Table R1: Summary of average run time in seconds (s) or minutes (m) and number of iterations of different methods on different datasets. We use Algorithm 1 for the synthetic experiments, and Algorithm 3 for the other two experiments.

Datasets	Metrics	LS	PMTL	EPO	FERERO
Synthetic, Figures 3(a-c)	Iterations	100	100	60	10
	Per-iteration run time	3.50E-4s	7.67E-3s	4.93E-3s	7.50E-3s
	Total run time	0.035s	0.767s	0.296s	0.075s
Synthetic, Figures 3(d-f)	Iterations	100	200	80	200
	Per-iteration run time	3.10E-4s	7.65E-3s	4.93E-3s	7.30E-3s
	Total run time	0.031s	0.153s	0.394s	0.146s
Multi-MNIST/Fashion/F+M	Epochs	100	100	100	100
	Per-epoch run time	3.54s	11.88s	9.66s	7.02s
	Total run time	5.9m	19.8m	16.1m	11.7m
Emotion	Epochs	200	200	200	200
	Per-epoch run time	9.5E-3s	0.496s	0.238s	0.039s
	Total run time	1.9s	99.1s	47.6s	7.70s

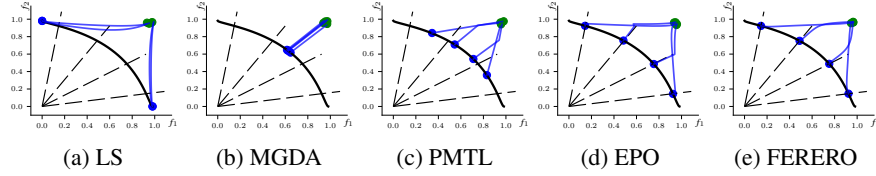(a) LS   (b) MGDA   (c) PMTL   (d) EPO   (e) FERERO

Figure R1: Revision of Figure 2 (a-e) by adding the green dots. Converging solutions (blue dots) and optimization trajectories (blue lines) on the objective space of different methods on synthetic objectives given in (5.1). Dashed arrows represent pre-specified preference vectors. The green dots represent initial objective values.

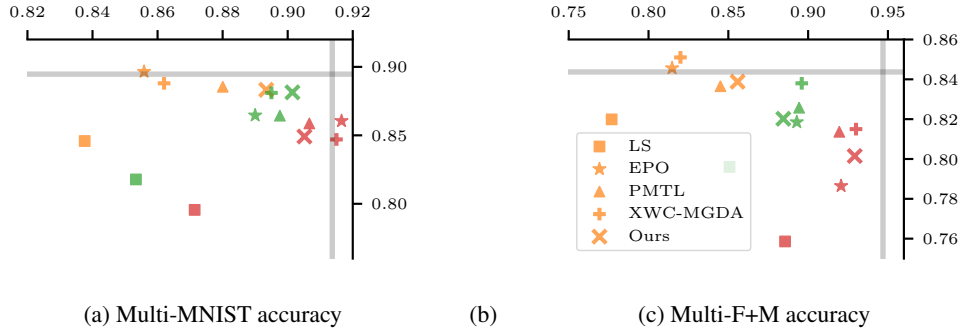(a) Multi-MNIST accuracy   (b)   (c) Multi-F+M accuracy

Figure R2: Revision of Figure 4 (a,c) by changing the range. Training losses and accuracies of various methods with different preferences across three image datasets. The horizontal and vertical axes represent results for objective 1 and objective 2, respectively. Different colored dashed arrows indicate various preference vectors. Different markers denote the solutions obtained by different methods, with marker colors matching the preferences.