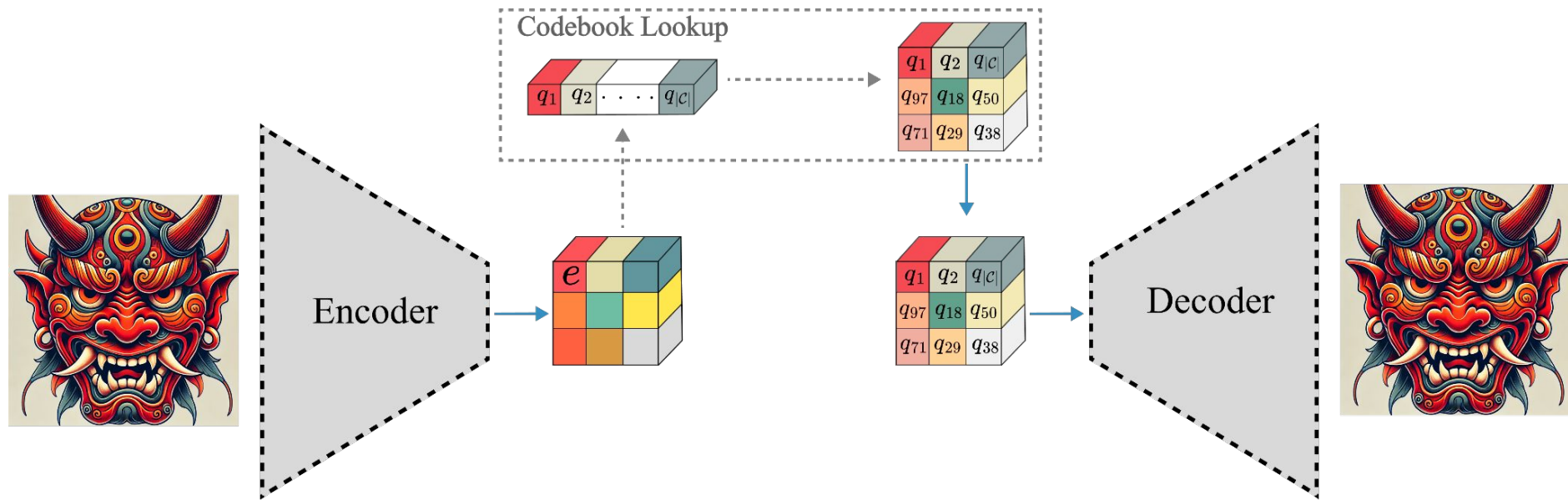


Restructuring Vector Quantization with the Rotation Trick

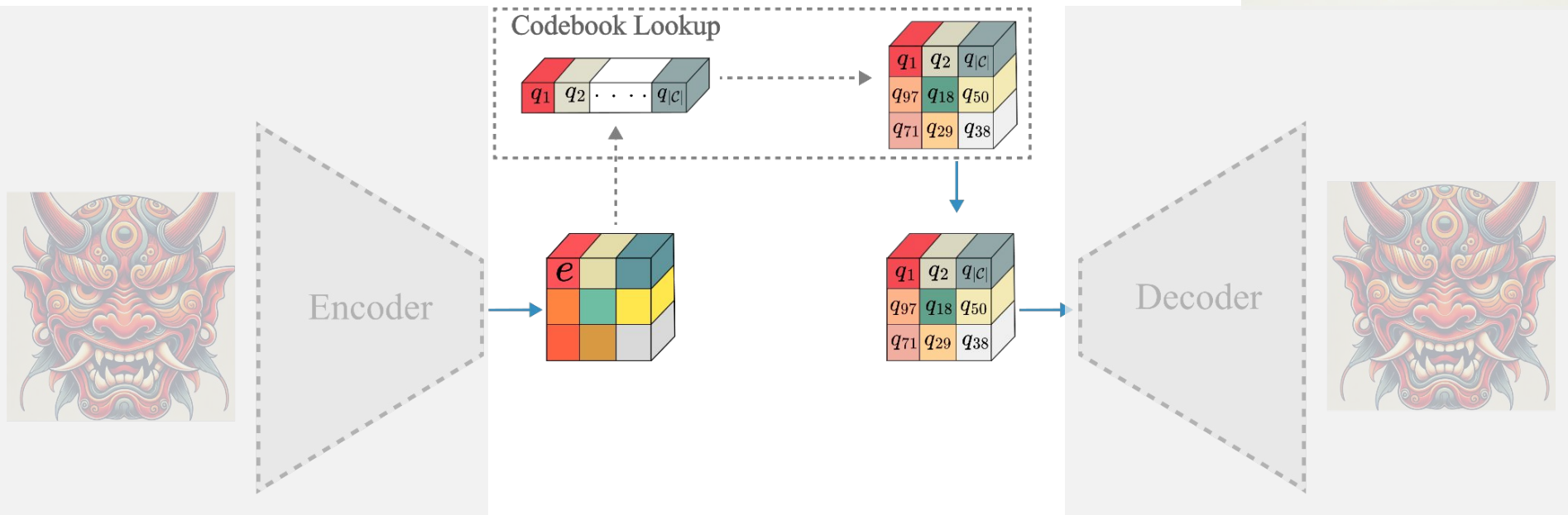
Christopher Fifty, Ronald G Junkins, Dennis Duan, Aniketh
Iyengar, Jerry W. Liu, Ehsan Amid, Sebastian Thrun,
Christopher Ré

