
Feasible Recourse Plan via Diverse Interpolation

Duy Nguyen
VinAI Research, Vietnam

Ngoc Bui
VinAI Research, Vietnam

Viet Anh Nguyen
The Chinese University of Hong Kong

Abstract

Explaining algorithmic decisions and recommending actionable feedback is increasingly important for machine learning applications. Recently, significant efforts have been invested in finding a diverse set of recourses to cover the wide spectrum of users' preferences. However, existing works often neglect the requirement that the recourses should be close to the data manifold; hence, the constructed recourses might be implausible and unsatisfying to users. To address these issues, we propose a novel approach that explicitly directs the diverse set of actionable recourses towards the data manifold. We first find a diverse set of prototypes in the favorable class that balances the trade-off between diversity and proximity. We demonstrate two specific methods to find these prototypes: either by finding the maximum a posteriori estimate of a determinantal point process or by solving a quadratic binary program. To ensure the actionability constraints, we construct an actionability graph in which the nodes represent the training samples and the edges indicate the feasible action between two instances. We then find a feasible path to each prototype, and this path demonstrates the feasible actions for each recourse in the plan. The experimental results show that our method produces a set of recourses that are close to the data manifold while delivering a better cost-diversity trade-off than existing approaches.

et al., 2020a; Verma et al., 2020; Stepin et al., 2021; Pawelczyk et al., 2021). For example, when a bank deploys an ML model to filter credit loan applicants, it is desirable to provide the denied applicants with certain explanations on why such a decision was made. One possible explanation comes in the form of a directive recommendation such as “*you need to do A to be granted a loan*”, where A is a concrete action to be implemented. This type of explanation is called an algorithmic *recourse* and a set of multiple recourses is called a *recourse plan*. The recourse plan provides a reasoning for the decision and suggests actions that need to perform if the users want to reapply in the future. Algorithmic recourse is becoming a powerful tool to enhance the reliability of the ML model and the engagement of the users, especially in the consequential domains such as loan approvals (Siddiqi, 2012), university admission (Waters and Miiikkulainen, 2014), and job hiring (Ajunwa et al., 2016), to name a few. Algorithmic recourse (Ustun et al., 2019) is also known in the literature of interpretable machine learning as counterfactual explanation (Wachter et al., 2017; Bui et al., 2022) or contrastive explanation (Karimi et al., 2020a).

In practice, many criteria should be considered when constructing a recourse plan. First, each recourse in the plan should be valid in the sense that the recourse should flip the unfavorable outcome of the predictive model. Second, the actions should be relatively small to alleviate the human efforts in implementing the recourses. Third, the recourses should be diverse to cover a wide spectrum of users' preferences. Finally, the recourses should be attainable: the recommended recourses should be actionable for the users who receive them (Ustun et al., 2019). A recourse plan that takes into account these four criteria is called feasible.

Several techniques have been invested in constructing a diverse recourse plan for a given classifier (Russell, 2019; Mothilal et al., 2020; Dandl et al., 2020; Bui et al., 2022). Russell (2019) proposes a mixed-integer programming method to generate a counterfactual plan for a linear classifier, in which the diversity is imposed using a rule-based approach. Dandl et al. (2020) propose a model-agnostic approach using a multi-objective evolutionary algorithm to construct a diverse recourse plan. Alternatively, a recourse plan can also be found by using iterative methods to mini-

1 Introduction

Algorithmic recourse is an emerging method to explain machine learning (ML) models by suggesting how to alter the predictive outcome of any given instance (Karimi

mize the weighted sum of three terms representing validity, proximity, and diversity (Mothilal et al., 2020).

To promote the actionability of the constructed recourses, existing approaches often use constraints to restrict the space of admissible recourses in the optimization problem (Russell, 2019; Dandl et al., 2020; Mothilal et al., 2020). The optimization problem is then solved by common constraint programming (CP) solvers or projected gradient descent algorithms. However, these approaches often require expert knowledge to design suitable actionability constraints. Furthermore, they often consider the input features independently; thus the constructed recourses might be implausible and unsatisfying to the users. For example, the recourse might suggest the user has an income of 10000 with a job that has an income of 1000 on average. To address this issue, one might use the structural causal model (SCM) and find the recourse through the minimal intervention on the SCM graph (Karimi et al., 2020b, 2021). However, this approach generates only a single recourse and requires access to the SCM, which is rarely available in practice (Verma et al., 2020).

FACE (Poyiadzi et al., 2020) is another method for constructing sequential and actionable recourses. To capture actionability constraints, FACE constructs a graph where the nodes represent the training samples and the edges indicate if there is a feasible action that can transform one node into another. FACE then finds the prototypes in a high-density region of the training data with favorable prediction and recommends sequential and actionable recourses as the shortest paths from the input to the prototypes.

Other approaches consider the data manifold when constructing recourse plans (Pawelczyk et al., 2020; Joshi et al., 2019; Looveren and Klaise, 2021). The recourse is considered attainable if it is close to the manifold of the training dataset. Joshi et al. (2019) and Pawelczyk et al. (2020) model the data manifold by learning a variational autoencoder (VAE) for the distribution of the training data. Looveren and Klaise (2021) use an example in the desired class, namely *prototype*, to guide the constructed recourse toward the data manifold. However, these works only consider a single recourse.

Problem Statement. We consider a binary classification setting with a p -dimensional covariate and a binary $\{0, 1\}$ response, where 0 represents a negative/unfavorable outcome and 1 represents a positive/favorable outcome. In real-world applications, the negative outcome may correspond to decisions such as “loan denied” or “application rejected”, while the positive outcome may correspond to “loan approved” or “application accepted”. We are given a classifier $\mathcal{C} : \mathbb{R}^p \rightarrow \{0, 1\}$, and we assume that we have access to N data samples $\hat{x}_i \in \mathbb{R}^p$, $i = 1, \dots, N$ which has positive predicted outcome: $\mathcal{C}(\hat{x}_i) = 1$. Given an input $x_0 \in \mathbb{R}^p$ which receives a negative predicted outcome by

the model, i.e., $\mathcal{C}(x_0) = 0$, we aim to provide a menu, or a plan, consisting K recourses $\{x_k^r\}_{k=1, \dots, K}$ for x_0 that balances multiple criteria such as cost, validity, diversity, and adherence to the data manifold. These criteria can be translated as follows: The cost criterion implies that x_k^r should be close to the input x_0 , where closeness is measured using a certain distance function. The validity requires that $\mathcal{C}(x_k^r) = 1$, while the diversity requires that $\{x_k^r\}$ should be sufficiently different from each other. The adherence to the data manifold here requires that the recourses x_k^r should be relatively close to the aforementioned samples \hat{x}_i .

Our approach and contributions. Our recourse plan generator consists of two stages:

- (i) Find a set of K prototypes $\{x_k\}_{k=1, \dots, K}$ from the available dataset $\{\hat{x}_i\}$.
- (ii) For each k , find the corresponding recourse x_k^r via an interpolation from x_0 to x_k .

Our generator is based on the premise that the interpolation mechanism preserves the desirable properties of the prototypes, including diversity and proximity. Under this premise, if $\{x_k\}$ are diverse then the recourses $\{x_k^r\}$ are also diverse. Moreover, if x_k is close to the input x_0 , it is also likely that the interpolation will lead to a recourse x_k^r that is close to x_0 . Thus, it is imperative to induce diversity and proximity in Step (i) above when we find the K prototypes. Nevertheless, we acknowledge that there is currently no general consensus on the “correct” definition of the diversity for a recourse plan, and different users may prefer different metrics to measure the diversity of a plan.

Guided by this thinking, we introduce two novel methods to find a set of K prototypes that balance the trade-off between the diversity and proximity from x_0 . In Section 2, we formulate a proximity-based determinantal point process and find the prototypes through a maximum a posteriori estimate using a greedy and local search heuristic. In Section 3, we take a different viewpoint on the trade-off, and we formulate a binary quadratic program (BQP) to select the prototypes. We then use an eigen-approximate hierarchy to approximate this BQP, which can be solved efficiently by either a best-response or a dual-ascent method. This hierarchy serves as a screening tool to identify potential samples and subsequently reduce the size of the BQP.

Our method relies on the interpolation module: given an input x_0 and a prototype x_k , the interpolation traverses on a path joining these two endpoints and returns a point x_k^r on this path that is closest to x_0 and has positive prediction $\mathcal{C}(x_k^r) = 1$. There are many possible ways to form a path emanating from x_0 to connect with x_k . If no actionability constraints are imposed, we can use a linear interpolation, which corresponds to tracing a straight line joining x_0 and x_k . If actionability is required, then we can

employ a graph representation similar to FACE (Poyiadzi et al., 2020) and use the shortest path to traverse from x_0 to x_k .

Our proposed scheme has several advantages vis-à-vis the state-of-the-art methods for generating diverse recourse plans, notably the DiCE (Mothilal et al., 2020), and methods for generating sequential and actionable recourses, such as FACE (Poyiadzi et al., 2020):

1. First, by identifying the prototypes and then conducting an interpolation, we guarantee that the recourse x_k^r is directed towards the samples with the favorable predicted value. Notice that DiCE does not assume to have access to the samples $\{\hat{x}_i\}$ from the favorable class, thus the recourses generated from DiCE may not flow towards $\{\hat{x}_i\}$. As a consequence, our recourses will adhere better to the data manifold than the recourses of DiCE, see Figure 1. In addition, by using an appropriate method of interpolation, we guarantee that our recourses are valid, in a sense that $\mathcal{C}(x_k^r) = 1$. Because DiCE is a gradient-based method, there is no guarantee that the recourses generated by DiCE are uniformly valid.
2. Second, our formulation takes the diversity measured with respect to the input x_0 . More specifically, we measure the diversity of prototypes by the diversity of the normalized direction vectors pointing from the input x_0 to the samples \hat{x}_i . In doing so, the diversity will be adjusted to the coordinate of the input vector x_0 . In contrast, the approach from DiCE measures the diversity directly from the samples \hat{x}_i and is not adjusted to x_0 .
3. Third, by using a prototype selection method and an actionability graph, we ensure that the recourses are actionable and diverse. In contrast, the FACE method does not consider the diversity criteria explicitly, thus it may fail to produce a sequential recourse plan that is diverse, see Figure 2 for an example.

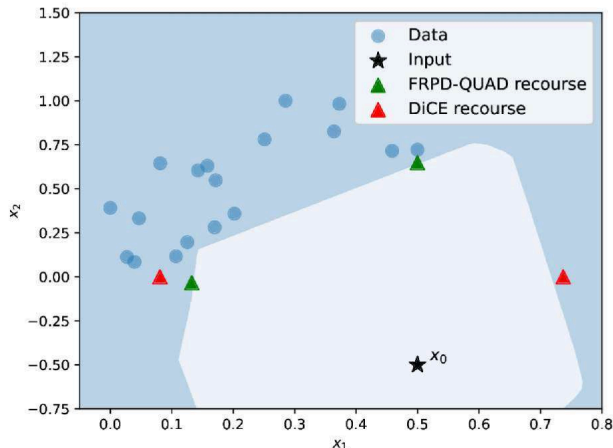


Figure 1: The right, red triangle is suggested by DiCE and it fails to adhere to the data manifold.

Figure 1 shows a two-dimensional example in which the DiCE method recommends a recourse that does not adhere to the data manifold. The positively predicted samples are drawn as blue circles, and the input x_0 is represented by a black star. The background colors represent the positive and negative classification regions. The DiCE recourses (red triangles) are diverse, but one recourse fails to be close to the data manifold (the recourse drawn on the right). The reason is that DiCE is a gradient-based method that does not take the data into account. On the contrary, our recourses found by solving a BQP (green triangles) are correctly directed towards the data manifold.

