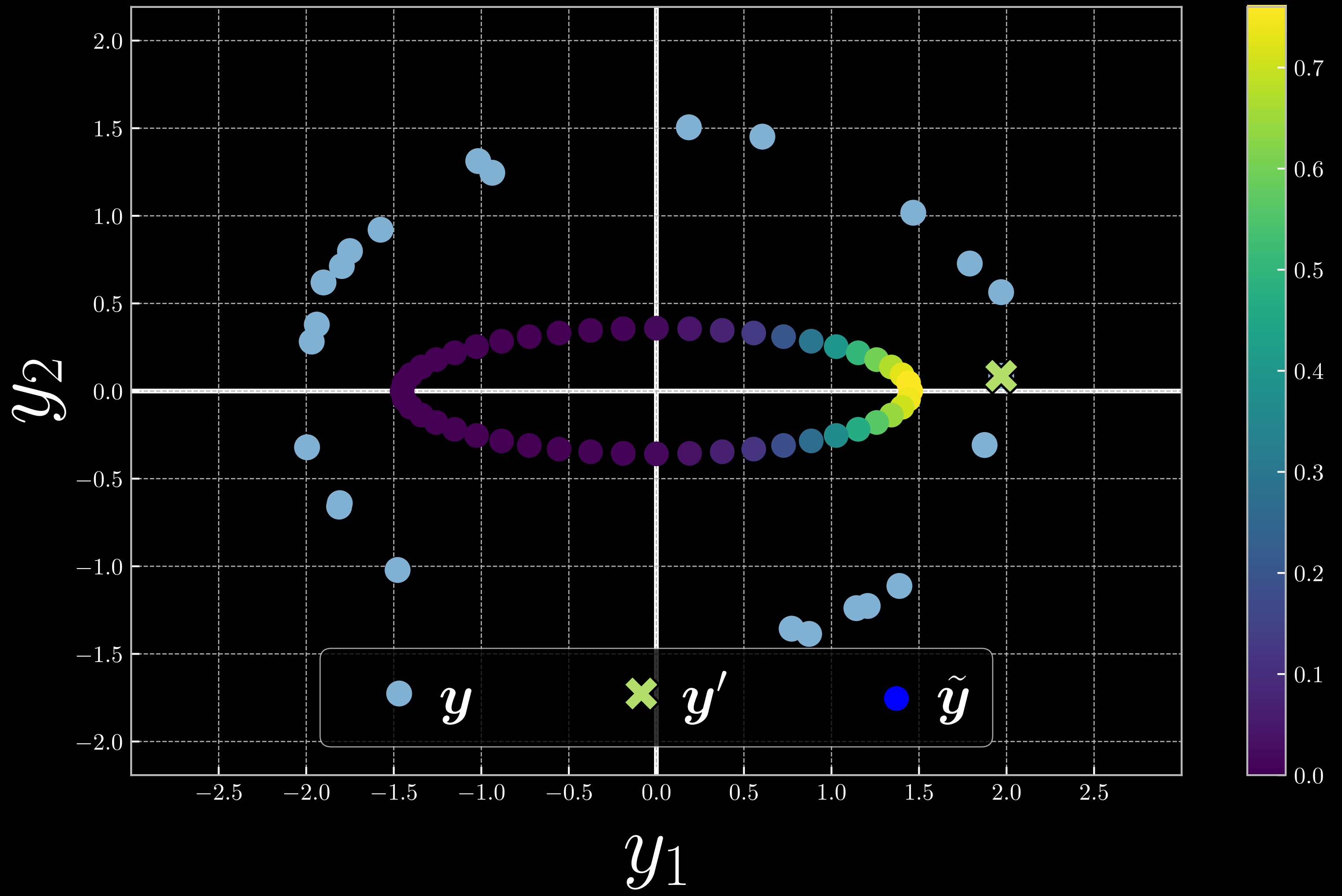
simple discrete approximation

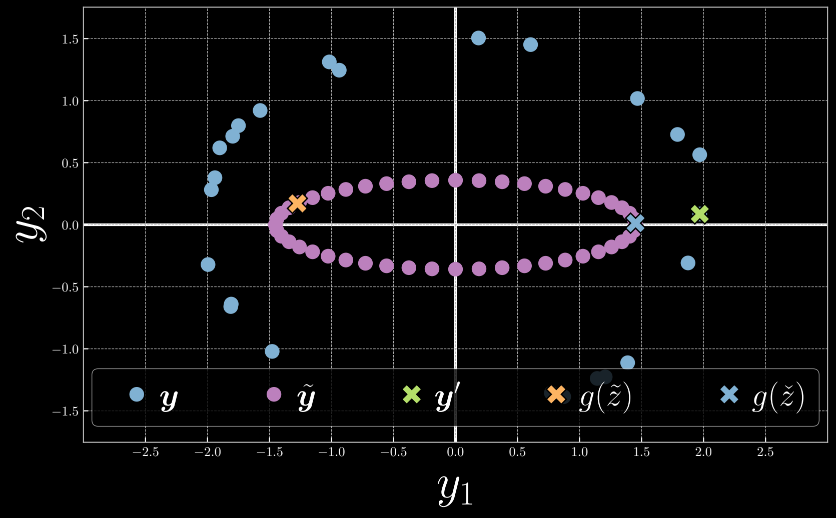
$$\tilde{F}_\beta(\boldsymbol{y}) = \underbrace{-\frac{1}{\beta} \log \frac{1}{|\mathcal{Z}|} \sum_{z \in \mathcal{Z}} \exp[-\beta E(\boldsymbol{y}, z)] \, \Delta z}_{\underset{z}{\mathrm{softmin}}_\beta [E(\boldsymbol{y}, z)]}$$

actual-softmin

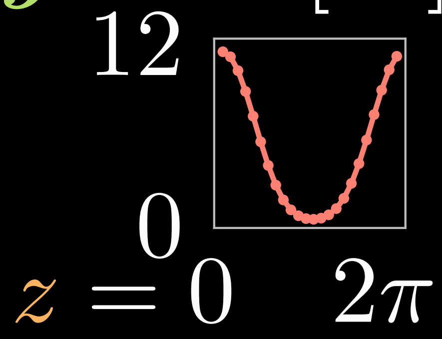
$$
\begin{aligned}
\lim_{\beta\to 0} F_\beta(\boldsymbol{y}) &= \lim_{\beta\to 0} -\frac{1}{\beta}\log\frac{1}{|\mathcal{Z}|}\int_{\mathcal{Z}} \exp[-\beta E(\boldsymbol{y}, z)]\,\mathrm{d}z = \\
&= \lim_{\beta\to 0} -\frac{\mathrm{d}}{\mathrm{d}\beta}\left[\log\frac{1}{|\mathcal{Z}|} + \log\int_{\mathcal{Z}} \exp[-\beta E(\boldsymbol{y}, z)]\,\mathrm{d}z\right] = \\
&= \lim_{\beta\to 0} -\frac{\mathrm{d}}{\mathrm{d}\beta}\log\int_{\mathcal{Z}} \exp[-\beta E(\boldsymbol{y}, z)]\,\mathrm{d}z = \\
&= \lim_{\beta\to 0} -\frac{1}{\int_{\mathcal{Z}} \exp[-\beta E(\boldsymbol{y}, z)]\,\mathrm{d}z}\frac{\mathrm{d}}{\mathrm{d}\beta}\int_{\mathcal{Z}} \exp[-\beta E(\boldsymbol{y}, z)]\,\mathrm{d}z = \\
&= \lim_{\beta\to 0} -\frac{1}{\int_{\mathcal{Z}} \exp[-\beta E(\boldsymbol{y}, z)]\,\mathrm{d}z}\int_{\mathcal{Z}} -E(\boldsymbol{y}, z)\exp[-\beta E(\boldsymbol{y}, z)]\,\mathrm{d}z = \\
&= \frac{1}{|\mathcal{Z}|}\int_{\mathcal{Z}} E(\boldsymbol{y}, z)\,\mathrm{d}z = \langle E(\boldsymbol{y}, z)\rangle
\end{aligned}
$$