Background | Vector Quantized Variational AutoEncoders VQ-VAEs

AutoEncoder that compresses to a set of discrete tokens (codewords)



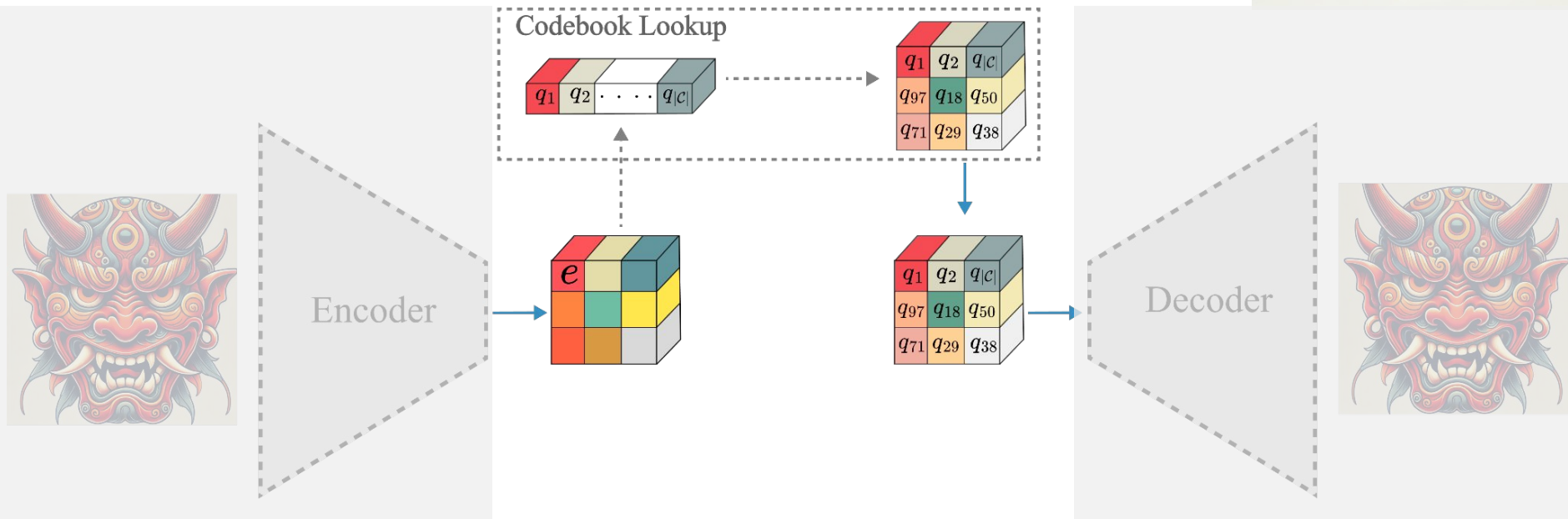
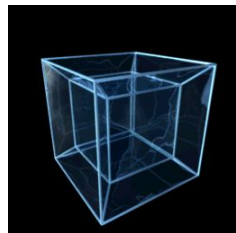
Motivation | Circles and Squares