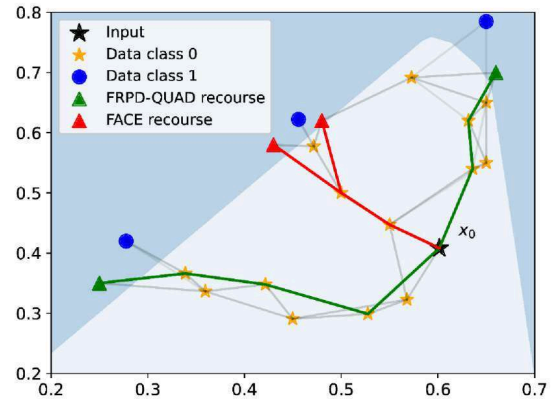


Figure 2: The red triangles are suggested by FACE and they are not sufficiently diverse.

Figure 2 shows a two-dimensional example of recourses generated under the actionability constraints. The grey lines represent all edges of the actionability graph. The green and red lines represent the recourse actions for input x_0 for our method and FACE, respectively. The FACE recourses are not diverse because FACE does not directly optimize this criterion.

Additional notations. Given x_0 and a sample \hat{x}_i , we set the *normalized direction vector* a_i as

$$a_i \triangleq (\hat{x}_i - x_0) / \|\hat{x}_i - x_0\|_2 \in \mathbb{R}^p \quad (1)$$

so that each a_i has unit l_2 -norm. The matrix $A = [a_1, \dots, a_N] \in \mathbb{R}^{p \times N}$ collects all the direction vectors, and we compute the (cosine) similarity matrix as $S = A^\top A$, where we can write each element in S by $S_{ij} = a_i^\top a_j$. The element S_{ij} is also closely related to the cosine angle between the direction vectors of \hat{x}_i and \hat{x}_j . The space of N -by- N symmetric positive definite matrices is denoted by \mathbb{S}_+^N . For any vector z , $\|z\|_0$ counts the number of non-zero elements in z . We write I to denote the identity matrix of the appropriate dimension.

2 Prototype Selection via Determinantal Point Processes

In this section, we construct the set of diverse prototypes using a maximum a posteriori (MAP) estimate of a determinantal point process (DPP). For the purpose of this paper, the choice of the prototypes should balance the trade-off between diversity and proximity with respect to the input x_0 . We first provide a brief introduction to DPP, a more thorough introduction can be found in Appendix B. Then, we construct a proximity-based DPP and discuss how to find its MAP estimate to select the prototypes.

DPPs is a family of probabilistic measures that arise from the field of quantum physics: it is particularly useful to model the repulsive behavior of Fermion particles (Macchi, 1975). Grounded by its probabilistic nature, DPPs have been applied in various machine learning tasks (Kulesza and Taskar, 2012; Affandi et al., 2014; Urschel et al., 2017), ranging from text and video summarization (Lin and Bilmes, 2012; Cho et al., 2019; Gong et al., 2014) to recommendation systems (Chen et al., 2018; Wilhelm et al., 2018; Gartrell et al., 2017).

Definition 2.1 (*L-ensemble DPP*). *Given a positive semidefinite N -by- N matrix $L \in \mathbb{S}_+^N$, an L -ensemble DPP is a distribution over all 2^N index subsets $J \subseteq \{1, \dots, N\}$ such that*

$$\text{Prob}(J) = \det(L_J) / \det(I + L),$$

where L_J denotes the $|J|$ -by- $|J|$ submatrix of L with rows and columns indexed by J .

We need to construct a DPP that aids the selection of K diverse prototypes from N data samples that have favorable predicted outcomes. Because an L -ensemble DPP can be identified via a matrix $L \in \mathbb{S}_+^N$, we focus on choosing an appropriate L that can balance between two criteria: the diversity of the prototypes and the distance of the prototypes to the input x_0 . Towards this end, we impose an additional parameter $\theta \in [0, 1]$ to capture this trade-off, and the matrix L^θ is defined as

$$L^\theta = \theta S + (1 - \theta)D,$$

where S is a similarity matrix among favorable data samples and D is a diagonal matrix capturing the locality structure around x_0 . One possible choice of S is the cosine similarity matrix $S = A^\top A \in \mathbb{S}_+^N$, and one possible choice of D is the diagonal matrix with

$$D_{ii} = \exp(-\text{dist}(x_0, \hat{x}_i)^2 / h^2) \quad \forall i = 1, \dots, N,$$

where $h > 0$ is the bandwidth and dist is a distance function. Notice that because both S and D chosen as such are positive semidefinite, L^θ is also positive semidefinite. We

then find the K diverse prototypes from the training data by solving the following problem

$$\max \{ \det(L_z^\theta) : z \in \{0, 1\}^N, \|z\|_0 = K \}, \quad (2)$$

where L_z^θ is a submatrix of L^θ restricted to rows and columns indexed by the one-components of z . It is well-known that the solution to problem (2) coincides with the MAP estimate of the DPP with a cardinality constraint (Kulesza and Taskar, 2012).

Let us now consider the two extremes: If $\theta = 0$, then $L^0 = D$, it is now easy to see that the optimal solution z^* of (2) will have non-zeros elements corresponding to the maximum values of D_{ii} . Thus, z^* translates to choosing K data samples that are closest to x_0 , measured by the distance function dist . On the other extreme, if $\theta = 1$, then $L^1 = S$, and we recover the canonical setting of DPP *without* the proximity constraints. The resulting optimal solution of (2) in this case will simply promote the diversity of the prototypes.

We now switch gears to discuss the solution procedure to solve (2). First, we highlight that problem (2) is a submodular maximization problem since the log-probability function in DPP $\log \det(L_z)$ is a submodular function (Gillenwater et al., 2012). Further, this problem is well-known to be NP-hard (Kulesza and Taskar, 2012), and thus it is notoriously challenging to solve (2) to optimality. In addition, in real-world applications of the machine learning model, we expect the number of samples N to be large. We thus resort to popular heuristics in order to find a good solution to (2) in a high-dimensional setting with low solution time. A common greedy algorithm to solve the MAP estimation problem (2) is to iteratively find at each incumbent set of prototypes z an index j by

$$j = \underset{i: z_i=0}{\text{argmax}} \log \det(L_{z \vee e_i}) - \log \det(L_z),$$

where \vee returns the element-wise maximum between two vectors and e_i is the vector of zeros with the i -th element being one. The algorithm then adds j to the set of prototypes until reaching the cardinality constraint, i.e., when $\|z\|_0 = K$. In doing so, each j guarantees to maximize the marginal gain to the incumbent set. An efficient implementation of this greedy algorithm that costs $\mathcal{O}(K^2 N)$ time for each inference is provided in Chen et al. (2018).

Local search for MAP inference. The greedy algorithm achieves an approximation ratio of $\mathcal{O}(\frac{1}{k!})$ (Civril and Magdon-Ismael, 2009). To improve its performance, we introduce in Appendix C a simple 2-neighborhood local search that switches one element from the incumbent set with one element from the complementary set.

3 Prototype Selection via Quadratic Binary Programming

We now describe a second approach to finding the set of prototypes that can balance the diversity and proximity trade-off. In stark contrast to the DPP method proposed in the previous section, this second method relies solely on formulating a binary quadratic program in an intuitive manner and then utilizing optimization techniques in order to scale up the problem to the high-dimensional setting.

Formulation. Given the normalized direction vectors a_i defined as in (1), we can measure the anti-diversity of a set of prototypes by a quadratic form

$$\text{Anti-Diversity}(z) \triangleq \sum_{i,j} a_i^\top a_j z_i z_j$$

that simply sums up the pairwise inner product of direction vectors. It is easy to see that if two prototypes i and j have similar directions a_i and a_j , then they are not diverse, which is further represented by the fact that their inner product $a_i^\top a_j$ are close to one. Thus, higher values of $\text{Anti-Diversity}(z)$ indicate that the solution z is not collectively diverse.

Given a distance function dist , we can now formulate a quadratic program whose objective function captures the diversity and proximity trade-off. For a weight parameter $\vartheta \in [0, 1]$, we define the following optimization problem

$$\begin{aligned} \min \quad & \vartheta \text{Anti-Diversity}(z) + (1 - \vartheta) \sum_i \text{dist}(\hat{x}_i, x_0) z_i \\ \text{s. t.} \quad & z \in \{0, 1\}^N, \|z\|_0 = K. \end{aligned}$$

In the matrix notations, we have the equivalent form

$$\begin{aligned} \min \quad & \vartheta z^\top S z + (1 - \vartheta) d^\top z \\ \text{s. t.} \quad & z \in \{0, 1\}^N, \|z\|_0 = K, \end{aligned} \quad (3)$$

where $S \in \mathbb{S}_+^N$ is the similarity matrix, and $d \in \mathbb{R}_+^N$ is the vector of distances from x_0 with $d_i = \text{dist}(\hat{x}_i, x_0)$. If $\vartheta = 0$, then problem (3) only takes the proximity into account and it collapses into a linear binary program. Its optimal solution can be identified using a greedy argument; indeed, the optimal solution z^* , in this case, has $z_i^* = 1$ if the sample \hat{x}_i is one of the K nearest samples from x_0 . When $\vartheta = 1$, problem (3) stresses only on minimizing the anti-diversity, or equivalently, maximizing the diversity of the prototypes. Because $S \in \mathbb{S}_+^N$, problem (3) has a convex quadratic objective function, and it can be solved by state-of-the-art solvers such as CPLEX (CPLEX, 2009), GUROBI (Gurobi Optimization, LLC, 2021) or MOSEK (MOSEK ApS, 2019). Nevertheless, it has N binary variables and is not easy to solve if N is large. Next, we explore approximation methods that find good quality solutions to (3) with low computing overhead.

Eigen-approximate hierarchy. We now delineate our approach to solve (3), which is inspired by recent advances in using the eigen-approximations to solve (convex) quadratic binary programs (Vreugdenhil et al., 2021).

Suppose that S admits the following eigendecomposition $S = \sum_{m=1}^N \sigma_m v_m v_m^\top$, where σ_m are nonnegative eigenvalues and v_m forms an orthogonal basis of \mathbb{R}^N . Without any loss of generality, we suppose that σ_m are sorted in decreasing order, that is, $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_N \geq 0$. Next, we approximate S using its top- M eigenspace approximation, i.e., $S \approx S_M \triangleq \sum_{m=1}^M \sigma_m v_m v_m^\top$, where we assume that $\sigma_M > 0$. In this case, S_M is a strictly positive definite matrix. Following Vreugdenhil et al. (2021), the resulting top- M eigen-approximation of problem (3) is

$$\begin{aligned} \min \quad & (1 - \vartheta) d^\top z + \vartheta z^\top S_M z \\ \text{s. t.} \quad & z \in \{0, 1\}^N, \|z\|_0 = K, \end{aligned} \quad (4)$$

where we emphasize that the objective function of (4) involves the matrix S_M . The eigen-approximation is particularly useful thanks to its following min-max representation.

Lemma 3.1 (Min-max equivalence). *Suppose that $\vartheta \in (0, 1]$. Problem (4) is equivalent to the min-max problem*

$$\min_{z \in \{0, 1\}^N, \|z\|_0 = K} \max_{\gamma \in \mathbb{R}^M} \mathcal{L}(z, \gamma) \quad (5)$$

where \mathcal{L} admits the following form:

$$\begin{aligned} \mathcal{L}(z, \gamma) = & ((1 - \vartheta) d - 2 \sum_{m=1}^M \gamma_m v_m)^\top z \\ & - \sum_{m=1}^M \gamma_m^2 / (\vartheta \sigma_m). \end{aligned} \quad (6)$$

The proof of Lemma 3.1 is included in Appendix D. The advantage of the min-max formulation is that its objective function is a linear function of the binary variables z . This linearity is beneficial because it leads to an analytical optimal solution in the variable z for any fixed value of γ .

