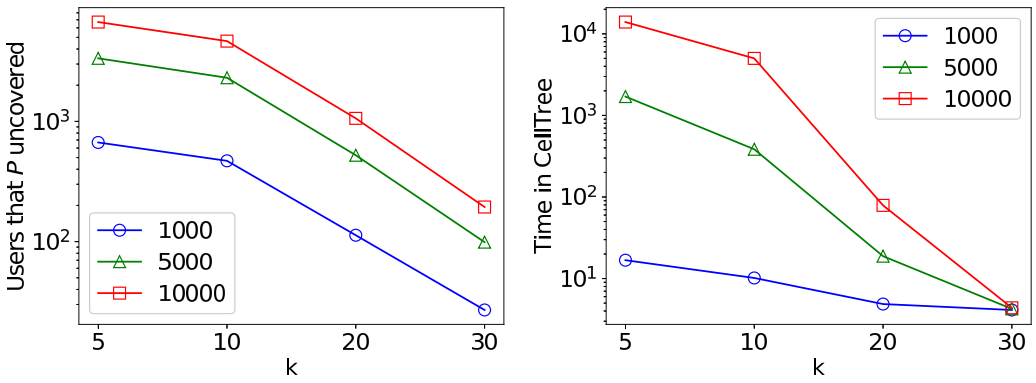


(a) User uncovered by P **(b)** Time spent in $CellTree$

Figure 6.10: Effect of k (HOTEL)

dataset P already covers a part of the users. With the increasing k, unchanged product dataset and user dataset, the products in dataset P are to cover more users easily because the the threshold to be a user's top-k is lowering down. With P covering more users, no matter which of the lemmas we propose or the baseline $CellTree$ approach will all greatly benefit from the reducing users. We show the experiment in Figure 6.10. The y axis of Figure 6.10a means the users number that we need to handle after removing the users that covered by P. We can see that user dataset with size 1000, 5000 or 10000, their uncovered users decrease exponentially with the change of k. And because of the decreasing uncovered users, our response time also decreases exponentially.

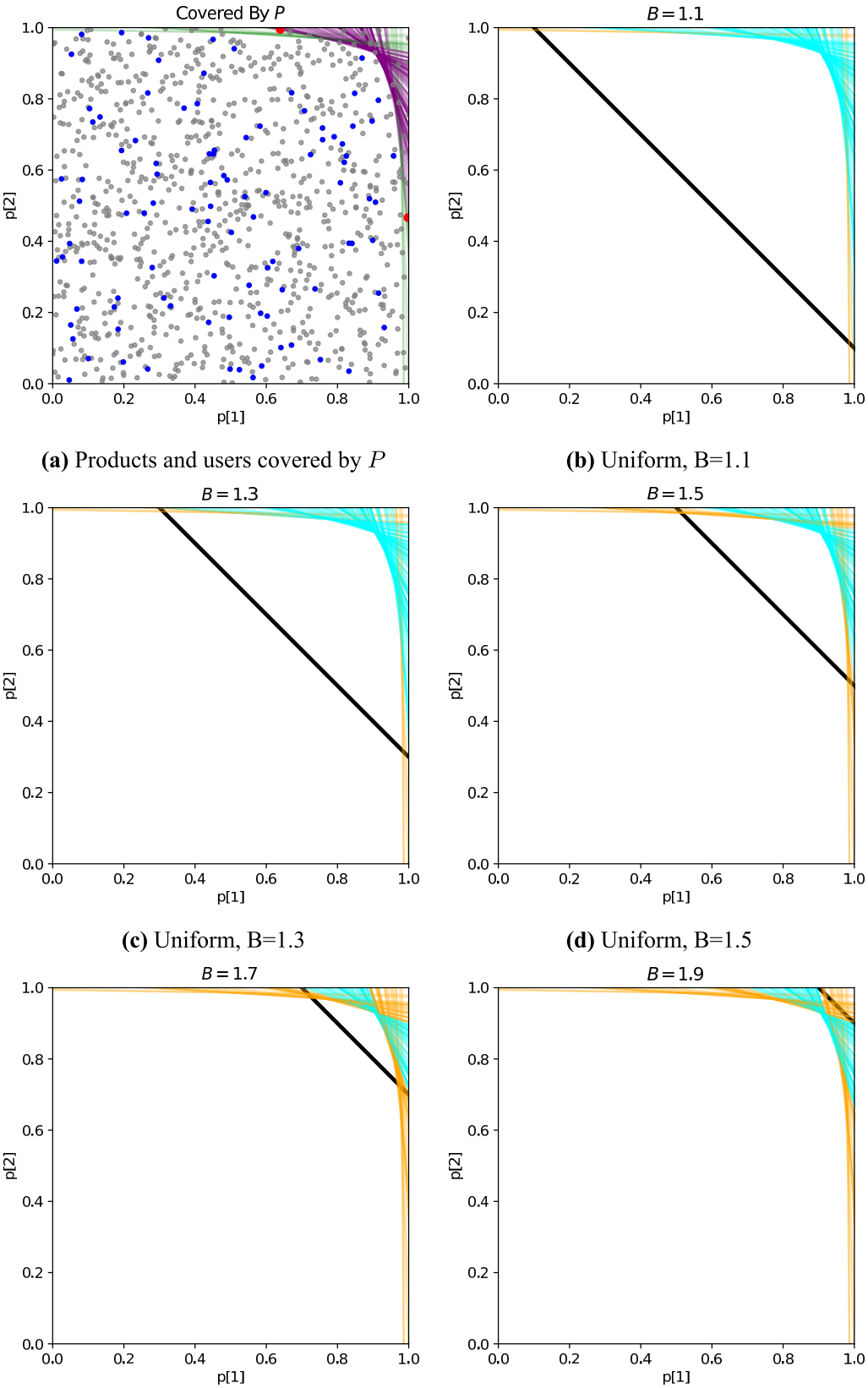
6.4 Effect of Product Dataset Distribution and B