Motivation | Circles and Squares

Finite Scalar
Quantization

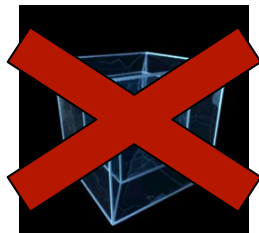
ICLR 2024



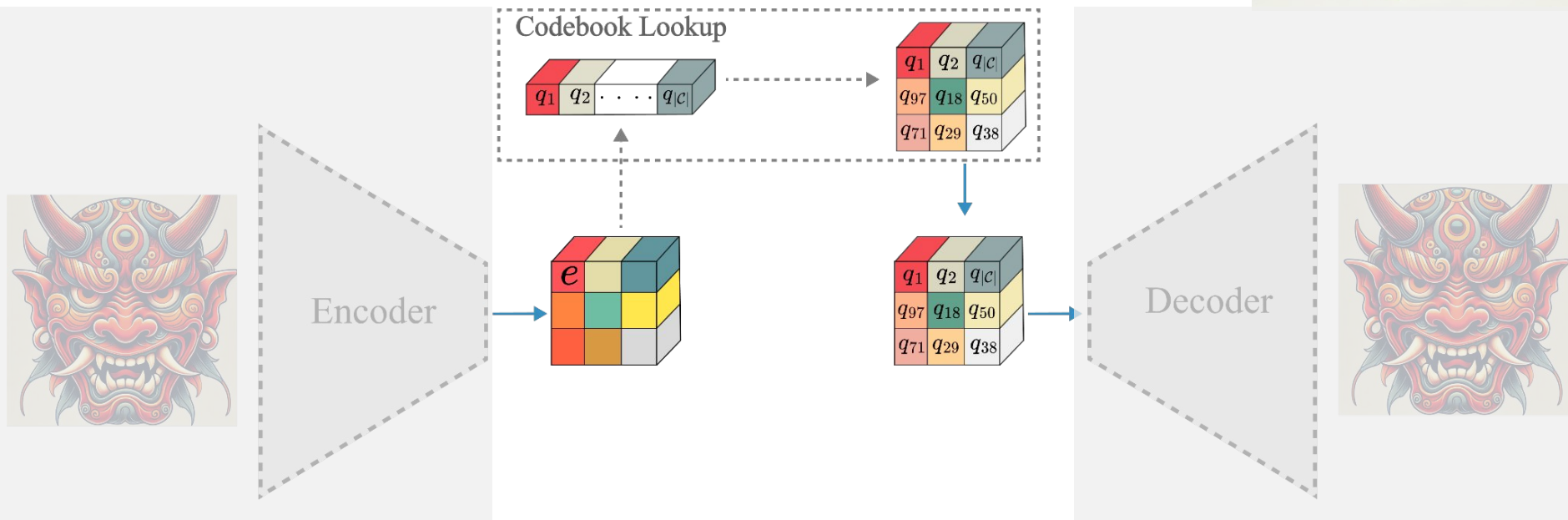
Motivation | Circles and Squares

Finite Scalar
Quantization

ICLR 2024

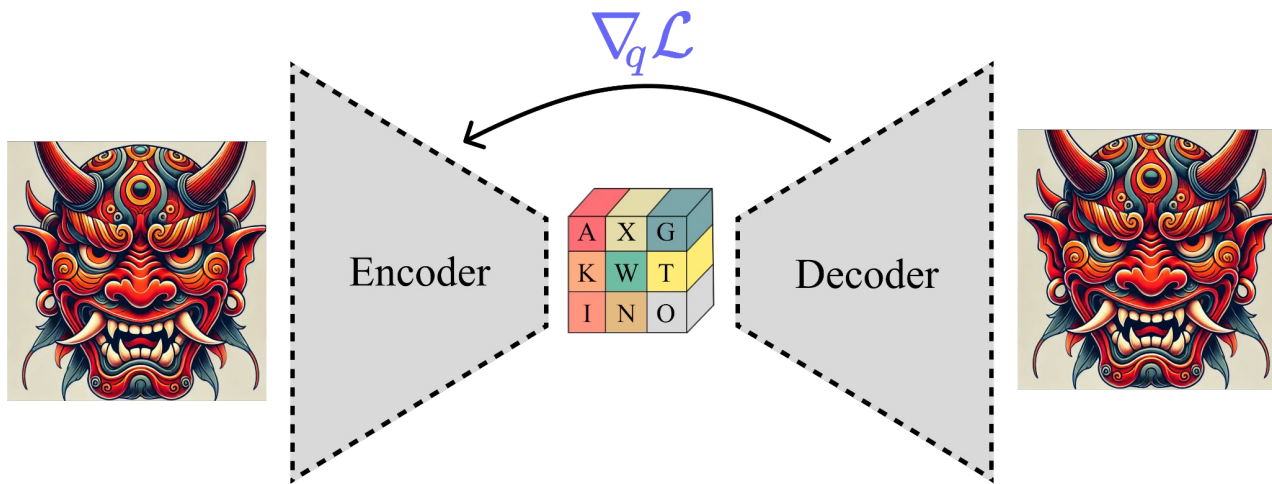


