Dimension-Free Parameterized Approximation Schemes for Hybrid Clustering

Ameet Gadekar

CISPA Helmholtz Center for Information Security
Saarbrücken, Germany

Tanmay Inamdar

Indian Institute of Technology Jodhpur
Jodhpur, India

07.03.2025

STACS 2025

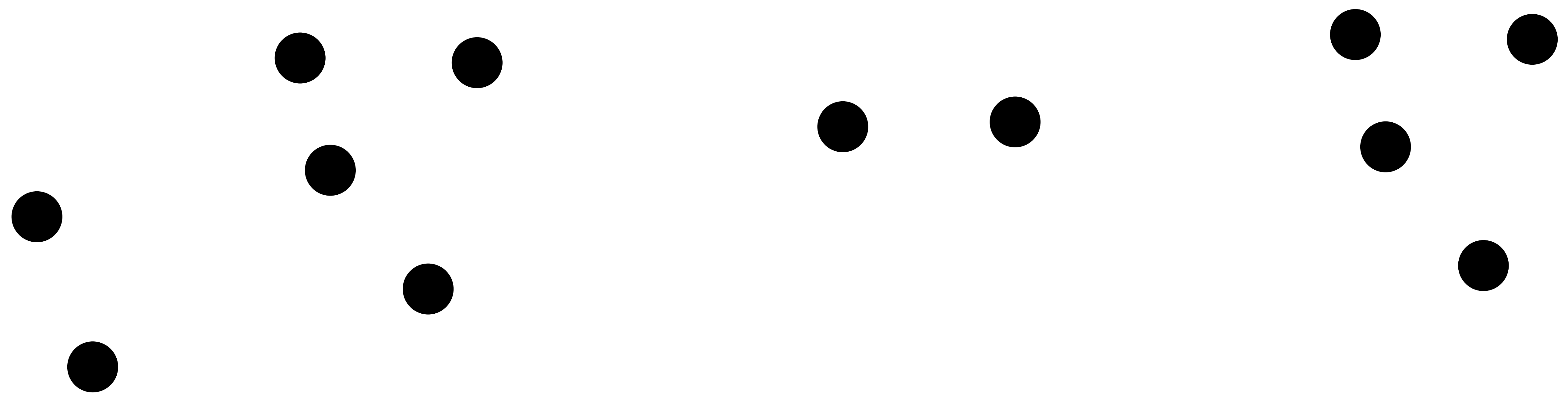
Jena, Germany

Clustering

Given a set of objects

Want to group them such that

objects in the same group are more “similar” to each other than to those in the other groups

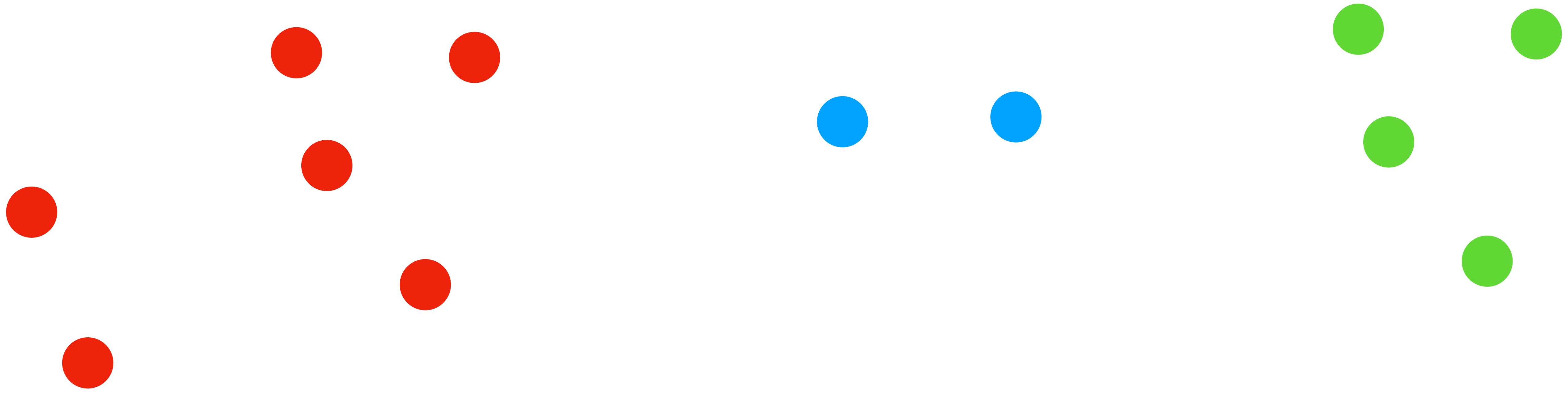


Clustering

Given a set of objects

Want to group them such that

objects in the same group are more “similar” to each other than to those in the other groups