To straight forward presents our algorithm and shows the interesting finding in experiments, we use a 2d experiment and visualize it as in Figure 6.11, 6.12 and 6.13.

First of all, the users are all generated uniformly from $w[1] + w[2] = 1$.

To explain Figure 6.11, 6.12 and 6.13, we take 6.11a as an example, this figure is drawn in product space. Product dataset D consists with the grey, blue and red points ; product dataset P consists with the blue and red points; the red points are the points that each of them at least covers one user. Specially, we have to expand the size of red points to clearly show them. We can see the red points of Figure 6.11a at about $(0.9, 0.9)$ and $(0.95, 0.8)$. The lines are all $w_i\dot{p} = S_{ik}$. The blue lines mean the corresponding $w_i\dot{p} = S_{ik}$ of users that covered by P and the rest users' halfspaces are orange lines. From Figure 6.11b to Figure 6.11f, the bold black line means the constraint $\Sigma_{i=1}^{d} p[i] \leq B$ when B equals 1.1, 1.3, 1.5, 1.7, 1.9 respectively. The orange lines mean the $w_i\dot{p} = S_{ik}$ that intersects with constraint and the blues aren't intersect with constraint.

In Figure 6.11a, because the distribution of products is uniform so for each user weight vector w_i, there is a proper product p_{wi} nearby the extend of w_i ranks top-k for it. And