Free energy $F_\beta(\boldsymbol{y}) \doteq -\frac{1}{\beta} \log \frac{1}{|\mathcal{Z}|} \int_{\mathcal{Z}} \exp[-\beta E(\boldsymbol{y}, z)] \, \mathrm{d}z$



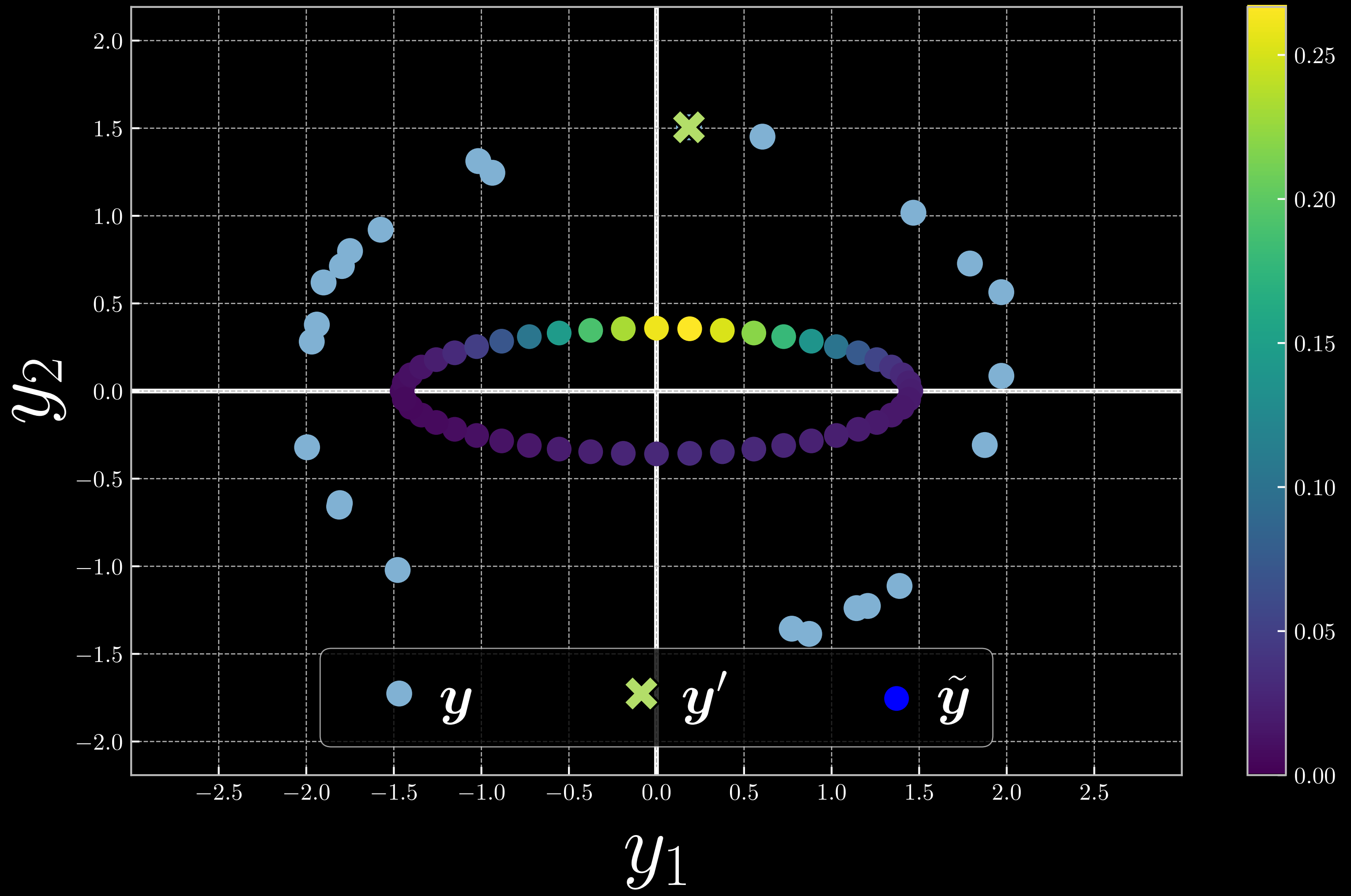


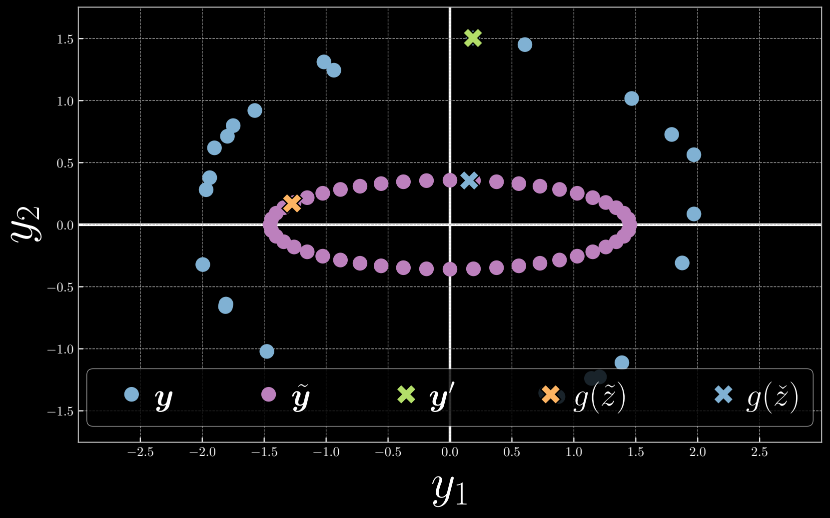
$\boldsymbol{y}' = \boldsymbol{Y}[23]$



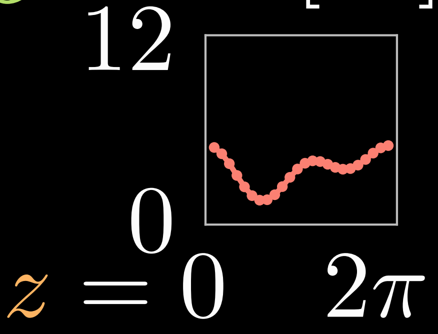
$\beta = 1$

Free energy $F_\beta(\boldsymbol{y}) \doteq -\frac{1}{\beta} \log \frac{1}{|\mathcal{Z}|} \int_{\mathcal{Z}} \exp[-\beta E(\boldsymbol{y}, z)] \, \mathrm{d}z$