Spherical
Codes
Quantization



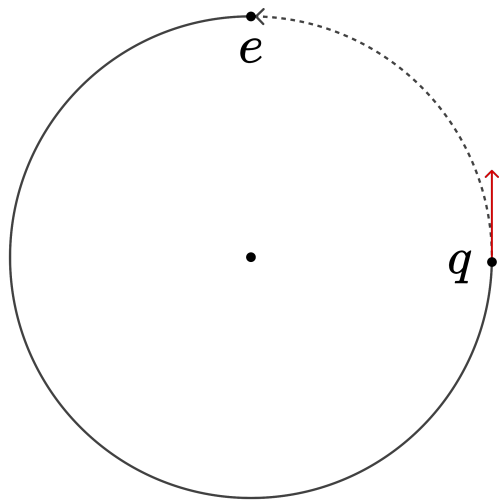
Question

How to get the gradient back to the encoder?



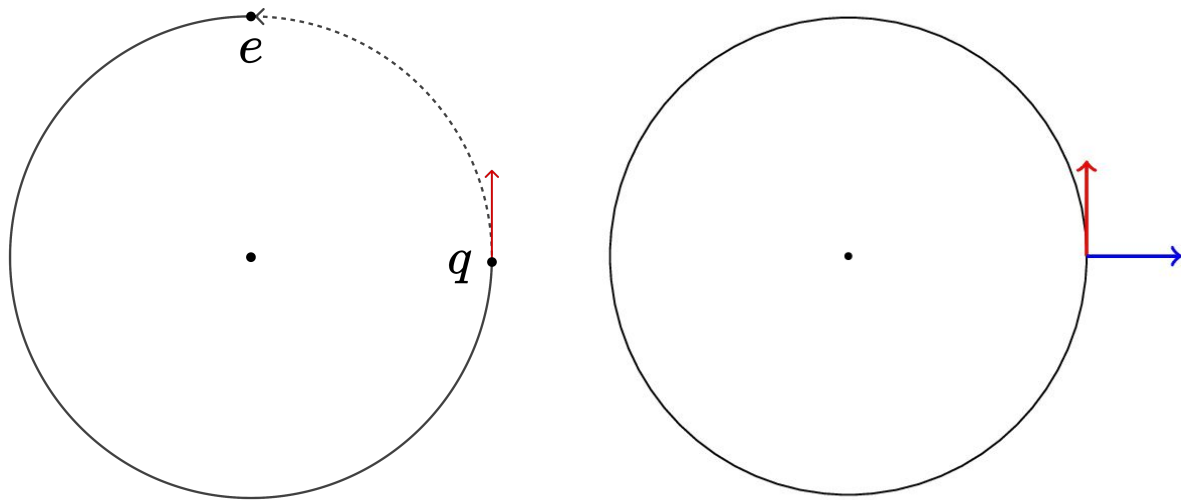
Question

How do you move vectors on a spherical surface?