Fix any vector $\gamma \in \mathbb{R}^M$, the optimization problem $\min_{z \in \{0, 1\}^N, \|z\|_0 = K} \mathcal{L}(z, \gamma)$ admits the solution

$$z^*(\gamma) = \mathbf{1}_N(\mathcal{I}(\gamma)), \quad (7)$$

where the function $\mathbf{1}_N(\mathcal{I}(\gamma))$ returns a binary vector whose elements in the index set $\mathcal{I}(\gamma)$ is one, and the set $\mathcal{I}(\gamma)$ is

$$\mathcal{I}(\gamma) = \left\{ i \in \{1, \dots, N\} : i \text{ is an index of the } k\text{-smallest elements of vector } (1 - \vartheta) d - 2 \sum_{m=1}^M \gamma_m v_m \right\}.$$

Similarly, for any $z \in \{0, 1\}^N$, the problem over the γ variable $\max_{\gamma \in \mathbb{R}^M} \mathcal{L}(z, \gamma)$ admits the solution

$$\gamma^*(z) = (-\vartheta \sigma_1 v_1^\top z; \dots; -\vartheta \sigma_M v_M^\top z) \in \mathbb{R}^M, \quad (8)$$

which can be verified by the first-order optimality condition. Equipped with this information and inspired by the algorithms in Vreugdenhil et al. (2021), we deploy two different algorithms to screen potential solutions for (4):

Algorithm 1 Best response iterations

Input: Eigenvalues $\sigma_1, \dots, \sigma_M > 0$ with eigenvectors v_1, \dots, v_M , distance vector d , weight ϑ , limit iteration T

Initialization: Set $z_0 \leftarrow 0, \gamma_0 \leftarrow 0$

for $t = 0, \dots, T - 1$ **do**

$$\gamma_{t+1} \leftarrow \begin{pmatrix} -\vartheta \sigma_1 v_1^\top z_t \\ \vdots \\ -\vartheta \sigma_M v_M^\top z_t \end{pmatrix}; z_{t+1} \leftarrow \mathbb{1}_N(\mathcal{I}(\gamma_{t+1}))$$

end for

Output: z_1, \dots, z_T

Algorithm 1 (Best response iterations). This algorithm leverages the best response functions in (7) and (8) to generate an alternating scheme. The iterations proceed by fixing an incumbent solution in z to find the corresponding optimal solution in the γ variable, then switch the role of these two variables to find the optimal solution in the z variable corresponding to the incumbent solution of γ . The pseudocode is presented in Algorithm 1.

Algorithm 2 (Dual ascent iterations). This algorithm proceeds with the dual form of the min-max problem (5), obtained by interchanging the max and min operators:

$$\max_{\gamma \in \mathbb{R}^M} \min_{z \in \{0,1\}^N, \|z\|_0=K} \mathcal{L}(\gamma, z) \quad (9)$$

where \mathcal{L} admits the following form:

$$\mathcal{L}(\gamma, z) = (1 - \vartheta)d^\top z - \sum_{m=1}^M \gamma_m^2 / (\vartheta \sigma_m) - 2 \sum_{m=1}^M \gamma_m v_m^\top z. \quad (10)$$

It is well-known that the optimal value of problem (9) constitutes a lower bound on that of problem (5). The dual ascent iterations leverage the solution (7) to solve the inner problem of (9), then it computes the gradient in the outer γ variable and takes a gradient ascent step with diminishing step sizes. The pseudocode is given in Algorithm 2.

Algorithm 2 Dual program iterations

Input: Similar to Algorithm 1

Parameters: λ (Default $\lambda = 0.1$)

Initialization: Set $z_0 \leftarrow 0, \gamma_0 \leftarrow 0$

for $t = 0, \dots, T - 1$ **do**

$$\gamma_{t+1} \leftarrow \gamma_t - \frac{2\lambda}{\sqrt{t+1}} \begin{pmatrix} \frac{\gamma_t}{\vartheta \sigma_1} + v_1^\top z_t \\ \vdots \\ \frac{\gamma_t}{\vartheta \sigma_m} + v_m^\top z_t \end{pmatrix}$$

$$z_{t+1} \leftarrow \mathbb{1}_N(\mathcal{I}(\gamma_{t+1})),$$

end for

Output: z_1, \dots, z_T

Screening and dimensionality reduction. Both the best response and the dual ascent iterations can serve as screening tools to reduce the dimension of the original problem (3). Let $\tau < T$ be an integer. Intuitively, we look at the last $\tau + 1$ iterate solutions in the variable z and select the indices of non-zero among them. This set of indices will identify the samples that are likely to constitute the optimal solution for the eigen-approximate problem (4). Formally, this set can be found as $\mathcal{Z} = \{i \in [N] : (z_{T-\tau} \vee \dots \vee z_T)_i > 0\}$. The set \mathcal{Z} will be injected to problem (3) to form a new problem

$$\begin{aligned} \min \quad & \vartheta z^\top S z + (1 - \vartheta)d^\top z \\ \text{s. t.} \quad & z \in \{0, 1\}^N, \|z\|_0 = K, z_i = 0 \forall i \notin \mathcal{Z}, \end{aligned} \quad (11)$$

where all indices of z which are not in the set \mathcal{Z} are pre-set to zero. This is equivalent to reducing the number of binary variables, and problem (11) becomes more amenable to commercial solvers. The solution z^* of (11) determines the set of K prototypes for interpolation.

4 Interpolation Schemes

Sections 2 and 3 provide K diverse prototypes to guide users for implementation. However, these recourses might be too costly and too far from the actual effort needed to flip the model prediction. A potential remedy is to use *linear interpolations* between the input instance x_0 and the prototypes $\{x_k\}$ to find the recourses on the decision boundary (Vlassopoulos et al., 2020). In this linear interpolation scheme, the recourses are the intersections between the decision boundary and the line segments joining the input x_0 and the prototypes.

The linear interpolation scheme will provide the plan with the lowest cost that preserves the diversity of K prototypes. However, it is hard to impose actionability constraints into the linear interpolation; thus the recommended recourses could become implausible and unsatisfactory for users. It is notable that in Euclidean spaces, linear interpolation can guarantee that the diversity of recourses is preserved if we use the Euclidean distance in the construction of the similarity matrices for both the DPP (Section 2) and the QUAD (Section 3) methods. Notice that state-of-the-art methods in the recourse literature, such as Poyiadzi et al. (2020) and Mothilal et al. (2020) rely on the Euclidean space setup, and we are not aware of any work that generates diverse recourses in a non-Euclidean space.

To promote actionable and sequential recourses, we leverage the ideas from FACE (Poyiadzi et al., 2020) to build a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$: each node $v \in \mathcal{V}$ corresponds to a sample in the training dataset, and an edge $(u, v) \in \mathcal{E}$ represents an actionable transition from node u to node v . The actionable transition should respect the cost threshold constraints as well as immutable feature constraints which are prescribed in each dataset. A sequential recourse is a di-

rected path from the input instance x_0 to a prototype, each transition in the path is considered an action. We then provide a diverse set of sequential recourses as the shortest paths connecting x_0 to $\{x_k\}$. Compared to FACE, our prototypes chosen from Sections 2 and 3 are already diverse, hence the paths suggested by our method are more likely to be diverse than those suggested by FACE. Further details about sequential recourse are relegated to Appendix A.

5 Numerical Experiments

We analyze our prototype selection variants: Greedy DPP (FRPD-DPP-GR), Local search DPP (FRPD-DPP-LS), and Quadratic (FRPD-QUAD) under non-actionability and actionability environments. For the non-actionability comparison, we use linear interpolation to find the recourses. In this line of comparison, we compare our approaches against DiCE (Mothilal et al., 2020), which is the most popular method to generate diverse recourses. For the actionability comparisons, we find the recourses based on the actionability graph, then compare our methods against FACE (Poyiadzi et al., 2020), which is the state-of-the-art method to generate actionable and sequential recourses. Appendix A reports details about experimental settings and additional numerical results.

Datasets. To construct a diverse recourse plan, we rely on the premise that the available data should be sufficiently diverse in order to find a set of diverse prototypes. Thus, we pick four real-world datasets in the domain of financial applications with potential consequential decisions (e.g., credit loan): German credit (Dua and Graff, 2017), Small Business Administration (SBA) (Li et al., 2018), Bank (Dua and Graff, 2017), Adult (Dua and Graff, 2017). The chosen datasets are widely used in the literature of diverse recourses (Mothilal et al., 2020; Bui et al., 2022; Haldar et al., 2021; Wang et al., 2021; Hasan and Talbert, 2022). We preprocess the data using the same min-max standardizer for the continuous features and one-hot encoding for categorical features as in Mothilal et al. (2020). The FACE-based interpolation scheme can easily handle one-hot features. For linear interpolation, we treat one-hot features in a probabilistic manner to relax them to continuous domains. The feature value with the highest probability will be activated. This approach is common in the recourse literature (Wachter et al., 2017; Mothilal et al., 2020; Upadhyay et al., 2021).

Classifier. We use a three-layer MLP classifier \mathcal{C} with the hidden sizes of 20, 50, and 20. For each dataset, we split uniformly at random 80% of the original dataset to train the classifier \mathcal{C} . The remaining data from the dataset is used to evaluate our methods and baselines.

We compare using the following metrics:

Cost. We measure the cost of a plan $\{x_k^r\}$ by the average

distance from the instance x_0 to each recourse in the plan:

$$\text{Cost}(\{x_k^r\}, x_0) \triangleq \frac{1}{K} \sum_{k=1}^K \text{dist}(x_k^r, x_0).$$

Validity. A plan is considered valid if every recourse in the plan flips the prediction of the underlying model. We then compute the validity metric as the fraction of instances for which the constructed recourse plan is valid.

Anti-Diversity. We compute the Anti-Diversity of a recourse plan $\{x_k^r\}$ as the sum of pairwise cosine similarity of the normalized direction vectors, which can be written

$$\text{Anti-Diversity}(\{x_k^r\}, x_0) \triangleq \sum_{1 \leq k, k' \leq K, k \neq k'} \frac{(x_k^r - x_0)^\top (x_{k'}^r - x_0)}{\|x_k^r - x_0\|_2 \|x_{k'}^r - x_0\|_2}.$$

DPP. We compute the DPP of a recourse plan using the same measure as in Mothilal et al. (2020), which can be written explicitly as

$$\text{DPP}(\{x_k^r\}) \triangleq \det(Q),$$

where $Q_{ij} = (1 + \text{dist}(x_i^r, x_j^r))^{-1} \quad \forall 1 \leq i, j \leq K.$

Distance to data manifold. The distance of a recourse plan to the data manifold is the maximum distance from each recourse in the plan to the data manifold. Notice that measuring the adherence to the data manifold using a latent distance is common in disentangled-based recourse methods such as Pawelczyk et al. (2020). In this experiment, we assume data from class 1 is the target data manifold:

$$\text{Manifold-Distance}(\{x_k^r\}) \triangleq \max_{k \in \{1, \dots, K\}} \min_{i \in \{1, \dots, N\}} \text{dist}(x_k^r, \hat{x}_i).$$

5.1 Experiments without Actionability Constraints

To evaluate the trade-off between cost and Anti-Diversity, and between cost and DPP, we vary $\theta = \vartheta \in [0.1, 1.0]$ for two methods: FRPD-QUAD, FRPD-DPP-LS. For DiCE’s parameters, we fix the proximity weight and change the diversity weight between 0 and 5.0. Because there are no actionability constraints, we choose the distance function as the Euclidean distance.