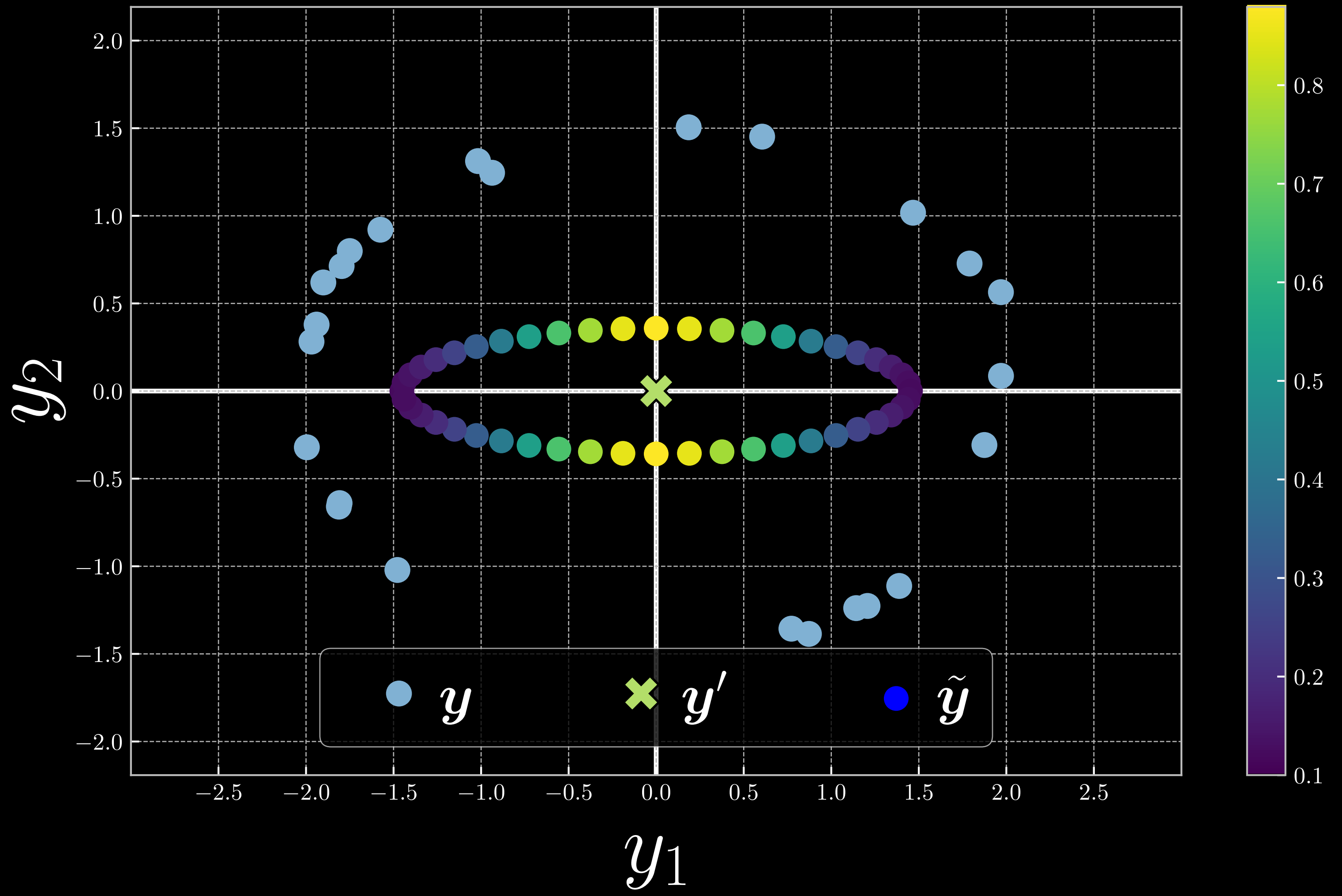


$\boldsymbol{y}' = \boldsymbol{Y}[10]$



$\beta = 1$

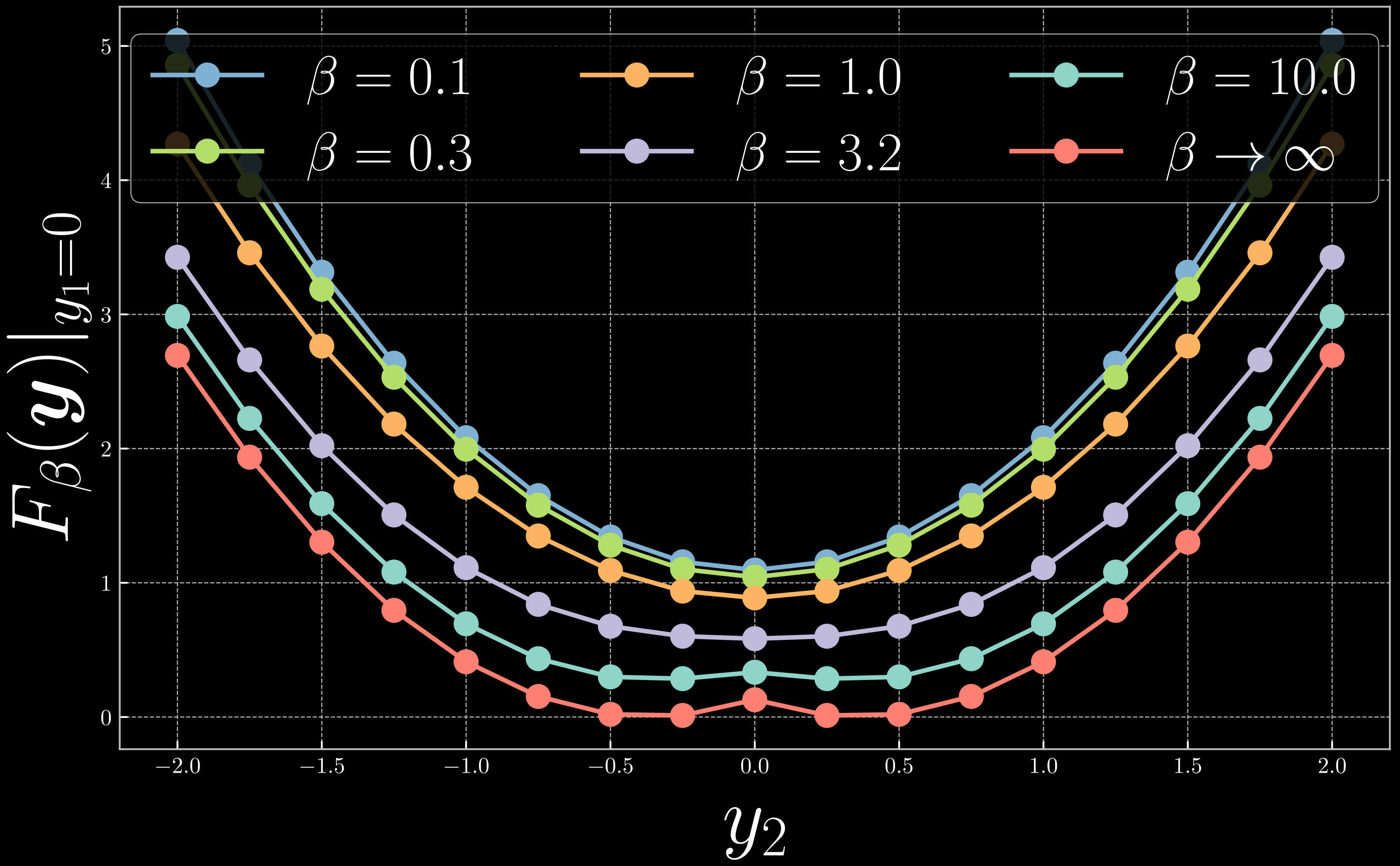
Free energy $F_\beta(\boldsymbol{y}) \doteq -\frac{1}{\beta} \log \frac{1}{|\mathcal{Z}|} \int_{\mathcal{Z}} \exp[-\beta E(\boldsymbol{y}, z)] \, \mathrm{d}z$



$\boldsymbol{y}' = (0, 0)$

$\beta = 1$

Free energy $F_\beta(\boldsymbol{y}) \doteq -\frac{1}{\beta} \log \frac{1}{|\mathcal{Z}|} \int_{\mathcal{Z}} \exp[-\beta E(\boldsymbol{y}, z)]\, \mathrm{d}z$



Nomenclature and PyTorch

actual-softmax

$$\operatorname*{softmax}_{z\,\beta}[E(\boldsymbol{y},z)] \doteq \frac{1}{\beta}\log\sum_{z\in\mathcal{Z}}\exp[\beta E(\boldsymbol{y},z)] - \underbrace{\frac{1}{\beta}\log N_z}$$

$$N_z \leftarrow \frac{|\mathcal{Z}|}{\Delta z}$$

$$= \frac{1}{\beta}\,\texttt{torch.logsumexp(}\beta E(\boldsymbol{y},z)\texttt{, dim=}z\texttt{)}$$

actual-softmin

$$\operatorname*{softmin}_{z\,\beta}[E(\boldsymbol{y},z)] \doteq -\frac{1}{\beta}\log\frac{1}{N_z}\sum_{z\in\mathcal{Z}}\exp[-\beta E(\boldsymbol{y},z)]$$

$$= -\operatorname*{softmax}_{z\,\beta}[-E(\boldsymbol{y},z)]$$

$$\texttt{torch.softmax(}l(j)\texttt{, dim=}j\texttt{)} = \operatorname*{softargmax}_{j\,\beta=1}[l(j)]$$