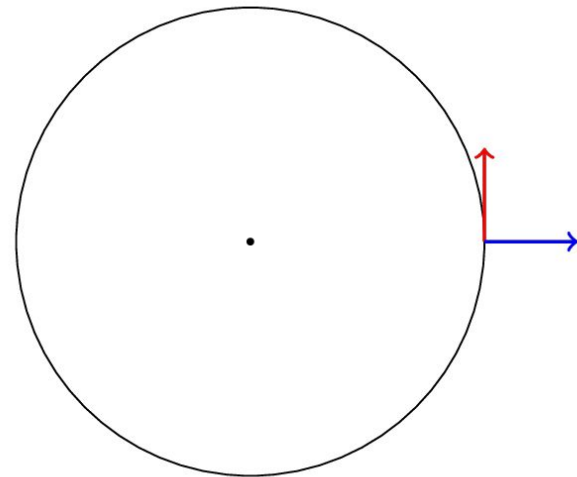
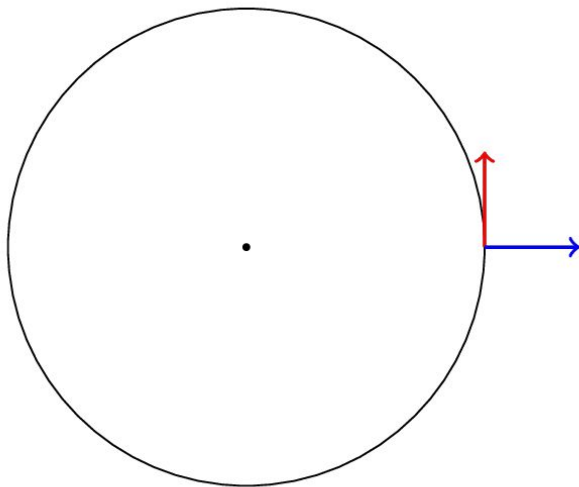
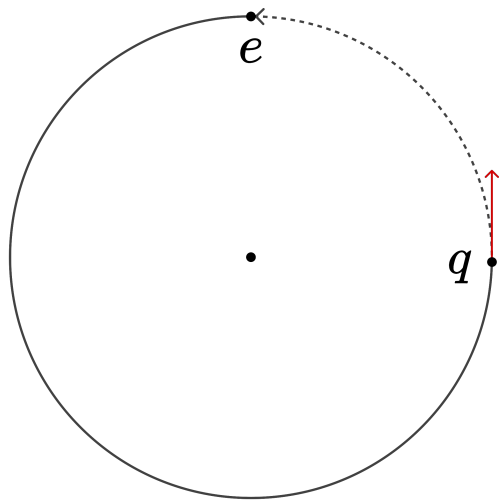
Question

How do you move vectors on a spherical surface?



Question

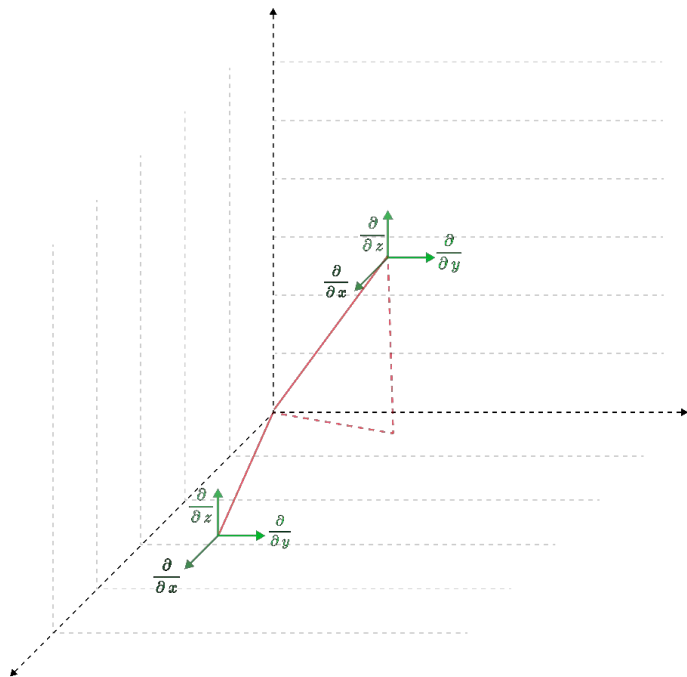
How do you move vectors on a spherical surface?



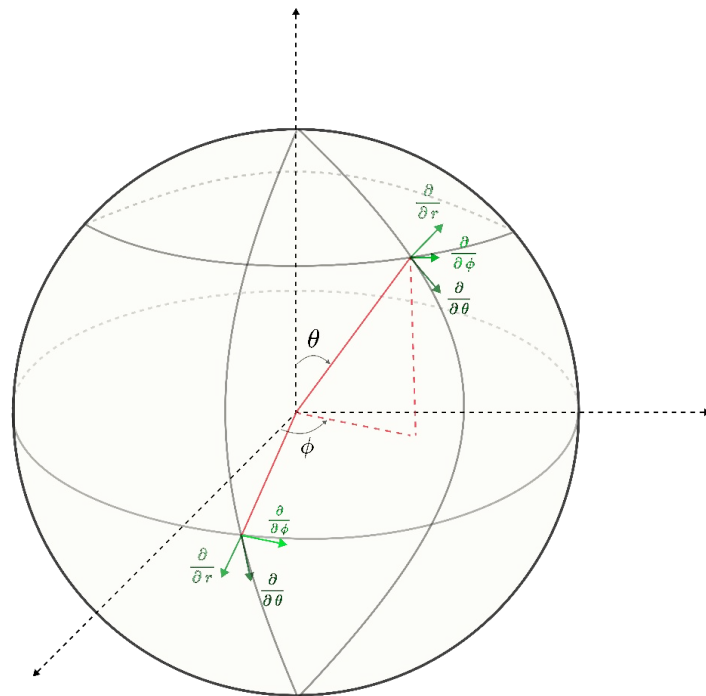
Question

How do you move vectors on a spherical surface?