(a) Products and users covered by P

(b) Uniform, B=1.1

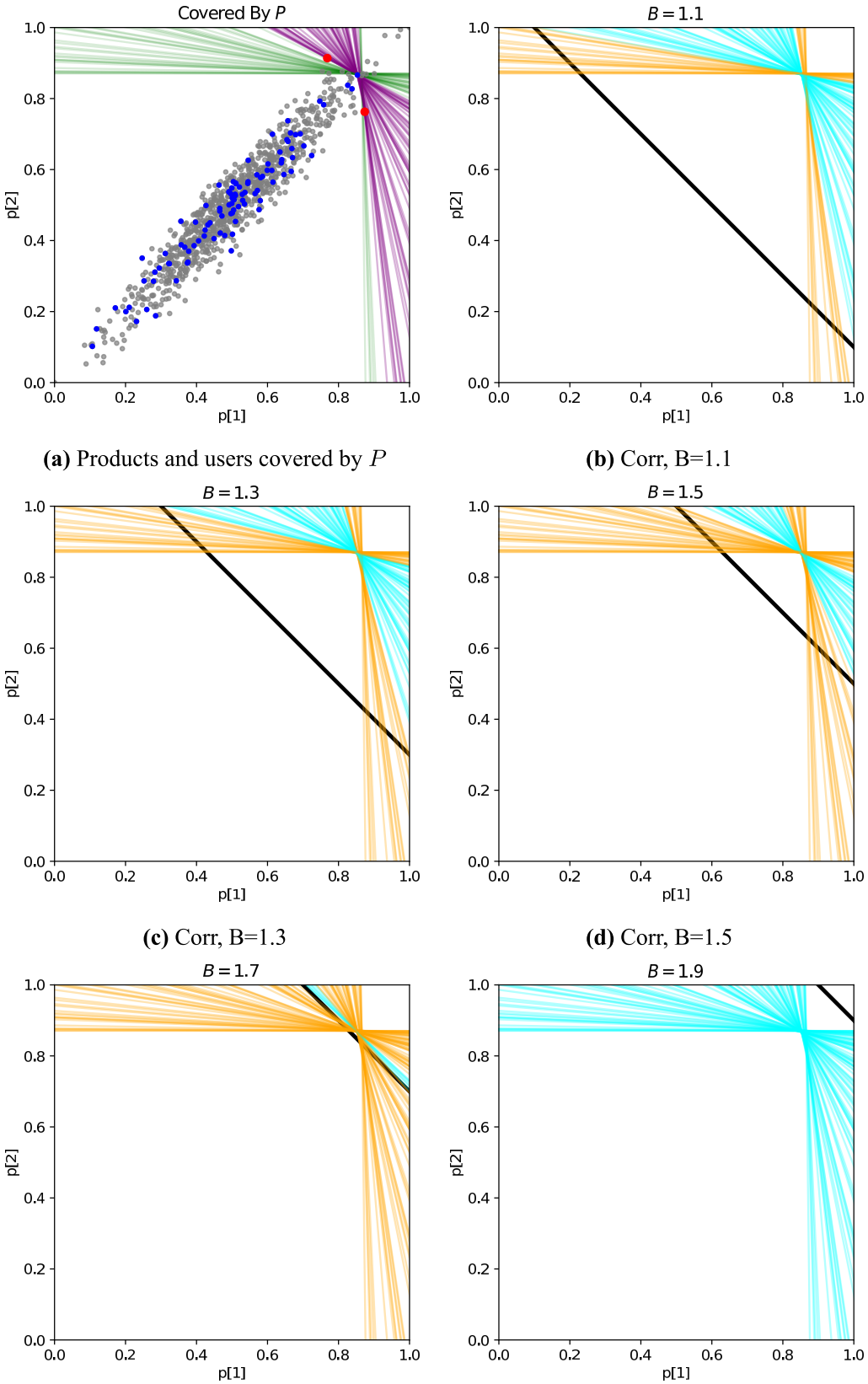
(c) Uniform, B=1.3

(d) Uniform, B=1.5

(e) Uniform, B=1.7

(f) Uniform, B=1.9

Figure 6.11: Effect of B on uniform products



(a) Products and users covered by P

(b) Corr, B=1.1

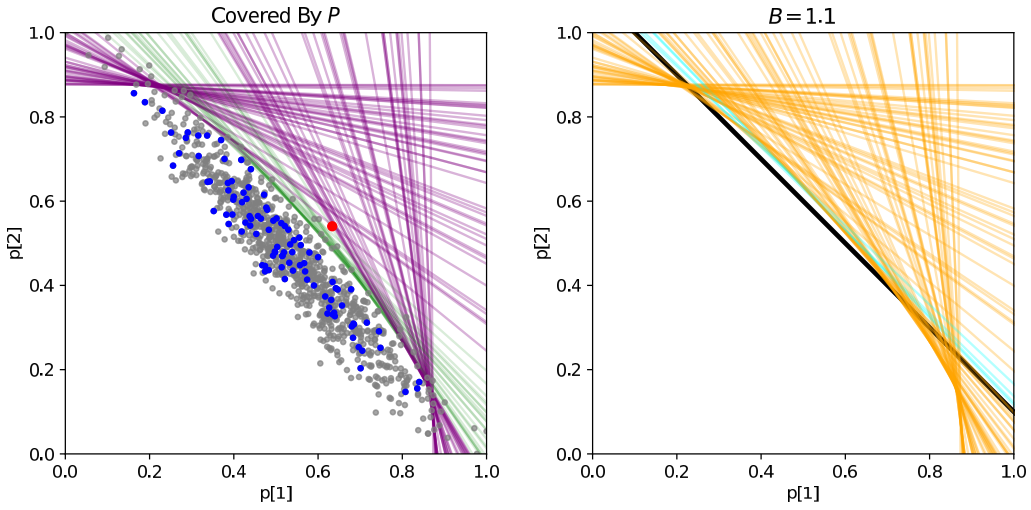
(c) Corr, B=1.3