TRAINING

Finding a well behaved energy function

Loss functional

$$\mathcal{L}(F(\,\cdot\,), \boldsymbol{Y}) = \frac{1}{N} \sum_{n=1}^{N} \ell(F(\,\cdot\,), \boldsymbol{y}^{(n)}) \in \mathbb{R}$$

$$\ell_{\text{energy}}\big(F(\,\cdot\,), \check{\boldsymbol{y}}\big) = F(\check{\boldsymbol{y}})$$

$$\ell_{\text{hinge}}\big(F(\,\cdot\,), \check{\boldsymbol{y}}, \hat{\boldsymbol{y}}\big) = \big(m - [F(\hat{\boldsymbol{y}}) - F(\check{\boldsymbol{y}})]\big)^{+}$$

$$\ell_{\log}\big(F(\,\cdot\,), \check{\boldsymbol{y}}, \hat{\boldsymbol{y}}\big) = \log\big(1 + \exp[F(\check{\boldsymbol{y}}) - F(\hat{\boldsymbol{y}})]\big)$$

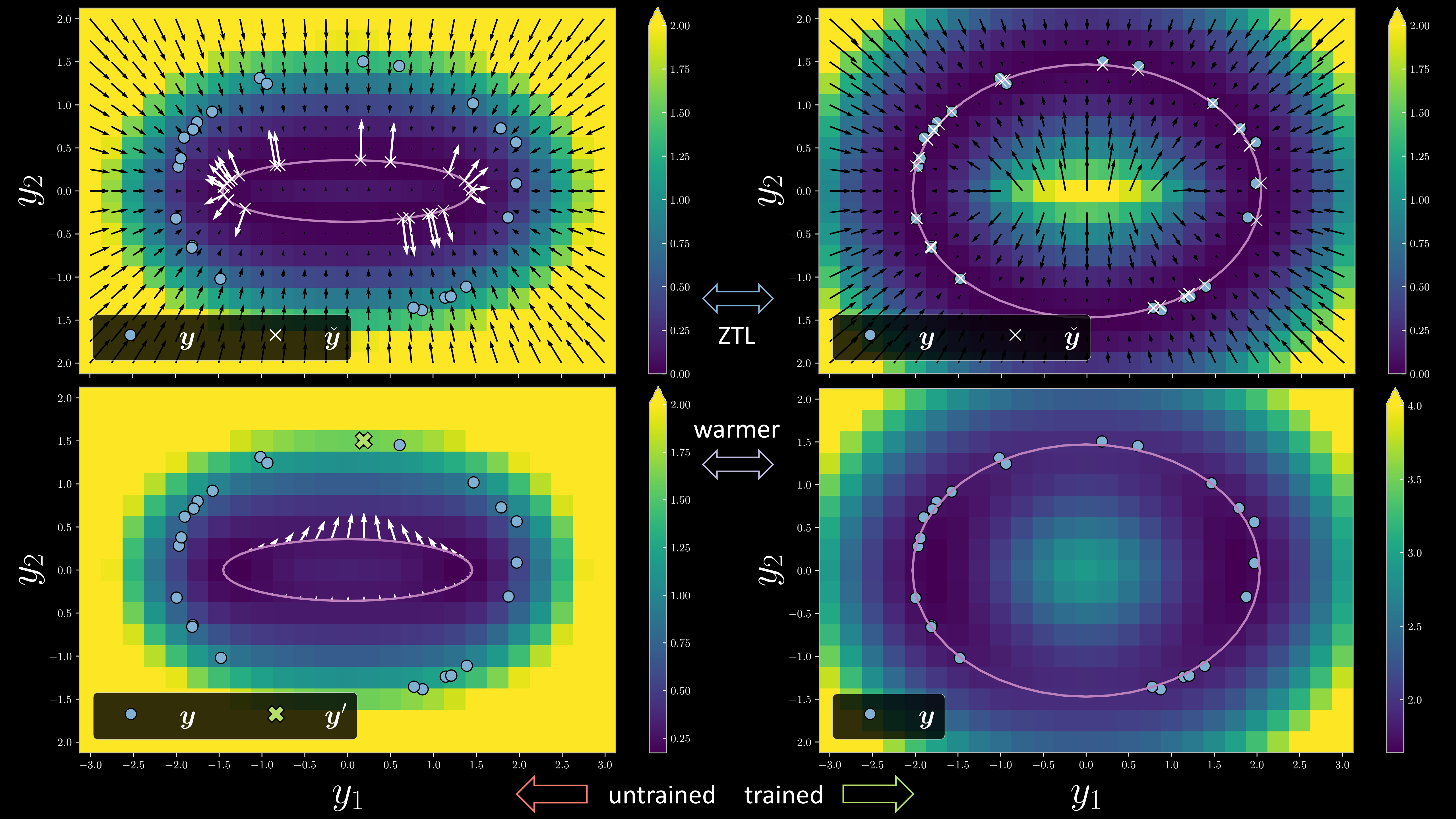
Loss functional

$$\mathcal{L}(F(\cdot), \boldsymbol{Y}) = \frac{1}{N} \sum_{n=1}^{N} \ell(F(\cdot), \boldsymbol{y}^{(n)})$$

$$\ell_{\text{energy}}\big(F(\cdot), \check{\boldsymbol{y}}\big) = F(\check{\boldsymbol{y}})$$

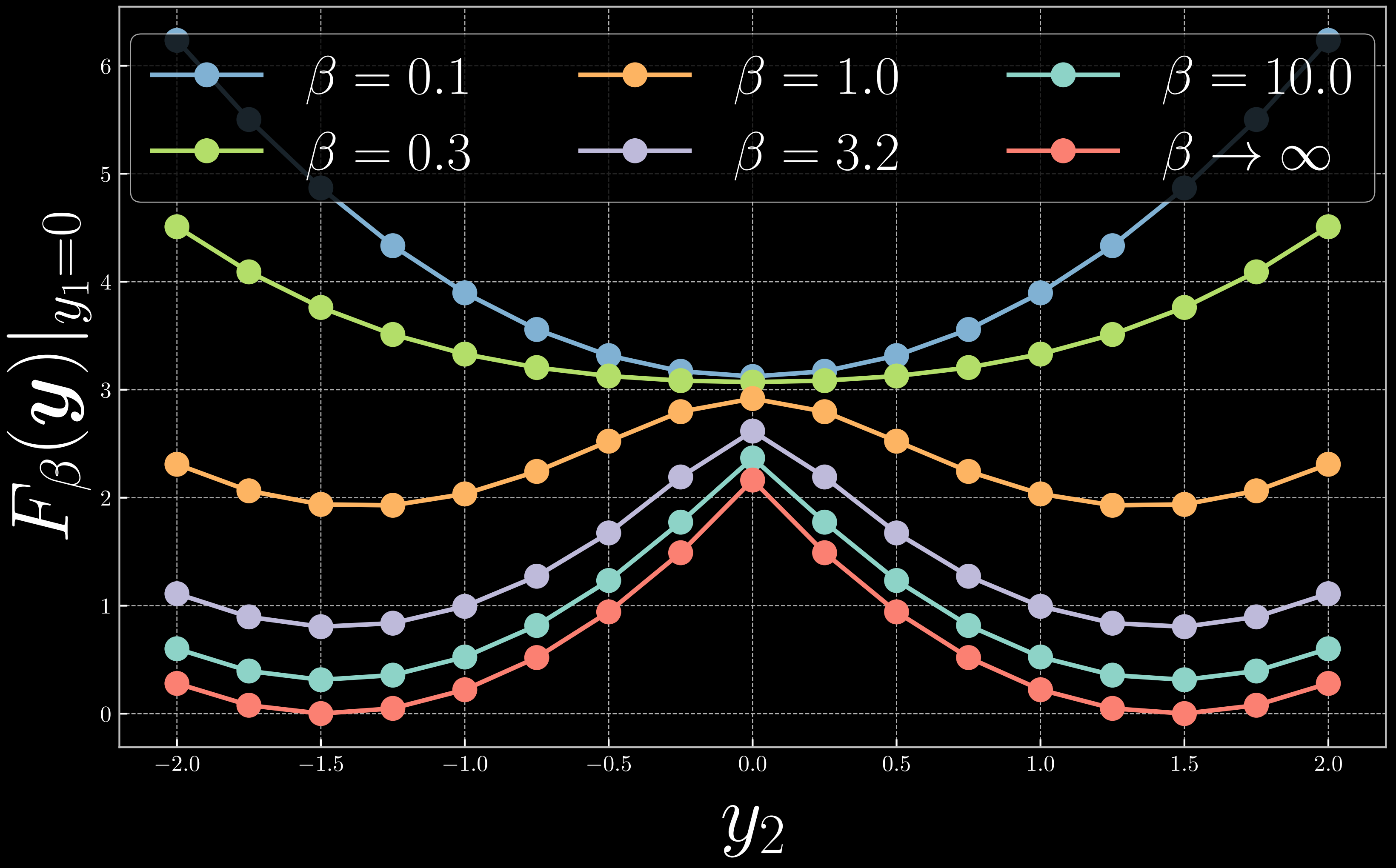
$$\ell_{\text{hinge}}\big(F(\cdot), \check{\boldsymbol{y}}, \hat{\boldsymbol{y}}\big) = \big(m - [F(\hat{\boldsymbol{y}}) - F(\check{\boldsymbol{y}})]\big)^{+}$$