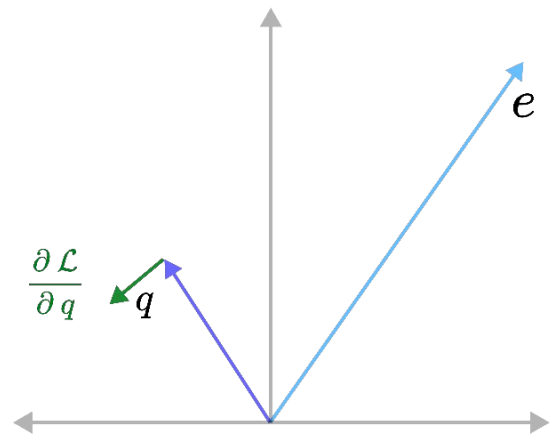
Cartesian Coordinate System



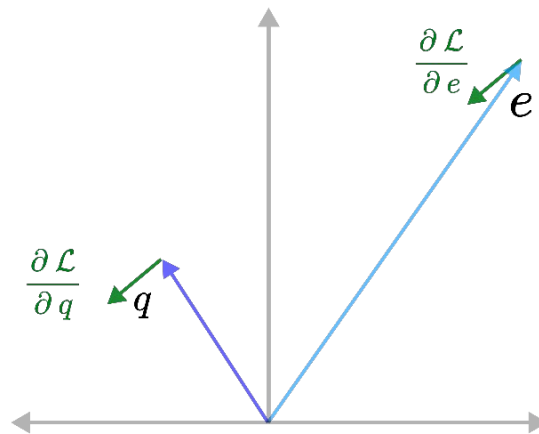
Spherical Coordinate System



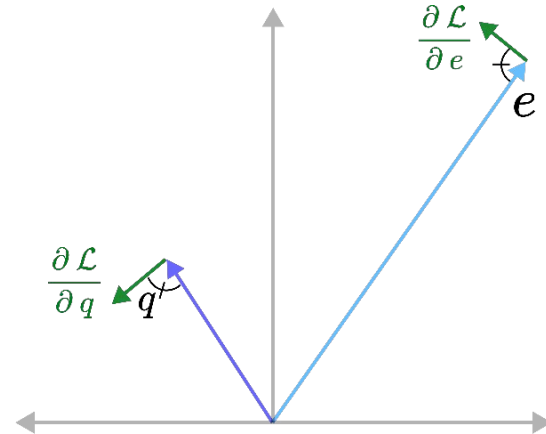
Question | Is there a “better” way to move the gradient from q to e ?



Straight-Through
Estimator (STE)



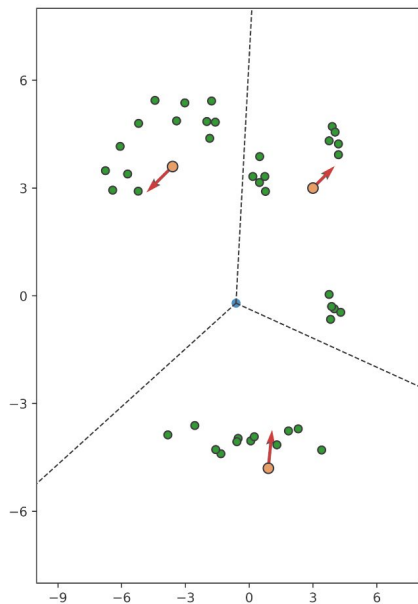
Rotation Trick



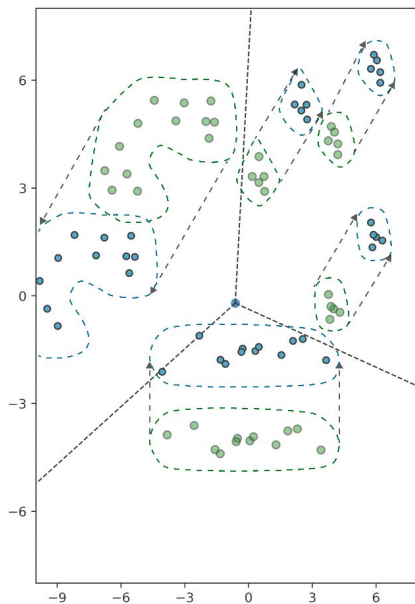
Question | Is there a “better” way to move the gradient from q to e?