We visualize the Pareto frontiers of the cost-Anti-Diversity and cost-DPP trade-offs of FRPD-QUAD, FRPD-DPP-LS, and DiCE in Figure 3. We report the additional results for FRPD-DPP-GR in Appendix A. The results in Figure 3 demonstrate that FRPD-QUAD, FRPD-DPP-GR, and FRPD-DPP-LS produce a diverse recourse plan at a far lower cost than DiCE. In addition, Table 1 demonstrates that the validity of DiCE drops to below one in the SBA dataset. Further, FRPD-QUAD has the lowest Anti-Diversity, which is natural because this method optimizes the Anti-Diversity metric explicitly. Similarly, if we use the DPP to measure the diversity, our DPP-based methods,

Feasible Recourse Plan via Diverse Interpolation

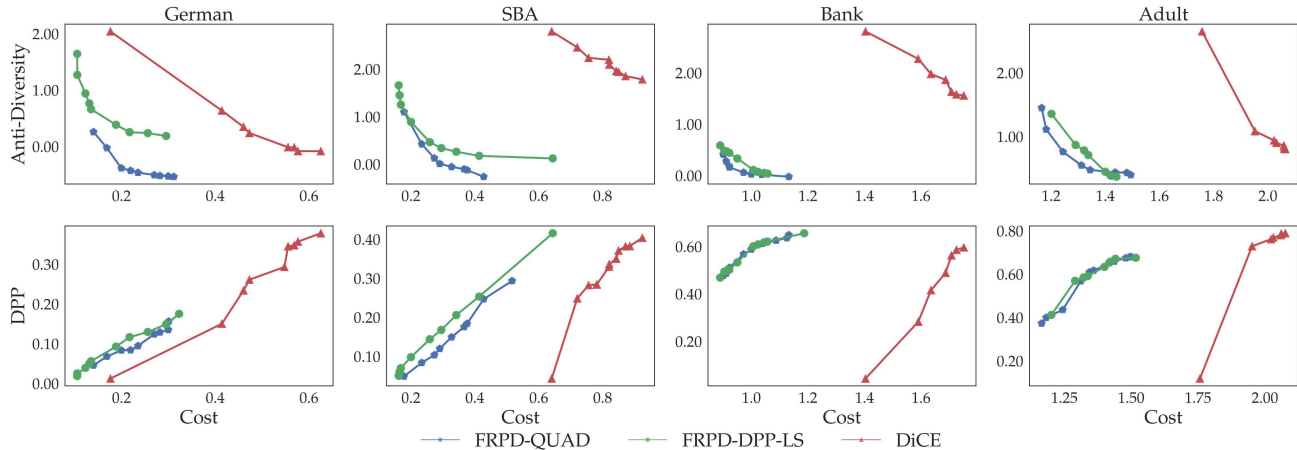


Figure 3: Pareto frontier of the cost vs. Anti-Diversity (top) and cost vs. DPP (bottom) trade-off on four real-world datasets.

Table 1: Benchmark of Cost, Validity, Anti-Diversity, DPP and Manifold-Distance between competing methods *without* actionability constraints. Details on experiment setting are provided in Appendix A.

Dataset	Methods	Cost	Validity	Anti Diversity	DPP	Manifold Distance
German	DiCE	0.41 ± 0.34	1.00 ± 0.00	0.71 ± 0.74	0.14 ± 0.09	0.15 ± 0.11
	FRPD-QUAD	0.30 ± 0.19	1.00 ± 0.00	-0.53 ± 0.70	0.14 ± 0.12	0.19 ± 0.19
	FRPD-DPP-GR	0.31 ± 0.22	1.00 ± 0.00	0.17 ± 0.20	0.16 ± 0.14	0.27 ± 0.23
	FRPD-DPP-LS	0.30 ± 0.18	1.00 ± 0.00	0.18 ± 0.20	0.15 ± 0.12	0.30 ± 0.23
SBA	DiCE	0.71 ± 0.22	0.95 ± 0.22	2.49 ± 0.61	0.23 ± 0.12	0.86 ± 0.24
	FRPD-QUAD	0.45 ± 0.30	1.00 ± 0.00	-0.17 ± 0.60	0.23 ± 0.19	0.23 ± 0.16
	FRPD-DPP-GR	0.41 ± 0.33	1.00 ± 0.00	0.18 ± 0.21	0.24 ± 0.23	0.26 ± 0.19
	FRPD-DPP-LS	0.41 ± 0.33	1.00 ± 0.00	0.18 ± 0.19	0.25 ± 0.23	0.27 ± 0.19
Bank	DiCE	1.62 ± 0.32	1.00 ± 0.00	2.14 ± 0.64	0.34 ± 0.26	0.48 ± 0.37
	FRPD-QUAD	1.09 ± 0.21	1.00 ± 0.00	0.03 ± 0.26	0.63 ± 0.11	0.51 ± 0.20
	FRPD-DPP-GR	1.06 ± 0.21	1.00 ± 0.00	0.04 ± 0.11	0.62 ± 0.11	0.43 ± 0.20
	FRPD-DPP-LS	1.06 ± 0.21	1.00 ± 0.00	0.04 ± 0.11	0.62 ± 0.11	0.43 ± 0.20
Adult	DiCE	1.95 ± 0.37	1.00 ± 0.00	1.08 ± 0.55	0.73 ± 0.13	1.54 ± 0.31
	FRPD-QUAD	1.48 ± 0.58	1.00 ± 0.00	0.43 ± 0.33	0.67 ± 0.18	0.60 ± 0.32
	FRPD-DPP-GR	1.44 ± 0.61	1.00 ± 0.00	0.37 ± 0.30	0.67 ± 0.19	0.66 ± 0.36
	FRPD-DPP-LS	1.44 ± 0.61	1.00 ± 0.00	0.37 ± 0.30	0.67 ± 0.19	0.66 ± 0.36

such as FRPD-DPP-GR and FRPD-DPP-LS outperform in this metric for three over four real datasets. Finally, our three methods are more likely to generate a recourse plan that is close to the data manifold.

5.2 Experiments with Actionability Constraints

Herein, we use the shortest path interpolation on the actionability graph and compare our recourses against FACE. We use CARLA’s source code (Pawelczyk et al., 2021) to construct the actionability graph as the common input.

In this benchmark, we choose the distance function as the shortest path on the actionability graph. Because each recourse has to be a node in the graph, we do not compare the distance to the manifold in this experiment. The results in Table 2 demonstrate that our three methods outperform FACE in terms of two diversity metrics, Anti-Diversity and

DPP. This result is natural because our methods directly optimize for the Anti-Diversity metric. In contrast, FACE does not take diversity as an explicit criterion. The FACE method, however, provides the smallest cost.

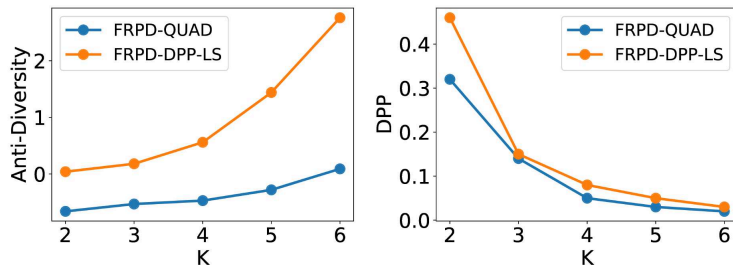
5.3 Sensitivity analysis of the number of recourses K

In this experiment, we study the impact of the number of selected prototypes K of a recourse plan on two diversity metrics: Anti-Diversity and DPP. First, we choose $\theta = \vartheta = 0.9$ and vary the number of recourses in a plan $K \in [2, 6]$ for our two methods: FRPD-QUAD and FRPD-DPP-LS.

Figure 4 reports the trade-off between K and two diversity measures on the German credit dataset. These results show that there is a trade-off between the number of selected prototypes and the diversity measures: as K increases, Anti-Diversity increases and DPP decreases.

Table 2: Benchmark of Cost, Validity, Anti-Diversity, and DPP between competing methods with actionability constraints on four real-world datasets. Details on experiment setting are provided in Appendix A.

Dataset	Methods	Shortest Path	Validity	Anti Diversity	DPP
German	FACE	0.24 \pm 0.17	1.00 \pm 0.00	2.18 \pm 0.77	0.04 \pm 0.04
	FRPD-QUAD	0.55 \pm 0.24	1.00 \pm 0.00	-0.13 \pm 0.24	0.21 \pm 0.07
	FRPD-DPP-GR	0.42 \pm 0.20	1.00 \pm 0.00	0.20 \pm 0.23	0.21 \pm 0.05
	FRPD-DPP-LS	0.42 \pm 0.20	1.00 \pm 0.00	0.22 \pm 0.23	0.23 \pm 0.07
SBA	FACE	0.25 \pm 0.20	1.00 \pm 0.00	2.08 \pm 0.78	0.02 \pm 0.03
	FRPD-QUAD	0.55 \pm 0.25	1.00 \pm 0.00	-0.04 \pm 0.41	0.17 \pm 0.09
	FRPD-DPP-GR	0.52 \pm 0.25	1.00 \pm 0.00	0.41 \pm 0.25	0.21 \pm 0.03
	FRPD-DPP-LS	0.52 \pm 0.25	1.00 \pm 0.00	0.41 \pm 0.25	0.21 \pm 0.03
Bank	FACE	1.29 \pm 0.29	1.00 \pm 0.00	1.42 \pm 0.86	0.31 \pm 0.27
	FRPD-QUAD	1.55 \pm 0.44	1.00 \pm 0.00	0.06 \pm 0.12	0.69 \pm 0.16
	FRPD-DPP-GR	1.52 \pm 0.42	1.00 \pm 0.00	0.16 \pm 0.26	0.74 \pm 0.04
	FRPD-DPP-LS	1.52 \pm 0.42	1.00 \pm 0.00	0.16 \pm 0.26	0.74 \pm 0.04
Adult	FACE	1.33 \pm 0.36	1.00 \pm 0.00	2.66 \pm 0.36	0.22 \pm 0.27
	FRPD-QUAD	1.99 \pm 0.59	1.00 \pm 0.00	0.52 \pm 0.17	0.78 \pm 0.06
	FRPD-DPP-GR	2.00 \pm 0.52	1.00 \pm 0.00	0.45 \pm 0.23	0.75 \pm 0.17
	FRPD-DPP-LS	2.00 \pm 0.52	1.00 \pm 0.00	0.45 \pm 0.23	0.75 \pm 0.17


 Figure 4: Impact of the number of recourses K to Anti-Diversity and DPP on the German Credit dataset.

Conclusions. We propose a novel method to generate a diverse recourse plan that is feasible in the sense that the plan takes into consideration multiple criteria such as cost, validity, diversity, and adherence to the data manifold. As diversity may possess different meanings and metrics, we provide two independent and possibly complementary approaches to select the prototypes for subsequent interpolation: one relies on MAP inference for DPPs and the other relies on a quadratic programming formulation. The prototypes can be used as targets for different interpolation schemes to generate diverse and sequential recourses. Finally, we remark that the goal of this paper is *not* about providing a normative answer to which diversity metric should be preferred. We instead relegate this interesting question to future research.

Remark 1 (Privacy of recourse). *The goal of this paper is to devise a diverse recourse plan that has a higher capacity of matching the different preferences of the subjects.*

A foundational assumption of this paper is the available access to the training data. In particular, to construct the set of diverse prototypes, we need access to positively-predicted training samples. The quality of our approach is dependent on the instances with positive predictions. Given that recourse is user-dependent, the training set for the

problem of recourse diversity should be sufficiently diverse to account for a wide range of user preferences.

Moreover, if we use the graph interpolation in Section 4, then we need additional access to the negatively-predicted training samples. Access to the training data is crucial to identify the manifold of the data and to generate actionable recourses. This access to the training data is leveraged in existing methods such as FACE (Poyiadzi et al., 2020). The required access to the training data is often criticized due to its potential breach of privacy. To resolve this issue, we can utilize multiple techniques and methods to guarantee privacy. For a concrete example, each sample can be modified by adding an independent noise vector, which is able to satisfy the notion of local differential privacy. The exact computation of the noise variance is left for future research.