$$\ell_{\log}\big(F(\cdot), \check{\boldsymbol{y}}, \hat{\boldsymbol{y}}\big) = \log\big(1 + \exp[F(\check{\boldsymbol{y}}) - F(\hat{\boldsymbol{y}})]\big)$$



y_2
y_1
y
$\check{y}$
y'
ZTL
warmer
untrained
trained

Free energy $F_\beta(\boldsymbol{y}) \doteq -\frac{1}{\beta} \log \frac{1}{|\mathcal{Z}|} \int_{\mathcal{Z}} \exp[-\beta E(\boldsymbol{y}, z)] \, \mathrm{d}z$

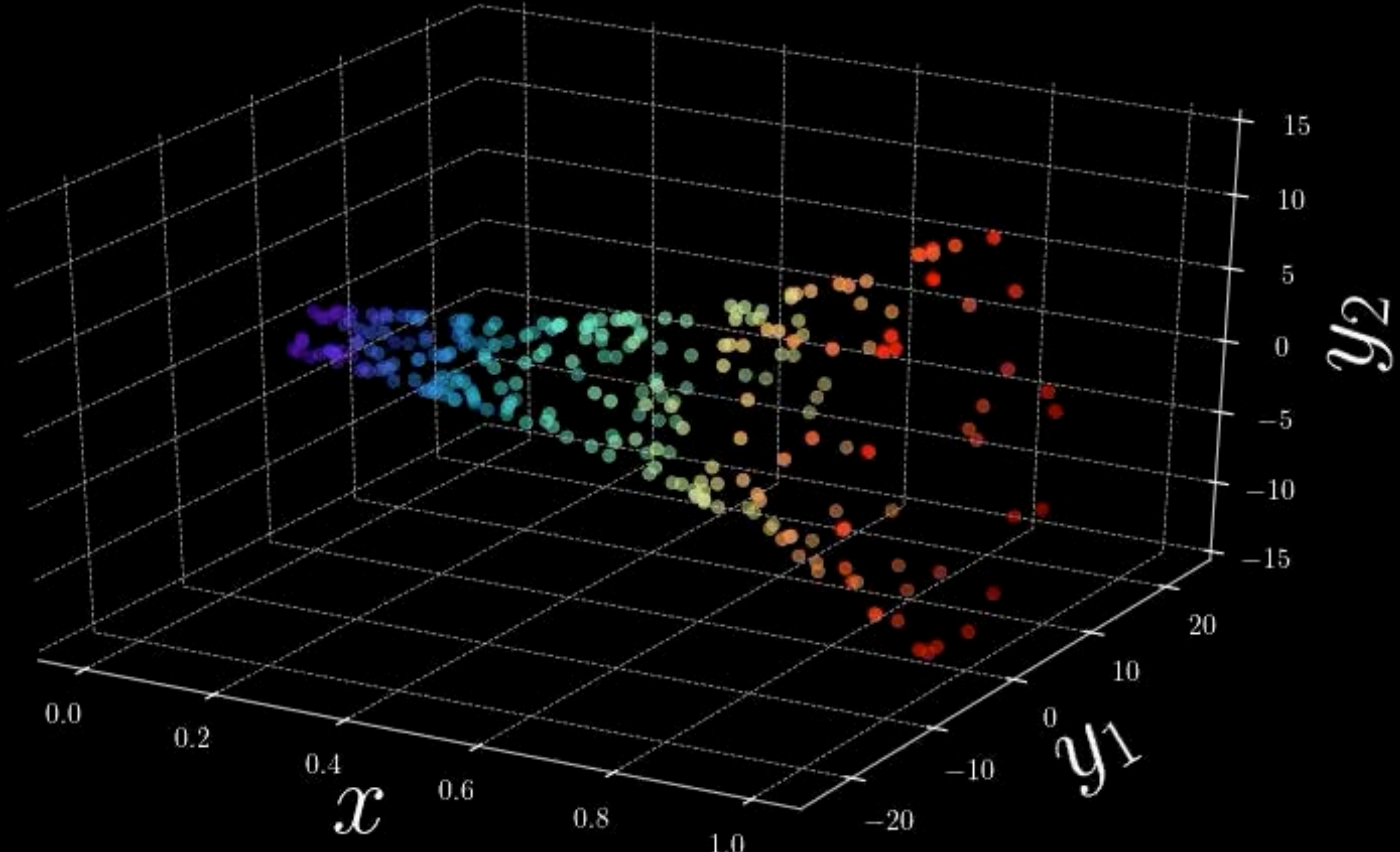


Self-supervised learning

Conditional case

Training samples

$$\alpha = 1.5$$
$$\beta = 2$$



$$\boldsymbol{y} = \begin{bmatrix} \rho_1(x)\cos(\theta) + \varepsilon \\ \rho_2(x)\sin(\theta) + \varepsilon \end{bmatrix}$$