Compression Theory: high capacity, low distortion

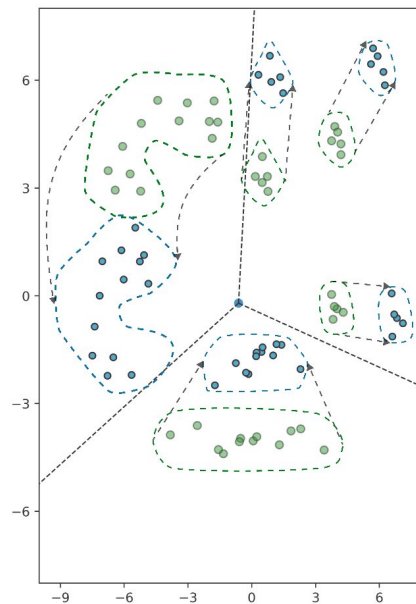
Voronoi Partition



STE Updates



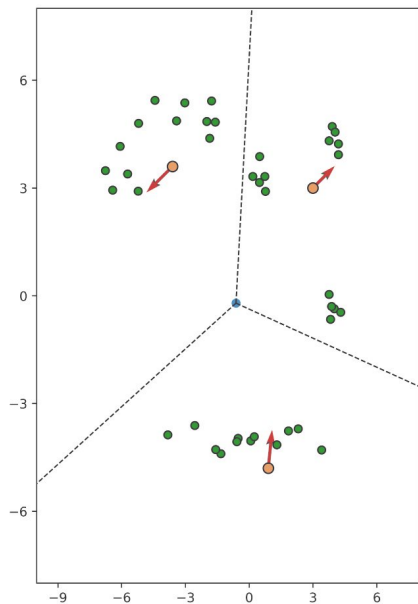
Rotation Trick Updates



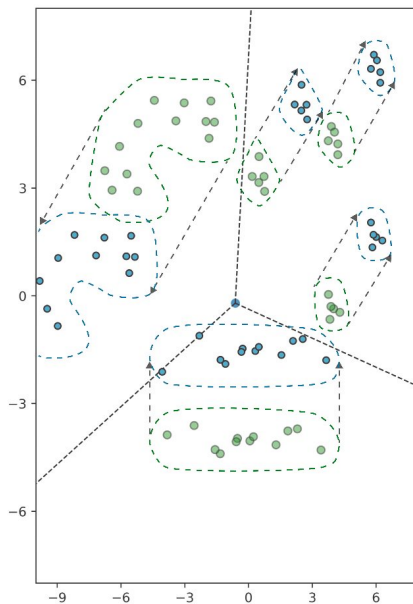
Question | Is there a “better” way to move the gradient from q to e ?

Compression Theory: high capacity, low distortion

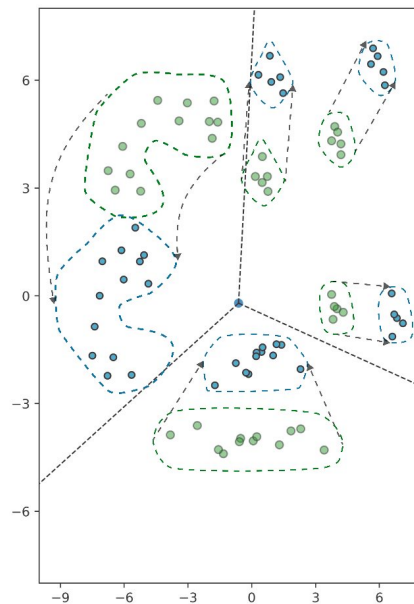
Voronoi Partition



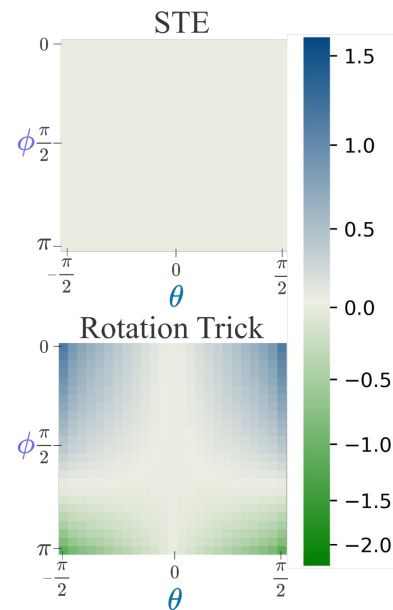
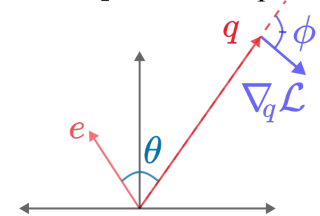
STE Updates