Remark 2 (Family of classifiers). *In this work, we treat the underlying classifier as a black-box model, so our methods can handle any family of classifiers (deep NNs, tree-based classifiers, etc.). Moreover, our framework can easily adapt to the multi-class classification problem. In this case, we treat the data from the desired class as potential candidates for selecting K prototypes. An interpolation technique between the input instance and K prototypes can be used to find recourses, for example, by linear interpolation.*

Acknowledgments. Viet Anh Nguyen acknowledges the generous support from the CUHK’s Improvement on Competitiveness in Hiring New Faculties Funding Scheme.

References

- Raja Hafiz Affandi, Emily Fox, Ryan Adams, and Ben Taskar. Learning the parameters of determinantal point process kernels. In *International Conference on Machine Learning*, pages 1224–1232. PMLR, 2014.
- Ifeoma Ajunwa, Sorelle Friedler, Carlos E Scheidegger, and Suresh Venkatasubramanian. Hiring by algorithm: Predicting and preventing disparate impact. Available at SSRN, 2016.
- Ngoc Bui, Duy Nguyen, and Viet Anh Nguyen. Counterfactual plans under distributional ambiguity. In *International Conference on Learning Representations*, 2022.
- Laming Chen, Guoxin Zhang, and Eric Zhou. Fast greedy MAP inference for determinantal point process to improve recommendation diversity. *Advances in Neural Information Processing Systems*, 31, 2018.
- Sangwoo Cho, Chen Li, Dong Yu, Hassan Foroosh, and Fei Liu. Multi-document summarization with determinantal point processes and contextualized representations. *arXiv preprint arXiv:1910.11411*, 2019.
- Theodoros Chondrogiannis, Panagiotis Bouros, Johann Gamper, Ulf Leser, and David B Blumenthal. Finding k-dissimilar paths with minimum collective length. In *Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 404–407, 2018.
- Ali Civril and Malik Magdon-Ismail. On selecting a maximum volume sub-matrix of a matrix and related problems. *Theoretical Computer Science*, 410(47-49):4801–4811, 2009.
- IBM ILOG CPLEX. V12. 1: User’s manual for CPLEX. *International Business Machines Corporation*, 46(53): 157, 2009.
- Susanne Dandl, Christoph Molnar, Martin Binder, and Bernd Bischl. Multi-objective counterfactual explanations. In *International Conference on Parallel Problem Solving from Nature*, pages 448–469. Springer, 2020.
- Michael Downs, Jonathan L Chu, Yaniv Yacoby, Finale Doshi-Velez, and Weiwei Pan. Cruds: Counterfactual recourse using disentangled subspaces. *ICML WHI*, 2020: 1–23, 2020.
- Dheeru Dua and Casey Graff. UCI machine learning repository, 2017. URL <http://archive.ics.uci.edu/ml>.
- Mike Gartrell, Ulrich Paquet, and Noam Koenigstein. Low-rank factorization of determinantal point processes. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2017.
- Jennifer Gillenwater, Alex Kulesza, and Ben Taskar. Near-optimal MAP inference for determinantal point processes. *Advances in Neural Information Processing Systems*, 25, 2012.
- Boqing Gong, Wei-Lun Chao, Kristen Grauman, and Fei Sha. Diverse sequential subset selection for supervised video summarization. *Advances in Neural Information Processing Systems*, 27, 2014.
- Riccardo Guidotti, Anna Monreale, Salvatore Ruggieri, Dino Pedreschi, Franco Turini, and Fosca Giannotti. Local rule-based explanations of black box decision systems. *arXiv preprint arXiv:1805.10820*, 2018.
- Gurobi Optimization, LLC. Gurobi Optimizer Reference Manual, 2021. URL <https://www.gurobi.com>.
- Swastik Haldar, Philips George John, and Diptikalyan Saha. Reliable counterfactual explanations for autoencoder based anomalies. In *8th ACM IKDD CODS and 26th COMAD*, pages 83–91, 2021.
- Md Golam Moula Mehedi Hasan and Douglas Talbert. Data augmentation using counterfactuals: Proximity vs diversity. In *The International FLAIRS Conference Proceedings*, volume 35, 2022.
- Shalmali Joshi, Oluwasanmi Koyejo, Warut Vijitbenjaronk, Been Kim, and Joydeep Ghosh. Towards realistic individual recourse and actionable explanations in black-box decision making systems. *arXiv preprint arXiv:1907.09615*, 2019.
- Amir-Hossein Karimi, Gilles Barthe, Bernhard Schölkopf, and Isabel Valera. A survey of algorithmic recourse: Definitions, formulations, solutions, and prospects. *arXiv preprint arXiv:2010.04050*, 2020a.
- Amir-Hossein Karimi, Julius Von Kügelgen, Bernhard Schölkopf, and Isabel Valera. Algorithmic recourse under imperfect causal knowledge: A probabilistic approach. *arXiv preprint arXiv:2006.06831*, 2020b.
- Amir-Hossein Karimi, Bernhard Schölkopf, and Isabel Valera. Algorithmic recourse: From counterfactual explanations to interventions. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pages 353–362, 2021.
- Alex Kulesza and Ben Taskar. Determinantal point processes for machine learning. *arXiv preprint arXiv:1207.6083*, 2012.
- Min Li, Amy Mickel, and Stanley Taylor. “Should this loan be approved or denied?”: A large dataset with class assignment guidelines. *Journal of Statistics Education*, 26(1):55–66, 2018.
- Hui Lin and Jeff Bilmes. Learning mixtures of submodular shells with application to document summarization. In *Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence*, 2012.

- Arnaud Van Looveren and Janis Klaise. Interpretable counterfactual explanations guided by prototypes. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 650–665. Springer, 2021.
- Odile Macchi. The coincidence approach to stochastic point processes. *Advances in Applied Probability*, 7(1): 83–122, 1975.
- MOSEK ApS. *MOSEK Optimizer API for Python 9.2.10*, 2019. URL <https://docs.mosek.com/9.2/pythonapi/index.html>.
- Ramaravind K Mothilal, Amit Sharma, and Chenhao Tan. Explaining machine learning classifiers through diverse counterfactual explanations. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pages 607–617, 2020.
- Gonzalo Navarro. A guided tour to approximate string matching. *ACM Computing Surveys (CSUR)*, 33(1):31–88, 2001.
- Martin Pawelczyk, Klaus Broelemann, and Gjergji Kasneci. Learning model-agnostic counterfactual explanations for tabular data. In *Proceedings of The Web Conference 2020*, pages 3126–3132, 2020.
- Martin Pawelczyk, Sascha Bielawski, Johannes van den Heuvel, Tobias Richter, and Gjergji Kasneci. CARLA: A Python library to benchmark algorithmic recourse and counterfactual explanation algorithms. *arXiv preprint arXiv:2108.00783*, 2021.
- Rafael Poyiadzi, Kacper Sokol, Raul Santos-Rodriguez, Tijn De Bie, and Peter Flach. FACE: Feasible and actionable counterfactual explanations. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, pages 344–350, 2020.
- Goutham Ramakrishnan, Yun Chan Lee, and Aws Albarghouti. Synthesizing action sequences for modifying model decisions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 5462–5469, 2020.
- Chris Russell. Efficient search for diverse coherent explanations. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pages 20–28, 2019.
- Naeem Siddiqi. *Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring*. John Wiley & Sons, 2012.
- Ronal Singh, Paul Dourish, Piers Howe, Tim Miller, Liz Sonenberg, Eduardo Velloso, and Frank Vetere. Directive explanations for actionable explainability in machine learning applications. *arXiv preprint arXiv:2102.02671*, 2021.
- Ilija Stepin, Jose M Alonso, Alejandro Catala, and Martín Pereira-Fariña. A survey of contrastive and counterfactual explanation generation methods for explainable artificial intelligence. *IEEE Access*, 9:11974–12001, 2021.
- Sohini Upadhyay, Shalmali Joshi, and Himabindu Lakkaraju. Towards robust and reliable algorithmic recourse. In *Advances in Neural Information Processing Systems*, 2021.
- John Urschel, Victor-Emmanuel Brunel, Ankur Moitra, and Philippe Rigollet. Learning determinantal point processes with moments and cycles. In *International Conference on Machine Learning*, pages 3511–3520. PMLR, 2017.
- Berk Ustun, Alexander Spangher, and Yang Liu. Actionable recourse in linear classification. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pages 10–19, 2019.
- Sahil Verma, John Dickerson, and Keegan Hines. Counterfactual explanations for machine learning: A review. *arXiv preprint arXiv:2010.10596*, 2020.
- Sahil Verma, Keegan Hines, and John P Dickerson. Amortized generation of sequential algorithmic recourses for black-box models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022.
- Georgios Vlassopoulos, Tim van Erven, Henry Brighton, and Vlado Menkovski. Explaining predictions by approximating the local decision boundary. *arXiv preprint arXiv:2006.07985*, 2020.
- Robbie Vreugdenhil, Viet Anh Nguyen, Armin Eftekhari, and Peyman Mohajerin Esfahani. Principal component hierarchy for sparse quadratic programs. In *Proceedings of the 38th International Conference on Machine Learning*, 2021.
- Sandra Wachter, Brent Mittelstadt, and Chris Russell. Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harvard Journal of Law & Technology*, 31:841, 2017.
- Yongjie Wang, Qinxu Ding, Ke Wang, Yue Liu, Xingyu Wu, Jinglong Wang, Yong Liu, and Chunyan Miao. The skyline of counterfactual explanations for machine learning decision models. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, pages 2030–2039, 2021.
- Austin Waters and Risto Miikkulainen. Grade: Machine learning support for graduate admissions. *AI Magazine*, 35(1):64–64, 2014.
- Mark Wilhelm, Ajith Ramanathan, Alexander Bonomo, Sagar Jain, Ed H Chi, and Jennifer Gillenwater. Practical diversified recommendations on Youtube with determinantal point processes. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pages 2165–2173, 2018.

Supplementary Materials for Paper: Feasible Recourse Plan via Diverse Interpolation

A Additional Experiment Results

Source code, datasets, and results can be accessed from <https://github.com/duykhuongnguyen/recourse-plan-diverse-interpolation>. We use the implementation of CARLA (Pawelczyk et al., 2021) for two baselines: DiCE and FACE.

A.1 Datasets

Real-world datasets. To form a diverse recourse plan, we rely on the premise that the available data should be sufficiently diverse in order to find a set of diverse prototypes. Thus, we pick four real-world datasets in the domain of financial applications with potential consequential decisions (e.g., credit loan) and these datasets are diverse in different perspectives:

- The German Credit dataset (Dua and Graff, 2017) contains 1,000 samples of loan applicants, each applicant is classified as either a good or a bad client. We choose five features: Status, Duration, Credit amount, Personal status, and Age. The Duration and Credit amount features represent key attributes of a loan application or, equivalently, of a financial product. This dataset exhibits product diversity.
- The Small Business Administration (SBA) dataset (Li et al., 2018) contains 2,102 samples of small company loans in California, the label indicates whether a business has defaulted on a loan. For the SBA dataset, we choose the following features: Selected, Term, NoEmp, CreateJob, RetainedJob, UrbanRural, ChgOffPrinGr, GrAppv, SBA Appv, New, RealEstate, Portion, Recession. Collected over 24 years from 1989 and 2012, this dataset exhibits temporal diversity.
- The Bank dataset (Dua and Graff, 2017) contains 4,522 samples of individuals undergoing direct marketing activities from a financial institution. The label represents whether the client subscribes to a bank term deposit (a financial product). We use the following features: Age, Education, Balance, Housing, Loan, Campaign, Previous, and Outcome. The Campaign and Previous feature the number of contacts with the client in this and the previous campaign, respectively, and they represent the aggressiveness in the marketing strategy. This dataset exhibits strategic diversity.
- The Adult dataset (Dua and Graff, 2017) contains 32,560 samples of individuals for income prediction. For this dataset, we use full features of the dataset: Age, Workclass, Fnlwgt, Education, Educational-num, Marital-status, Occupation, Relationship, Race, Gender, Capital-gain, Capital-loss, Hours-per-week, Native-country, and Income. This dataset collects a diverse set of individual-level features, and it exhibits intersectional diversity.