$$\rho : \mathbb{R} \to \mathbb{R}^2$$

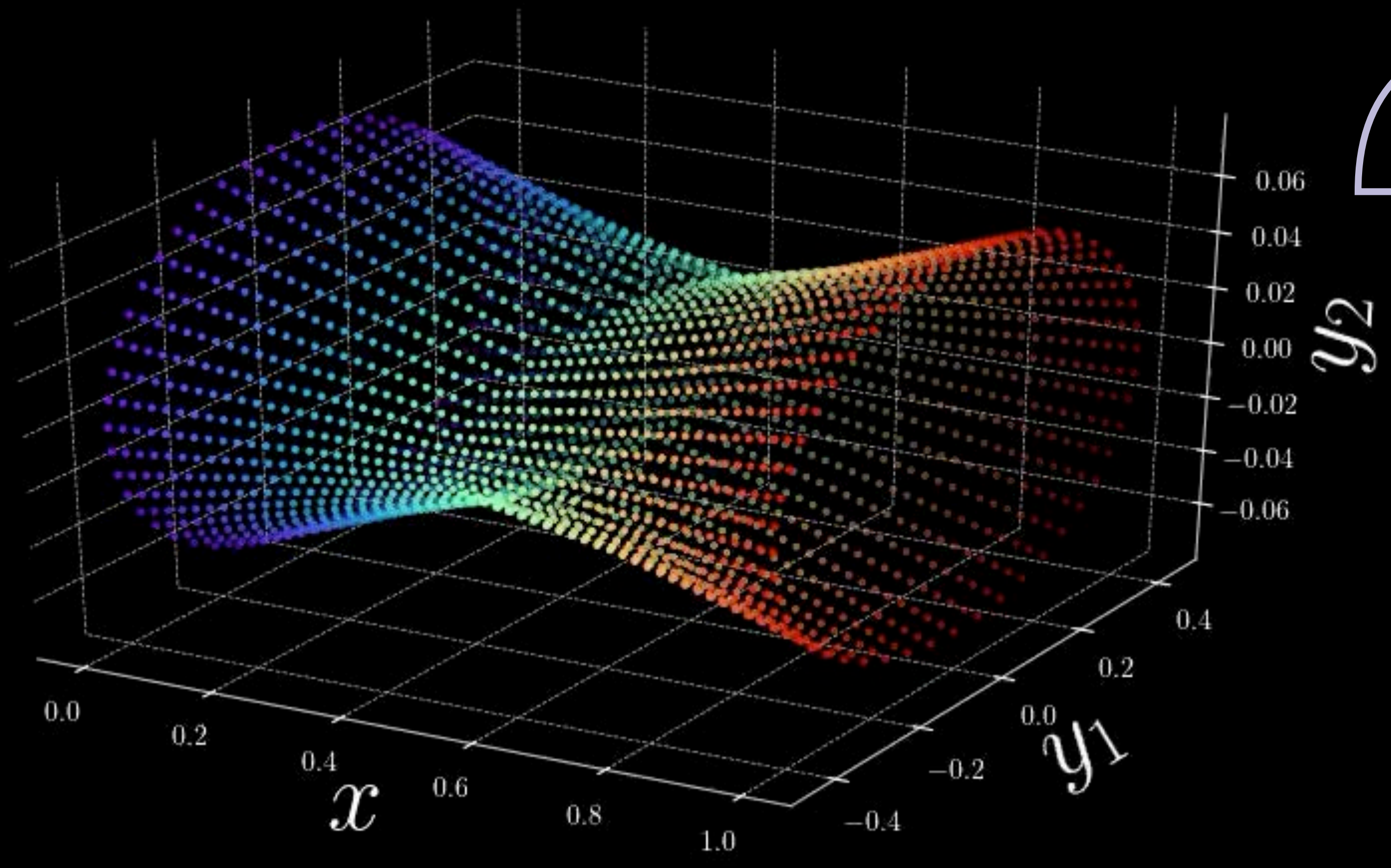
$$x \mapsto \begin{bmatrix} \alpha x + \beta(1-x) \\ \beta x + \alpha(1-x) \end{bmatrix} \cdot \exp(2x)$$