Rotation Trick Updates



Change in Distance Between e and q After an Update



Algorithm | Simply rotate e to q in forward pass

Algorithm 1 The Rotation Trick

Require: input example x

$e \leftarrow \text{Encoder}(x)$

$q \leftarrow$ nearest codebook vector to e

$R \leftarrow$ rotation matrix that aligns e to q

$\tilde{q} \leftarrow \text{stop-gradient} \left[\frac{\|q\|}{\|e\|} R \right] e$

$\tilde{x} \leftarrow \text{Decoder}(\tilde{q})$

$\text{loss} \leftarrow \mathcal{L}(x, \tilde{x})$

return loss

Results | Improved Image Reconstructions

Approach	Latent Shape	Codebook Size	Codebook Usage	Quantization Error ($\downarrow$)	r-FID ($\downarrow$)	r-IS ($\uparrow$)
VQGAN	$64 \times 64 \times 3$	8192	15%	2.5e-3	0.53	220.6
Gumbel VQGAN	$64 \times 64 \times 3$	8192	4%	—	0.60	219.7
VQGAN w/ Rotation Trick	$64 \times 64 \times 3$	8192	86%	1.7e-4	0.27	228.0
VQGAN	$32 \times 32 \times 4$	16384	2%	1.2e-2	5.0	141.5
Gumbel VQGAN	$32 \times 32 \times 4$	16384	12%	—	1.7	189.5
VQGAN w/ Rotation Trick	$32 \times 32 \times 4$	16384	27%	2.4e-4	1.1	200.2

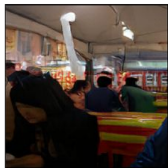
Orig



STE



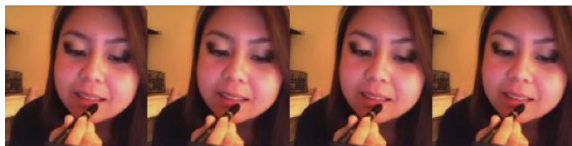
ROT



Results | Enables new applications in video modeling

Approach	Dataset	Codebook Usage	Train Loss ($\downarrow$)	Quantization Error ($\downarrow$)	Valid Loss ($\downarrow$)	r-FVD ($\downarrow$)
TimeSformer [†]	BAIR	0.4%	0.221	0.03	0.28	1661.1
TimeSformer w/ Rotation Trick	BAIR	43%	0.074	3.0e-3	0.074	21.4
TimeSformer [†]	UCF-101	0.1%	0.190	0.006	0.169	2878.1
TimeSformer w/ Rotation Trick	UCF-101	30%	0.111	0.020	0.109	229.1

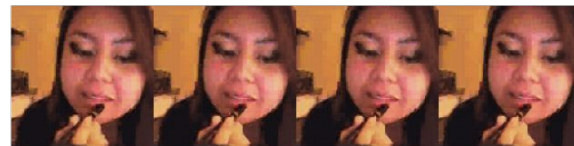
Original Video



STE
Reconstructions



Rotation Trick
Reconstructions



Acknowledgements



Ronald
Junkins



Dennis
Duan



Aniketh
Iyengar



Jerry
Liu



Ehsan
Amid



Sebastian
Thrun



Christopher
Ré

Acknowledgements



CITADEL | Securities



Ronald
Junkins



Dennis
Duan



Aniketh
Iyengar



Jerry
Liu



Ehsan
Amid



Sebastian
Thrun



Christopher
Ré

fifty@cs.stanford.edu

Acknowledgements



CITADEL | Securities



Ronald
Junkins



Dennis
Duan



Aniketh
Iyengar



Jerry
Liu



Ehsan
Amid



Sebastian
Thrun



Christopher
Ré

[https://github.com/lucidrains
/vector-quantize-pytorch](https://github.com/lucidrains/vector-quantize-pytorch)

fifty@cs.stanford.edu