The chosen datasets are widely used in the literature of diverse recourses (Mothilal et al., 2020; Bui et al., 2022; Haldar et al., 2021; Wang et al., 2021; Hasan and Talbert, 2022). We preprocess the data using the same min-max standardizer for the continuous features and one-hot encoding for categorical features as in Mothilal et al. (2020).

To construct the graph for actionability methods, we choose the below set of immutable features for each dataset:

- For German Credit, we select “Personal status” since it is challenging to compel changes in an individual’s personal status (male, female, single, married) (Ustun et al., 2019; Karimi et al., 2020b).
- For SBA, we select “Recession” and “UrbanRural” as they are problematic to change in the foreseeable future for the SBA dataset.
- For Bank, we select “previous”, “campaign”, and “outcome” since they are historical attributes.
- For Adult, we select “marital status” and “gender”.

A.2 Experimental setup

Classifier. For each dataset, we first do the 80-20 split (80% for training, 20% for testing) and train an MLP classifier on the training set. We report the Accuracy and AUC of the MLP classifier on each dataset in Table 3.

Table 3: Accuracy and AUC of the MLP classifiers on four real-world datasets.

Dataset	Accuracy	AUC
German Credit	0.73	0.71
SBA	0.97	0.99
Bank	0.89	0.68
Adult	0.83	0.88

Settings for Figure 3. In this experiment, we generate a recourse plan with $K = 3$ recourses. We vary $\theta = \vartheta \in [0.1, 1.0]$ for three methods: FRPD-QUAD, FRPD-DPP-GR and FRPD-DPP-LS. We choose $h = 1.0$ for FRPD-DPP-GR and FRPD-DPP-LS. We fix the proximity weight 0.5 as the default setting and change the diversity weight between 0 and 5.0 for DiCE.

Settings for Table 1 and Table 2. In this experiment, we choose $K = 3$ and $\theta = \vartheta = 0.9$ for our three methods: FRPD-QUAD, FRPD-DPP-GR and FRPD-DPP-LS. We choose $h = 1.0$ for FRPD-DPP-GR and FRPD-DPP-LS. We use the default setting for proximity weight and diversity weight of DiCE with 0.5 and 1.0, respectively. We use the default settings in CARLA (Pawelczyk et al., 2021) for FACE.

A.3 Additional numerical results

Dual program iterations. In the main paper, we utilize the best response iterations in Algorithm 1 (denoted FRPD-QUAD) to solve problem (5). In this section, we use the dual program iterations in Algorithm 2 (denoted FRPD-QUAD-DP) to solve the same problem and compare it with FRPD-QUAD. We evaluate different trade-offs: cost vs. Anti-Diversity and cost vs. DPP trade-off on four real-world datasets, and report the results in Figure 5. We also compute cost, Anti-Diversity, DPP, manifold distance, and discuss how each metric is correlated, and report the results in Table 4 and Table 5. The settings for these experiments are the same as in Figure 3, Table 1, and Table 2.

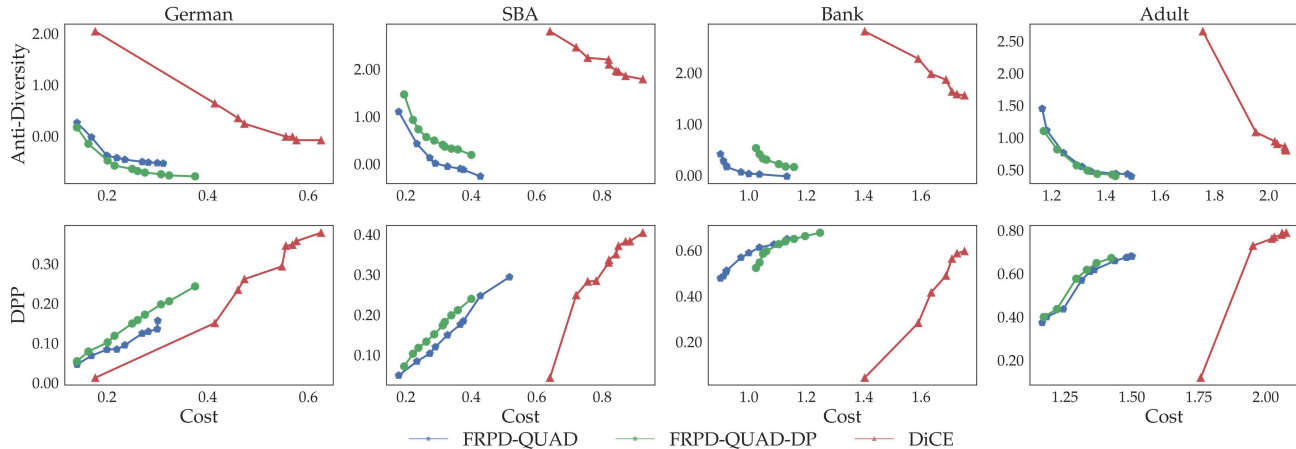


Figure 5: Pareto frontier of the cost vs. Anti-Diversity (top) and cost vs. DPP (bottom) trade-off between FRPD-QUAD, FRPD-QUAD-DP using linear interpolation, and DiCE on four real-world datasets.

The results in Figure 5 indicate that FRPD-QUAD-DP has higher DPP than FRPD-QUAD in three datasets. Further, the recourse plans generated by FRPD-QUAD are more directly correlated with the data manifold in the SBA and Bank datasets. Our two approaches also produce a recourse plan that closely resembles the data manifold. Table 5 shows that in all four datasets, FRPD-QUAD-DP outperforms FACE in terms of Anti-Diversity and DPP. This implies that FRPD-QUAD-DP recourses are more diverse than FACE recourses.

Cost-diversity trade-off of FRPD-DPP-GR. Here, we provide the additional results for FRPD-DPP-GR: the cost and

Feasible Recourse Plan via Diverse Interpolation

Table 4: Benchmark of Cost, Validity, Anti-Diversity, DPP and Manifold-Distance between FRPD-QUAD, FRPD-DP using linear interpolation, and DiCE on four real-world datasets.

Dataset	Methods	Cost	Validity	Anti Diversity	DPP	Manifold Distance
German	DiCE	0.41 ± 0.34	1.00 ± 0.00	0.71 ± 0.74	0.14 ± 0.09	0.15 ± 0.11
	FRPD-QUAD	0.30 ± 0.19	1.00 ± 0.00	-0.53 ± 0.70	0.14 ± 0.12	0.19 ± 0.19
	FRPD-QUAD-DP	0.31 ± 0.23	1.00 ± 0.00	-0.75 ± 0.61	0.20 ± 0.15	0.17 ± 0.18
SBA	DiCE	0.71 ± 0.22	0.95 ± 0.22	2.49 ± 0.61	0.23 ± 0.12	0.86 ± 0.24
	FRPD-QUAD	0.45 ± 0.30	1.00 ± 0.00	-0.17 ± 0.60	0.23 ± 0.19	0.23 ± 0.16
	FRPD-QUAD-DP	0.34 ± 0.28	1.00 ± 0.00	0.32 ± 0.65	0.20 ± 0.18	0.35 ± 0.25
Bank	DiCE	1.62 ± 0.32	1.00 ± 0.00	2.14 ± 0.64	0.34 ± 0.26	0.48 ± 0.37
	FRPD-QUAD	1.09 ± 0.21	1.00 ± 0.00	0.03 ± 0.26	0.63 ± 0.11	0.51 ± 0.20
	FRPD-QUAD-DP	1.20 ± 0.25	1.00 ± 0.00	0.16 ± 0.31	0.66 ± 0.09	0.60 ± 0.22
Adult	DiCE	1.95 ± 0.37	1.00 ± 0.00	1.08 ± 0.55	0.73 ± 0.13	1.54 ± 0.31
	FRPD-QUAD	1.48 ± 0.58	1.00 ± 0.00	0.43 ± 0.33	0.67 ± 0.18	0.60 ± 0.32
	FRPD-QUAD-DP	1.44 ± 0.59	1.00 ± 0.00	0.40 ± 0.36	0.67 ± 0.18	0.59 ± 0.31

Table 5: Benchmark of Cost, Validity, Anti-Diversity, and DPP between FRPD-QUAD, FRPD-QUAD-DP, and FACE with actionability constraints on four real-world datasets.

Dataset	Methods	Shortest Path	Validity	Anti Diversity	DPP
German	FACE	0.24 ± 0.17	1.00 ± 0.00	2.18 ± 0.77	0.04 ± 0.04
	FRPD-QUAD	0.55 ± 0.24	1.00 ± 0.00	-0.13 ± 0.24	0.21 ± 0.07
	FRPD-QUAD-DP	0.33 ± 0.11	1.00 ± 0.00	0.20 ± 0.23	0.24 ± 0.05
SBA	FACE	0.25 ± 0.20	1.00 ± 0.00	2.08 ± 0.78	0.02 ± 0.03
	FRPD-QUAD	0.55 ± 0.25	1.00 ± 0.00	-0.04 ± 0.41	0.17 ± 0.09
	FRPD-QUAD-DP	0.68 ± 0.32	1.00 ± 0.00	0.35 ± 0.25	0.21 ± 0.09
Bank	FACE	1.29 ± 0.29	1.00 ± 0.00	1.42 ± 0.86	0.31 ± 0.27
	FRPD-QUAD	1.55 ± 0.44	1.00 ± 0.00	0.06 ± 0.12	0.69 ± 0.16
	FRPD-QUAD-DP	1.48 ± 0.55	1.00 ± 0.00	0.07 ± 0.16	0.72 ± 0.03
Adult	FACE	1.33 ± 0.36	1.00 ± 0.00	2.66 ± 0.36	0.22 ± 0.27
	FRPD-QUAD	1.99 ± 0.59	1.00 ± 0.00	0.52 ± 0.17	0.78 ± 0.06
	FRPD-QUAD-DP	1.98 ± 0.52	1.00 ± 0.00	0.56 ± 0.19	0.75 ± 0.04

Anti-Diversity, cost, and DPP trade-off of FRPD-DPP-GR without actionability constraints. We use the same settings for FRPD-DPP-GR: fix $h = 1.0$ and vary $\vartheta \in [0.1, 1.0]$. The results in Figure 6 indicate that FRPD-QUAD and FRPD-DPP-GR produce a diverse recourse plan at a far lower cost than DiCE.

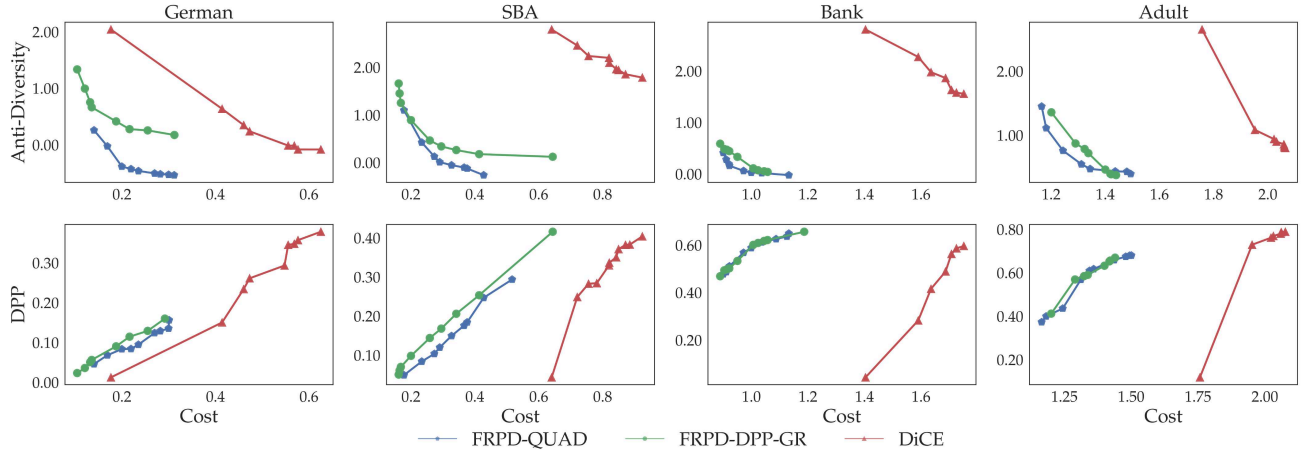


Figure 6: Pareto frontier of the cost vs. Anti-Diversity (top) and cost vs. DPP (bottom) trade-off between FRPD-QUAD, FRPD-DPP-GR using linear interpolation, and DiCE on four real-world datasets.