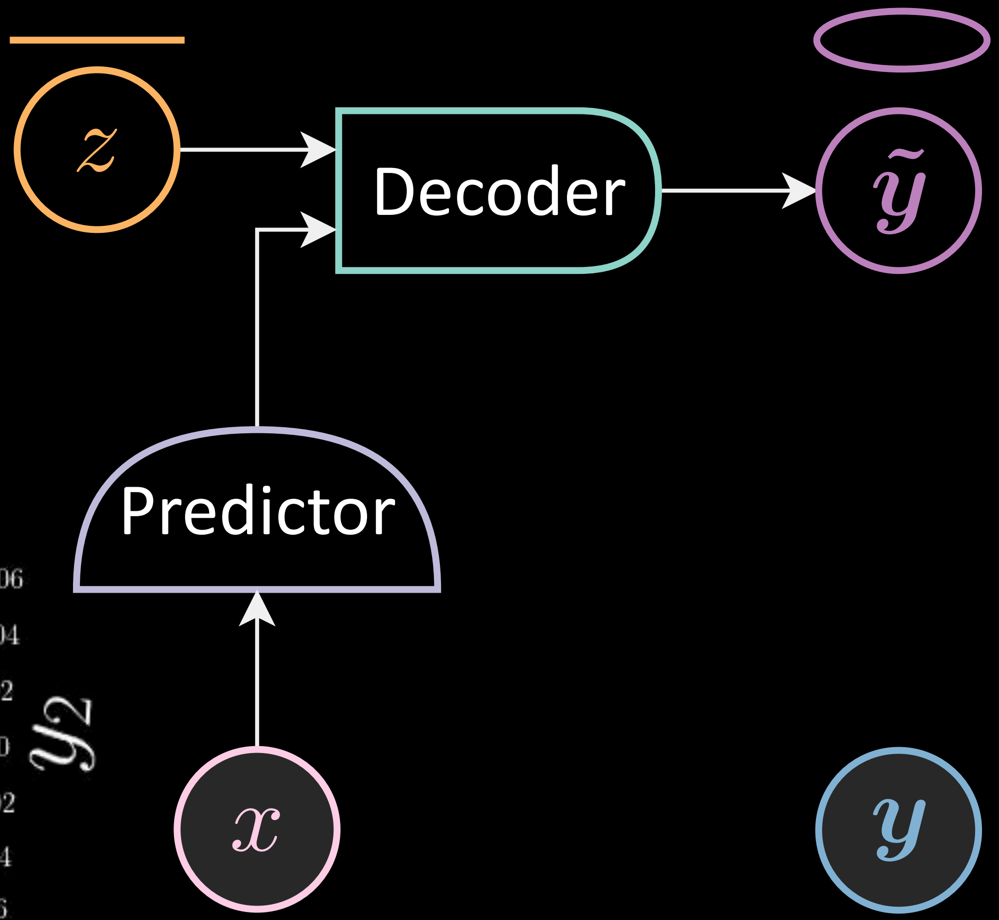
$$x \sim \mathcal{U}(0,1)$$

$$\theta \sim \mathcal{U}(0,2\pi)$$

$$\varepsilon \sim \mathcal{N}\left[0, \left(\tfrac{1}{20}\right)^2\right]$$

Untrained model manifold

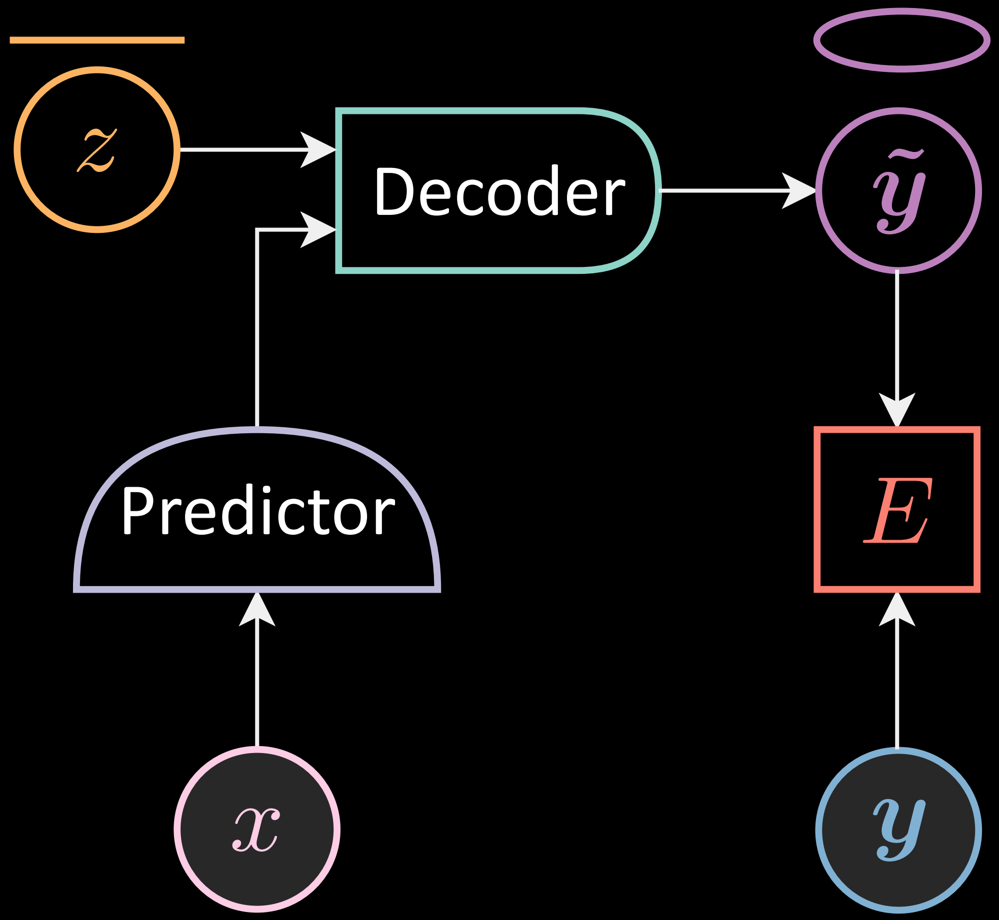




$$z = \left[0 : \frac{\pi}{24} : 2\pi\right[$$

$$x = \left[0 : \frac{1}{50} : 1\right]$$

Energy function



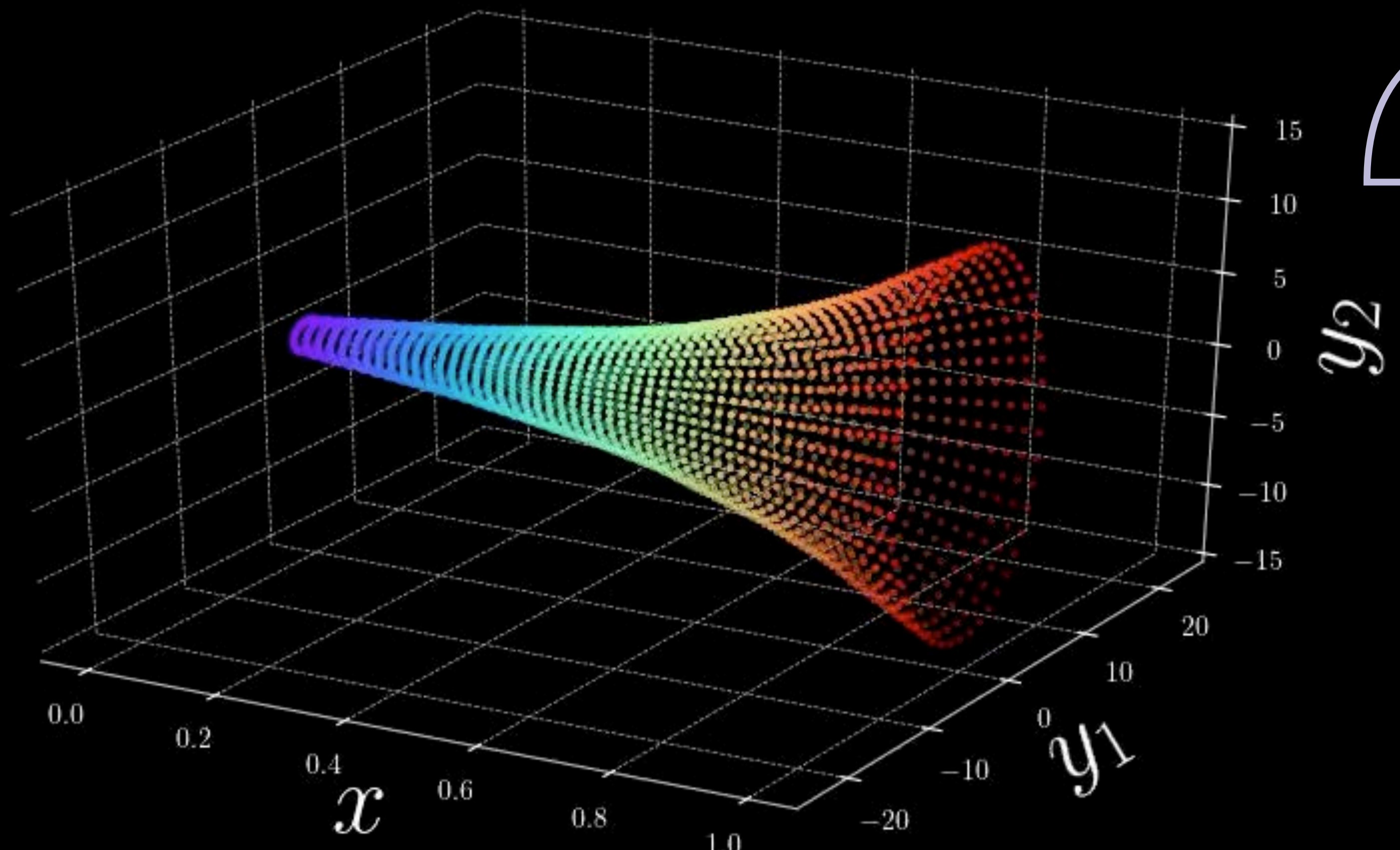
$$f, g : \mathbb{R} \to \mathbb{R}^2$$

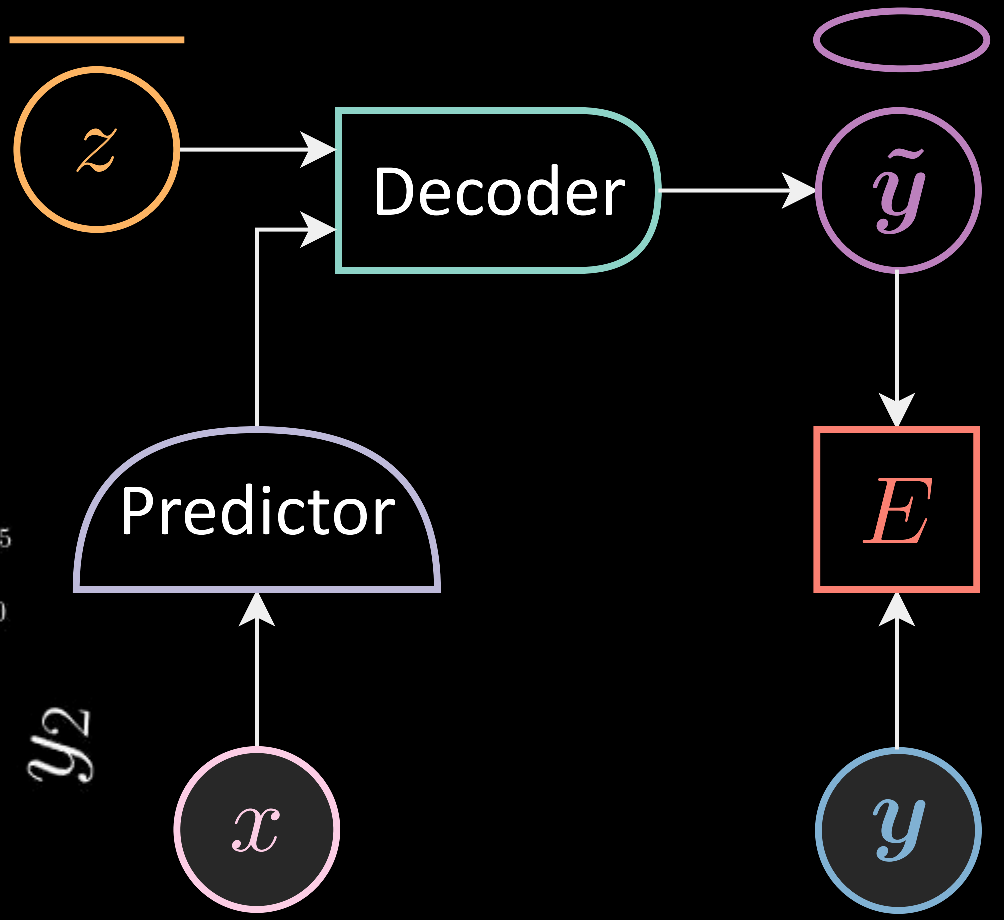
$$x \overset{f}{\mapsto} x \xrightarrow{\mathtt{L}^{+}} 8 \xrightarrow{\mathtt{L}^{+}} 8 \xrightarrow{\mathtt{L}} 2$$