(d) Corr, B=1.5

(e) Corr, B=1.7

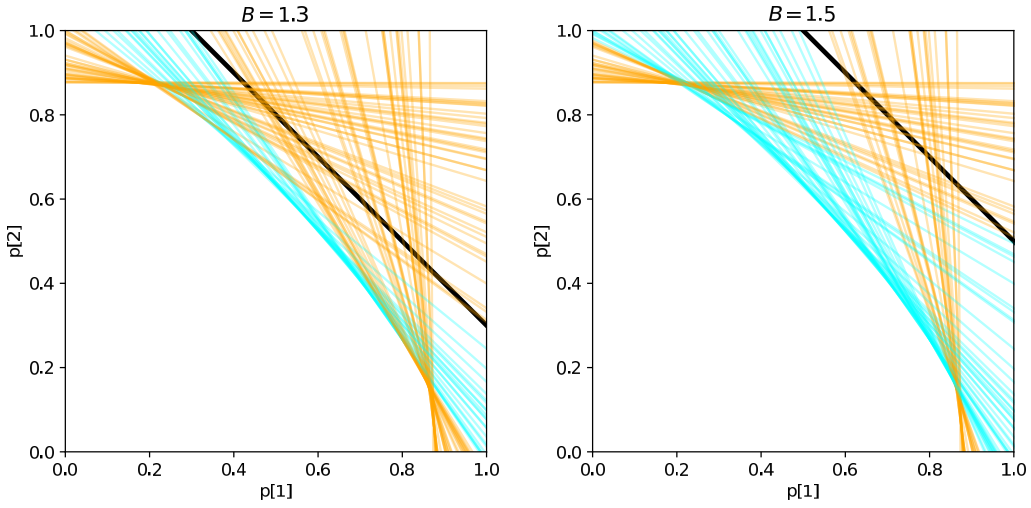
(f) Corr, B=1.9

Figure 6.12: Effect of B on corr products



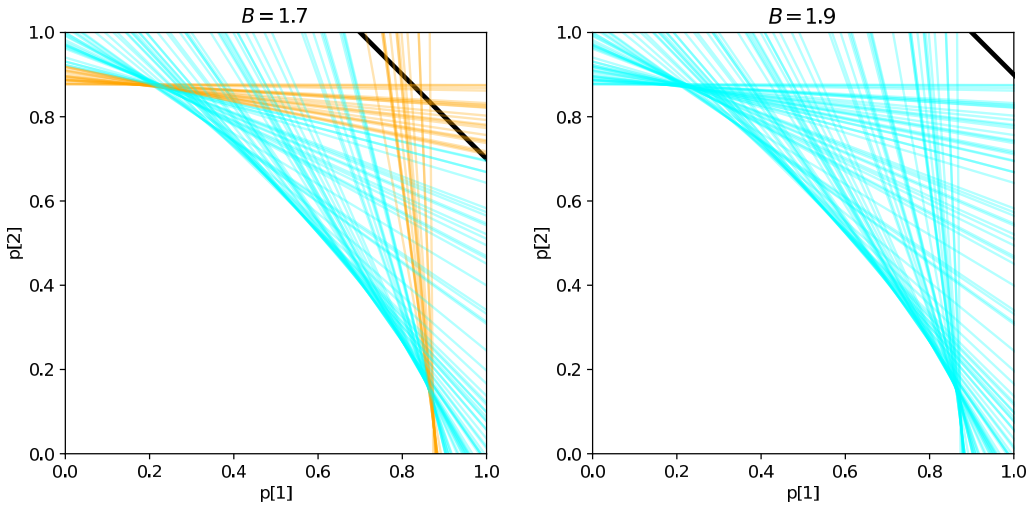
(a) Products and users covered by P

(b) Anti, B=1.1



(c) Anti, B=1.3

(d) Anti, B=1.5



(e) Anti, B=1.7

(f) Anti, B=1.9

Figure 6.13: Effect of B on anti products