Results on more baselines and COMPAS dataset.

We provide additional experiments with a new dataset: Compas. We also conduct the experiments with three additional baselines: LORE (Guidotti et al., 2018) and CCHVAE (Pawelczyk et al., 2020) for non-actionability comparison and CRUDS (Downs et al., 2020) for actionability comparison. We report the results for the German and Compas datasets in Table 6 and Table 7. Our method outperforms LORE, CCHVAE, and CRUDS in terms of Anti-Diversity and DPP.

Table 6: Comparing our method (QUAD) versus LORE and CCHVAE on German and Compas data *without* actionability constraint.

Dataset	Methods	Cost	Validity	Anti Diversity	DPP	Manifold Distance
German	LORE	0.19 ± 0.05	0.87 ± 0.27	1.05 ± 0.74	0.03 ± 0.02	0.12 ± 0.10
	CCHVAE	0.49 ± 0.12	1.00 ± 0.00	0.95 ± 0.42	0.05 ± 0.02	0.25 ± 0.18
	FRPD-QUAD	0.30 ± 0.19	1.00 ± 0.00	-0.53 ± 0.70	0.14 ± 0.12	0.19 ± 0.19
Compas	LORE	0.31 ± 0.22	0.93 ± 0.19	1.13 ± 0.91	0.03 ± 0.05	0.08 ± 0.11
	CCHVAE	0.55 ± 0.45	1.00 ± 0.00	1.45 ± 1.05	0.07 ± 0.06	0.13 ± 0.17
	FRPD-QUAD	0.37 ± 0.25	1.00 ± 0.00	-0.71 ± 0.31	0.24 ± 0.20	0.07 ± 0.09

Table 7: Comparing our method (QUAD) versus FACE and CRUDS on two datasets with actionability constraint

Dataset	Methods	Shortest Path	Validity	Anti Diversity	DPP
German	FACE	0.24 ± 0.17	1.00 ± 0.00	2.18 ± 0.77	0.04 ± 0.04
	CRUDS	0.78 ± 0.32	1.00 ± 0.00	1.56 ± 0.42	0.11 ± 0.07
	FRPD-QUAD	0.55 ± 0.24	1.00 ± 0.00	-0.13 ± 0.24	0.21 ± 0.07
Compas	FACE	0.35 ± 0.26	1.00 ± 0.00	2.23 ± 0.83	0.07 ± 0.05
	CRUDS	0.89 ± 0.47	1.00 ± 0.00	1.92 ± 0.56	0.14 ± 0.09
	FRPD-QUAD	0.59 ± 0.34	1.00 ± 0.00	-0.25 ± 0.17	0.32 ± 0.12

A recourse plan example. We present an example of a recourse plan using the German credit dataset with $K = 3$ recourses between two methods without actionability constraints (DiCE and FRPD-QUAD using linear interpolation). Two approaches (FRPD-QUAD and DiCE) might present users with a diverse set of recourses to choose from. However, we can observe in Table 8 that recourse plans generated by our method are more realistic and suited for users to make changes. For example, DiCE suggests a recourse that involves changing the age from 22 to 20.1 years old, which is not a feasible change.

Table 8: An example of a recourse plan generated by FRPD-QUAD using linear interpolation and DiCE on the German Credit dataset.

	Duration	Credit Amount	Personal Status	Age
Input	12	1007	A94	22
FRPD-QUAD	24.2	2918.9	A94	22
	10.9	943.9	A94	22.2
	10.7	682.1	A94	22
DiCE	11.9	1304.4	A91	20.1
	4	250	A94	22
	15.4	1073.5	A94	32.5

A.4 Sequential Recourse

The original recourse with single shot recommendations (Russell, 2019; Mothilal et al., 2020) provide ‘what-if’ feedback for users. However, in practice, it is more desirable to provide a sequence of directive actions that users should take in order to achieve that recourse (Singh et al., 2021; Verma et al., 2022). Sequential actions are also more realistic than a one-step change in real-world applications (Ramakrishnan et al., 2020). We present an example of a sequential recourse plan with $K = 2$ recourses on the Bank dataset in Figure 7.

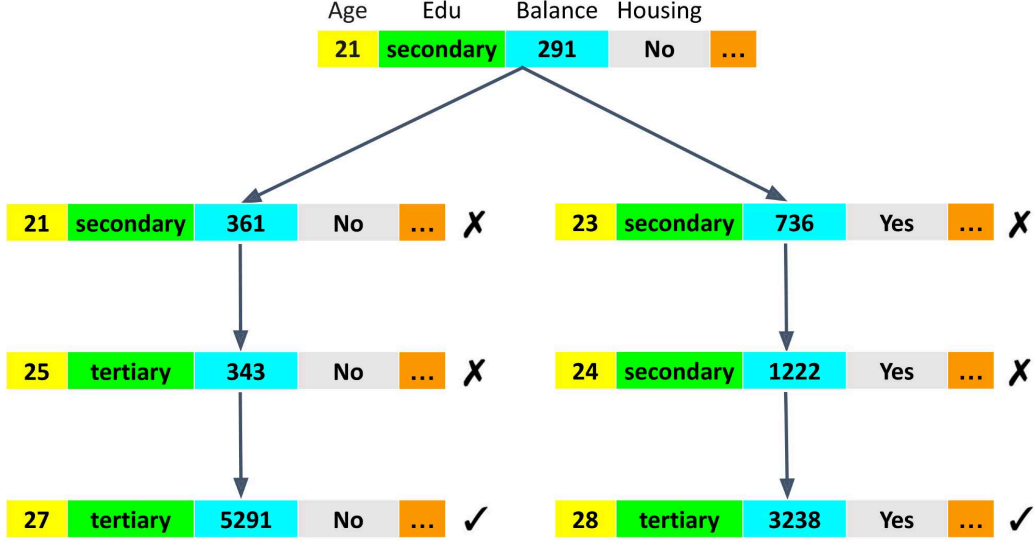


Figure 7: An example of a sequential recourse plan with $K = 2$ recourses on the Bank dataset. \times and \checkmark denote the unfavorable and favorable predicted outcomes, respectively.

Diversity of Sequential Recourses. Similar to the original recourse plan (Mothilal et al., 2020), it is also desirable to promote diversity among suggested sequential recourses in order to capture multiple users’ preferences. To evaluate the diversity of sequential recourse plans, we measure Path-Diversity using Levenshtein distance (Navarro, 2001) and Path-Anti-Diversity using Jaccard coefficient (Chondrogiannis et al., 2018).

- **Path Diversity:** The diversity of a sequential recourse plan is computed by

$$\text{Path-Diversity}(P_1, \dots, P_k) = \frac{1}{\binom{K}{2}} \sum_{\substack{i, j \in [K] \\ i < j}} \text{lev}(P_i, P_j).$$

Here, $\text{lev}(\cdot, \cdot)$ is the Levenshtein distance between two paths defined by sequences of nodes $P = (u_0, \dots, u_l)$ and $Q = (v_0, \dots, v_h)$. The Levenshtein distance is given by

$$\text{Lev}(P, Q) = \min \left\{ \begin{array}{l} d(v_h, v_{h-1}) + \text{Lev}(P, Q \setminus v_h), \\ d(u_l, u_{l-1}) + \text{Lev}(P \setminus u_l, Q), \\ d(u_l, v_h) + \text{Lev}(P \setminus u_l, Q \setminus v_h) \end{array} \right\},$$

where u_i, v_j are nodes on the graph \mathcal{G} and the function $d(\cdot, \cdot)$ is the Euclidean distance between node features. We set $\text{lev}(\emptyset, \emptyset) = 0$.

- **Path Anti Diversity:** The anti-diversity of a sequential recourse plan is the average Jaccard coefficient of all pairs of sequential recourses:

$$\text{Path-Anti-Diversity}(P_1, \dots, P_k) = \frac{1}{\binom{K}{2}} \sum_{\substack{i, j \in [K] \\ i < j}} \text{Jac}(P_i, P_j).$$

Here, the Jaccard coefficient defined on edges of two paths $P = (e_0, \dots, e_l)$ and $Q = (o_0, \dots, o_h)$ is given by

$$\text{Jac}(P, Q) = \frac{\sum_{e \in P \cap Q} d(e)}{\sum_{e \in P \cup Q} d(e)},$$

where $d(e)$ is the Euclidean distance between two vertices of the edge e .

Table 9 reports the Path-Diversity, Path-Anti-Diversity, and cost of recourse by the shortest paths connecting the input instance to the prototypes. It can be seen that sequential recourse plans generated by our methods (FRPD-QUAD, FRPD-DPP-GR, and FRPD-DPP-LS) are more diverse than those generated by FACE.

Table 9: Benchmark of Path-Diversity and Path-Anti-Diversity between competing methods with actionability constraints on four real-world datasets.

Dataset	Methods	Shortest Path Cost	Path Diversity	Path Anti Diversity
German	FACE	0.24 \pm 0.17	1.07 \pm 0.20	0.06 \pm 0.12
	FRPD-QUAD	0.55 \pm 0.24	2.00 \pm 0.67	0.05 \pm 0.07
	FRPD-DPP-GR	0.42 \pm 0.20	1.77 \pm 0.72	0.02 \pm 0.06
	FRPD-DPP-LS	0.42 \pm 0.20	1.77 \pm 0.72	0.02 \pm 0.06
SBA	FACE	0.25 \pm 0.20	1.90 \pm 0.70	0.06 \pm 0.08
	FRPD-QUAD	0.55 \pm 0.25	5.03 \pm 1.43	0.00 \pm 0.00
	FRPD-DPP-GR	0.52 \pm 0.25	4.47 \pm 1.54	0.00 \pm 0.00
	FRPD-DPP-LS	0.52 \pm 0.25	4.47 \pm 1.54	0.00 \pm 0.00
Bank	FACE	1.29 \pm 0.29	2.47 \pm 0.99	0.04 \pm 0.07
	FRPD-QUAD	1.55 \pm 0.44	4.17 \pm 0.92	0.01 \pm 0.02
	FRPD-DPP-GR	1.52 \pm 0.42	4.50 \pm 1.38	0.00 \pm 0.00
	FRPD-DPP-LS	1.52 \pm 0.42	4.50 \pm 1.38	0.00 \pm 0.00
Adult	FACE	1.33 \pm 0.36	1.57 \pm 0.83	0.12 \pm 0.28
	FRPD-QUAD	1.99 \pm 0.59	3.80 \pm 0.50	0.01 \pm 0.02
	FRPD-DPP-GR	2.00 \pm 0.52	3.87 \pm 0.43	0.05 \pm 0.13
	FRPD-DPP-LS	2.00 \pm 0.52	3.87 \pm 0.43	0.05 \pm 0.13

A.5 Comparisons of Prototype Selection Algorithms

Greedy and Local search DPP. Figure 8 presents the additional improvement of the solution quality, measured by the determinant value in the objective function of (2), that is generated by the local neighborhood search. We observe that across all datasets, the local search constantly improves the quality of the greedy heuristics.