$$z \overset{g}{\mapsto} \begin{bmatrix} \cos(z) & \sin(z) \end{bmatrix}^\top$$

$$E(x, \boldsymbol{y}, z) = [y_1 - f_1(x) g_1(z)]^2 + [y_2 - f_2(x) g_2(z)]^2$$

Trained model manifold

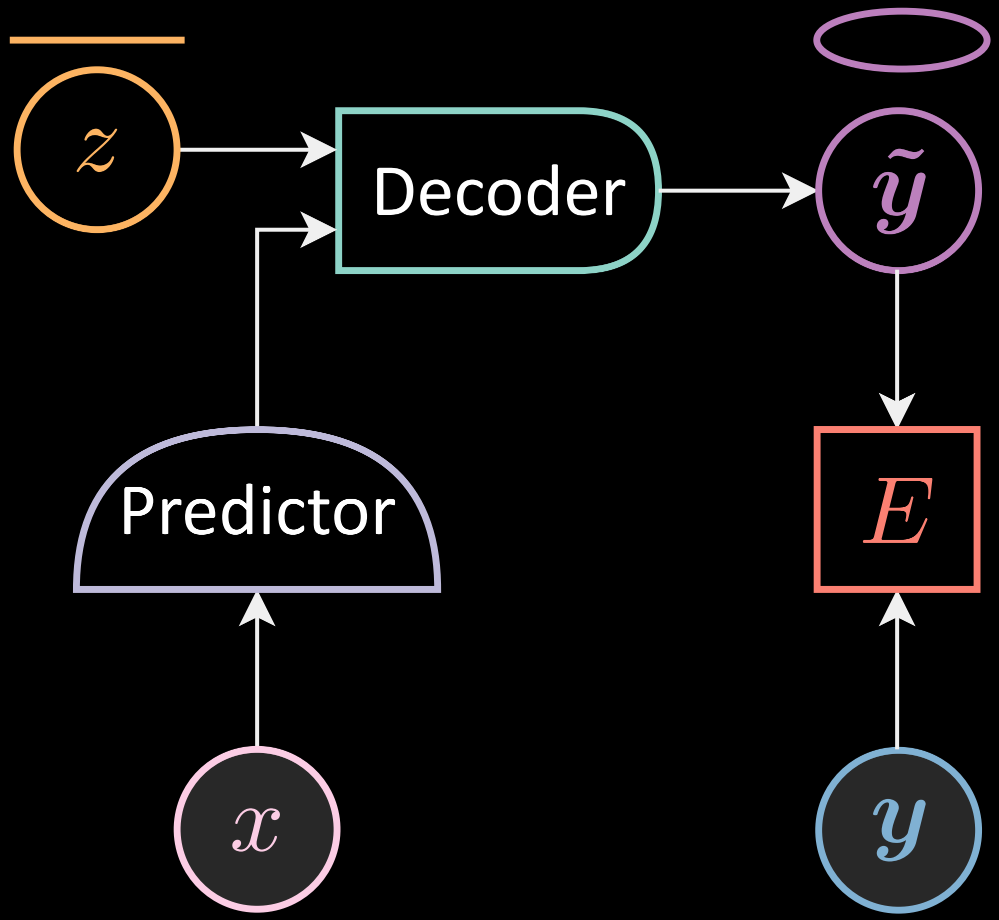




$$z = \left[0 : \frac{\pi}{24} : 2\pi\right[$$

$$x = \left[0 : \frac{1}{50} : 1\right]$$

Energy function (II)

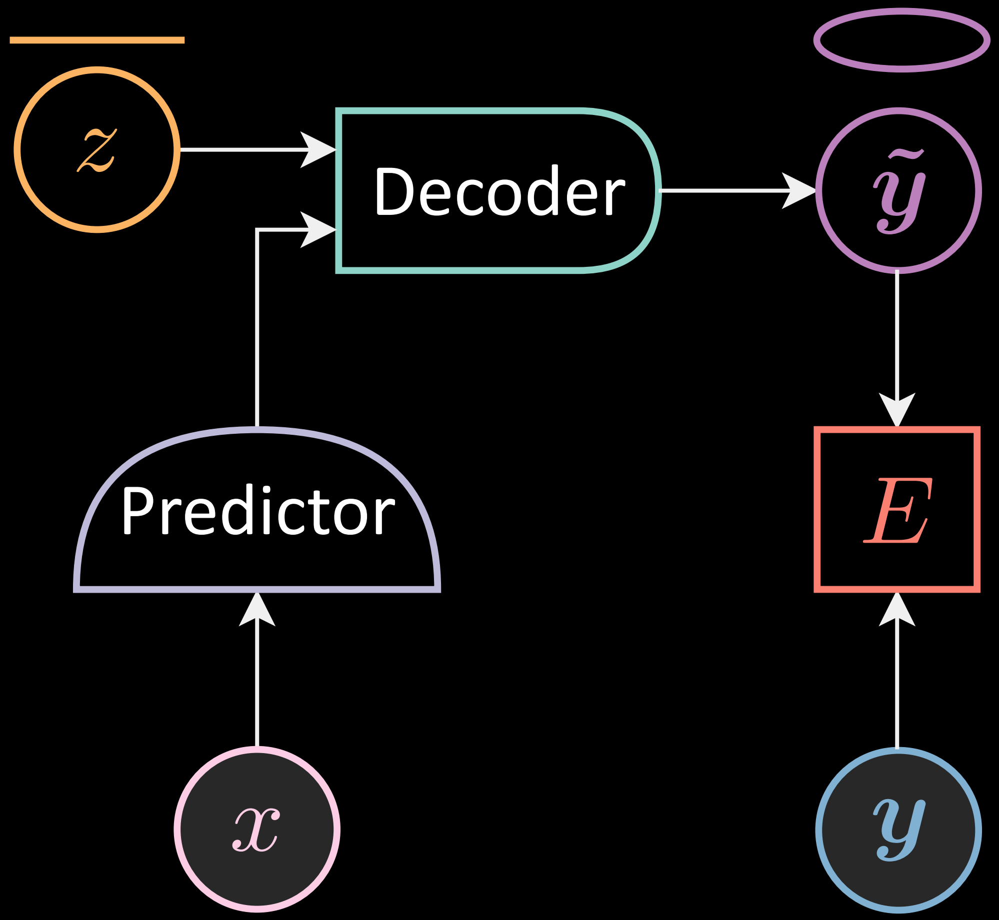


$$f : \mathbb{R} \to \mathbb{R}^{\dim(f)}$$

$$g : \mathbb{R}^{\dim(f)} \times \mathbb{R} \to \mathbb{R}^2$$

$$E(x, \boldsymbol{y}, z) = [y_1 - g_1(f(x), z)]^2 + [y_2 - g_2(f(x), z)]^2$$

Energy function (III)



$$f : \mathbb{R} \to \mathbb{R}^{\dim(f)}$$

$$g : \mathbb{R}^{\dim(f)} \times \mathbb{R}^{\dim(\boldsymbol{z})} \to \mathbb{R}^2$$

$$E(x, \boldsymbol{y}, \boldsymbol{z}) = [y_1 - g_1(f(x), \boldsymbol{z})]^2 + [y_2 - g_2(f(x), \boldsymbol{z})]^2$$