



Figure 1: Illustration of the equivalence between the binary constraint and the intersection of solid cube and ℓ_2 norm sphere.

Table 1: Results of STG, FIRDL, and our method in downstream tasks.

Task	Datasets	STG	FIRDL	Our
Classification (ACC \uparrow)	Madelon	0.82 (15)	0.54 (15)	0.90 (20)
	PCMAC	0.91 (300)	0.72 (300)	0.79 (300)
	COIL-20	1.00 (150)	1.00 (200)	1.00 (200)
	Yale	0.75 (200)	0.76 (150)	0.81 (300)
	Jaffe	0.98 (200)	0.98 (75)	1.00 (150)
	PIX10	0.98 (25)	0.99 (150)	1.00 (25)
	warpPIE10P	0.97 (200)	0.98 (200)	0.99 (300)
	GLIOMA	0.77 (200)	0.76 (75)	0.81 (300)
	LUNG	0.90 (150)	0.90 (300)	0.94 (50)
	PROSTATE	0.85 (200)	0.85 (150)	0.89 (300)
	SRBCT	0.97 (200)	0.95 (200)	0.98 (150)
	SMK	0.68 (100)	0.68 (150)	0.75 (200)
	Average rank	2.1	2.3	1.1
	Top-1 counts	2	1	11
Clustering (ACC \uparrow)	Madelon	0.60 (5)	0.53 (15)	0.60 (20)
	PCMAC	0.52 (25)	0.52 (75)	0.53 (200)
	COIL-20	0.70 (300)	0.62 (300)	0.66 (200)
	Yale	0.61 (200)	0.62 (300)	0.62 (300)
	Jaffe	0.81 (75)	0.84 (50)	0.87 (200)
	PIX10	0.82 (200)	0.83 (50)	0.87 (300)
	warpPIE10P	0.45 (75)	0.45 (50)	0.51 (25)
	GLIOMA	0.66 (25)	0.65 (50)	0.75 (75)
	LUNG	0.69 (200)	0.69 (200)	0.72 (150)
	PROSTATE	0.64 (100)	0.65 (75)	0.68 (50)
	SRBCT	0.58 (300)	0.56 (100)	0.63 (50)
	SMK	0.60 (50)	0.61 (300)	0.64 (25)
	Average rank	2.3	2.3	1.1
	Top-1 counts	2	1	11
Reconstruction (RMSE \downarrow)	Madelon	0.97 (20)	0.99 (20)	0.98 (10)
	PCMAC	1.04 (25)	1.01 (25)	1.30 (25)
	COIL-20	0.41 (300)	0.40 (300)	0.38 (300)
	Yale	0.55 (300)	0.56 (300)	0.54 (300)
	Jaffe	0.26 (300)	0.26 (300)	0.22 (300)
	PIX10	0.42 (300)	0.43 (300)	0.39 (300)
	warpPIE10P	0.26 (300)	0.26 (300)	0.25 (300)
	GLIOMA	0.69 (300)	0.69 (300)	0.69 (300)
	LUNG	0.78 (300)	0.78 (300)	0.80 (300)
	PROSTATE	0.72 (300)	0.72 (300)	0.73 (300)
	SRBCT	0.77 (300)	0.77 (300)	0.75 (300)
	SMK	0.68 (300)	0.68 (300)	0.69 (300)
	Average rank	1.7	1.8	1.8
	Top-1 counts	5	5	7