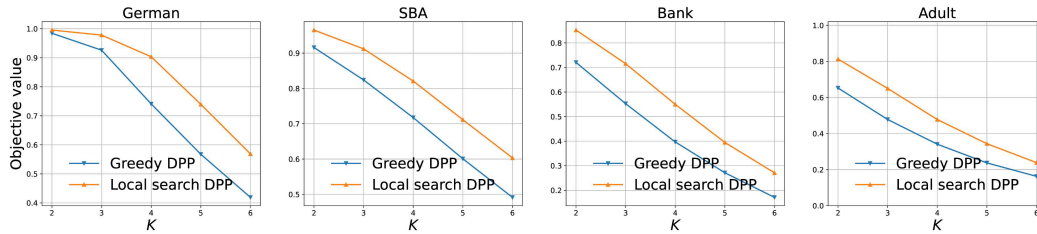


Figure 8: Average objective value (higher is better) of Greedy solution and Greedy+Local search solution with a different number of prototypes K , the average is taken over 100 randomly chosen input x_0 from the test set.

Greedy and Eigen-Approximate Binary Quadratic Program. We also implement a greedy heuristic search to solve problem (3). This greedy algorithm proceeds by finding iteratively for each incumbent set of prototypes z an index j by

$$j = \underset{i: z_i=0}{\operatorname{argmin}} \vartheta(z'_i)^\top S z'_i + (1 - \vartheta) d^\top z'_i, \quad \text{where } z'_i = z \vee \bar{e}_i$$

with \vee denoting the element-wise maximum between two vectors and \bar{e}_i is the vector of zeros with the i -th element being one. The algorithm then adds j to the set of prototypes until reaching the cardinality constraint, i.e., when $\|z\|_0 = K$. In doing so, each j guarantees to maximize the marginal gain to the incumbent set.

We compare the solution of the greedy method against our approximate solution where the quality of the solution is measured by the objective value of problem (3).

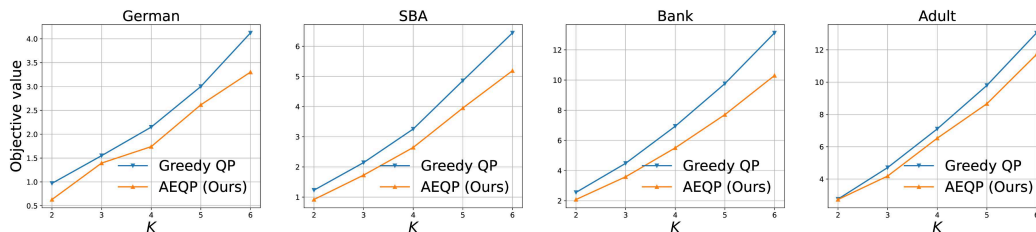


Figure 9: Average objective value (lower is better) of Greedy solution and Eigen-Approximate Binary Quadratic Program solution with a different number of prototypes K , the average is taken over 100 randomly chosen input x_0 from the test set.

Comparison of run time. We compare the run time of our three approaches (FRPD-QUAD, FRPD-DPP-GR, FRPD-DPP-LS) on synthetic datasets. We generate 2-dimensional data samples by sampling uniformly in a rectangle $x = (x_1, x_2) \in [-2, 4] \times [-2, 7]$ with the following binary labelling function f :

$$f(x) = \begin{cases} 1 & \text{if } x_2 \geq 1 + x_1 + 2x_1^2 + x_1^3 - x_1^4, \\ 0 & \text{otherwise.} \end{cases}$$

First, we synthesize N 2-dimensional data samples for each value $N = 100, \dots, 10000$. For each negatively predicted instance, we find $K = 3$ prototypes in the positively predicted class. The process is repeated independently five times, then we take the average run time of our three methods and report the results in Figure 10. These results indicate that the Greedy DPP has the smallest run time. The run time of the Local search DPP increases significantly when the number of samples increases. Furthermore, these findings demonstrate that our FRPD-QUAD approach is appropriate for the Anti-Diversity measure, as it has a comparable run time to FRPD-DPP-GR and the lowest Anti-Diversity in all the experiments.

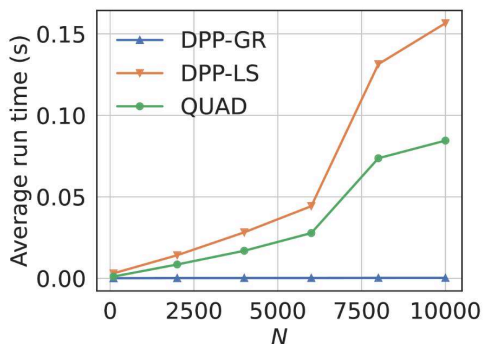


Figure 10: Average run time(s) of three different approaches with different sample sizes N over 5 independent replications for FRPD-QUAD (QUAD), FRPD-DPP-GR (DPP-GR) and FRPD-DPP-LS (DPP-LS).

Additionally, we evaluate the run time of FRPD-QUAD (QUAD), QUAD-Greedy (QUAD-GR), and QUAD-Local search (QUAD-LS) and report the results in Figure 11. These results demonstrate that the Greedy method has the smallest run time. The run time of the Local search quadratic increases significantly when the number of samples increases.

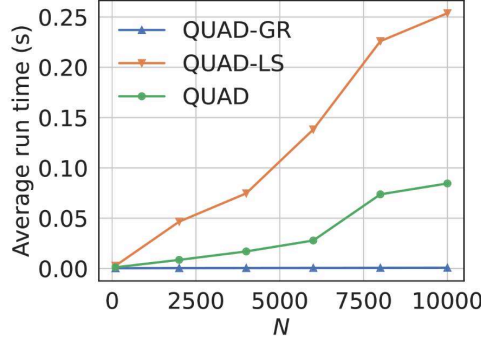


Figure 11: Average run time(s) of three different approaches with different sample sizes N over 5 independent replications for FRPD-QUAD (QUAD), QUAD-Greedy (QUAD-GR), and QUAD-Local search (QUAD-LS).

B Background on Determinant Point Processes

We first provide a definition of a DPP.

Definition B.1 (Determinantal point process). *Given a set $[N] \triangleq \{1, \dots, N\}$ containing N items, a DPP defines a probability measure over all subsets of $[N]$ that is parametrized by a matrix $P \in \mathbb{S}_+^N$, $0 \preceq P \preceq I$. If \tilde{R} is a random subset drawn according to this DPP, then for every subset $J \subseteq [N]$,*

$$\text{Probability}_P(J \subseteq \tilde{R}) = \det(P_J),$$

where P_J is a submatrix of P obtained by restricting to rows and columns indexed in J .

The bound conditions $0 \preceq P \preceq I$ ensure that all principal minors of P are nonnegative and smaller than one, which is sufficient to define a proper probability distribution. The matrix P is called the marginal kernel. If $J = \{i\}$ is a singleton set, then $\text{Probability}_P(i \in \tilde{R}) = P_{ii}$. If $J = \{i, j\}$ has two items, then

$$\text{Probability}_P(i, j \in \tilde{R}) = \det \begin{pmatrix} P_{ii} & P_{ij} \\ P_{ji} & P_{jj} \end{pmatrix} = P_{ii}P_{jj} - P_{ij}P_{ji}.$$

The term $-P_{ij}P_{ji}$ captures the relationship between two items i and j . Since P is symmetric, we have $-P_{ij}P_{ji} = -P_{ij}^2 \leq 0$; thus DPPs are able to model the negative correlations between items. The larger value of P_{ij} , the lower probability of items i and j co-occurs. If $P_{ij} = 0$, there is no interaction between the two items.

In practice, it is difficult to deal with a generic correlation kernel P . Therefore, L -ensemble, a subclass of DPPs, is often used instead since it provides several simpler formulas. For a positive semidefinite matrix $L \in \mathbb{S}_+^N$, an L -ensemble DPP specifies the atomic probabilities for every possible instantiation of \tilde{R} via

$$\text{Probability}_L(\tilde{R} = J) \propto \det(L_J) \quad \forall J \subseteq [N].$$

The normalization constant can be computed as $\sum_{J \subseteq [N]} \det(L_J) = \det(L + I)$. Here, a positive semidefinite matrix L is admissible to define a DPP with a marginal kernel matrix $P = I - (L + I)^{-1}$. Typically, the matrix L is chosen as a similarity matrix, in which the diagonal element L_{ii} represents the quality of item i while the off-diagonal element L_{ij} is the similarity measure between i and j .

C Local Search for the Maximum A Posteriori (MAP) Inference

We describe in this section a neighborhood heuristic that can be employed to improve the greedy algorithm of the MAP inference problem in Section 2.

Let $z \in \{0, 1\}^N$, $\|z\|_0 = K$ be the current solution of the local search algorithm. We define the local move for the solution z by removing the element j^- that has the *smallest* marginal decrease in the objective value and then adding the element

j^+ that engenders the *largest* marginal increase in the objective value. In detail, we define

$$j^- \triangleq \operatorname{argmin}_{i:z_i=1} \log \det(L_{z \wedge \bar{e}_i}) - \log \det(L_z)$$

and compute

$$j^+ \triangleq \operatorname{argmax}_{i:(z \wedge \bar{e}_{j^-})_i=0} \log \det(L_{z \wedge \bar{e}_{j^-} \vee e_i}) - \log \det(L_{z \wedge \bar{e}_{j^-}}),$$

where \bar{e}_i is the vector of ones with the i -th element being zero. The operator \wedge is the bitwise OR and \vee is the bitwise AND operator. Computing j^- costs $\mathcal{O}(K^4)$ in running time while we can find j^+ in $\mathcal{O}(NK^2)$ time using similar procedure as in Chen et al. (2018). For $K \ll N$, the total complexity for this local move is $\mathcal{O}(NK^2)$.

For the local search procedure, we first initialize the solution using the output of the greedy algorithm and then iteratively perform the local moves. This procedure terminates when there is no neighboring solution that can improve the objective value.

D Proof

We provide here the proof of Lemma 3.1 that is omitted in the main text.

Proof of Lemma 3.1. If we add the auxiliary variables h_m and the resulting constraints $h_m = v_m^\top z$, then problem (4) becomes

$$\begin{aligned} \min \quad & (1 - \vartheta)d^\top z + \vartheta \sum_{m=1}^M \sigma_m h_m^2 \\ \text{s. t.} \quad & h \in \mathbb{R}^M, z \in \{0, 1\}^N \\ & \|z\|_0 = K, \quad h_m = v_m^\top z \quad \forall m = 1, \dots, M. \end{aligned}$$

This optimization problem can be decomposed into a two-layer problem of the form

$$\begin{aligned} \min_{z \in \{0, 1\}^N, \|z\|_0=K} \quad & \min (1 - \vartheta)d^\top z + \vartheta \sum_{m=1}^M \sigma_m h_m^2 \\ \text{s. t.} \quad & h \in \mathbb{R}^M, h_m = v_m^\top z \quad \forall m = 1, \dots, M. \end{aligned}$$

For any feasible solution z , the inner minimization problem over h is a convex optimization problem. By strong duality, we have the equivalent problem

$$\min_{\substack{z \in \{0, 1\}^N \\ \|z\|_0=K}} \max_{\gamma \in \mathbb{R}^M} \min_{h \in \mathbb{R}^M} (1 - \vartheta)d^\top z + \vartheta \sum_{m=1}^M \sigma_m h_m^2 + 2 \sum_{m=1}^M \gamma_m (h_m - v_m^\top z)$$

The optimization problem in h is separable, and because $\sigma_m > 0$, the optimal solution is

$$h_m^* = -\frac{\gamma_m}{\vartheta \sigma_m} \quad \forall m = 1, \dots, M.$$

By replacing this optimal solution with the objective function, we thus obtain the equivalent problem. This completes the proof. \square