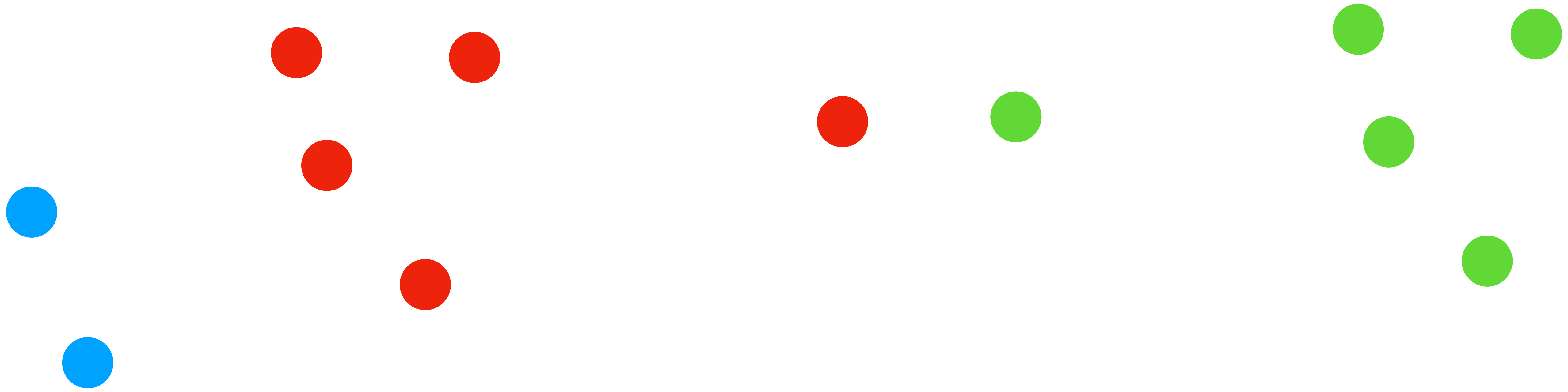


Clustering

Given a set of objects

Want to group them such that

objects in the same group are more “similar” to each other than to those in the other groups



Groups are called Clusters

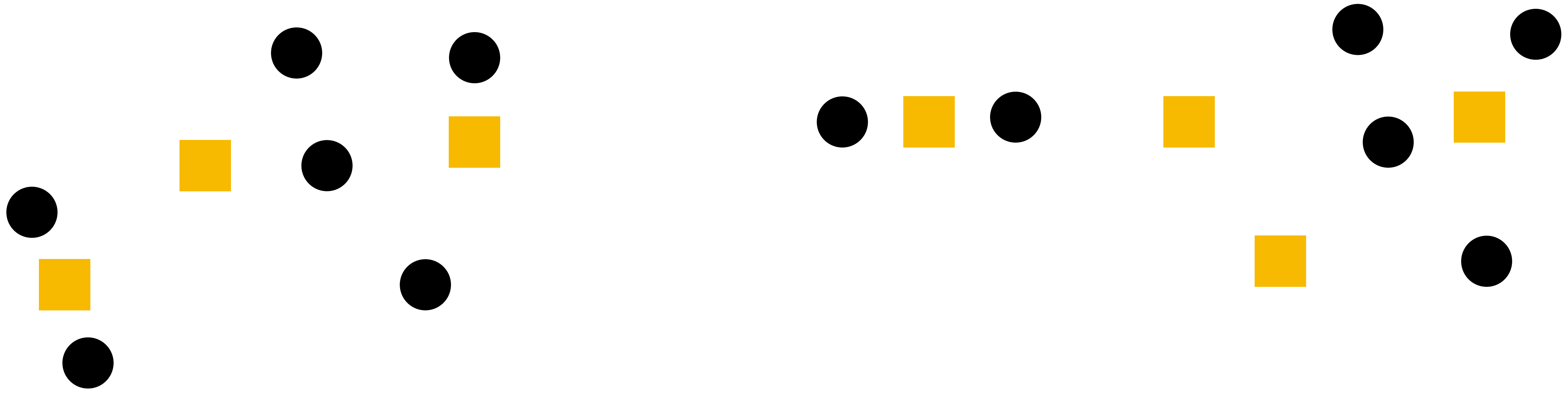
Center-based Clustering

Set of Objects — Points/Clients

Set of potential centers — Facilities

Want to choose centers and assign every point to a closest center

to minimize a clustering objective



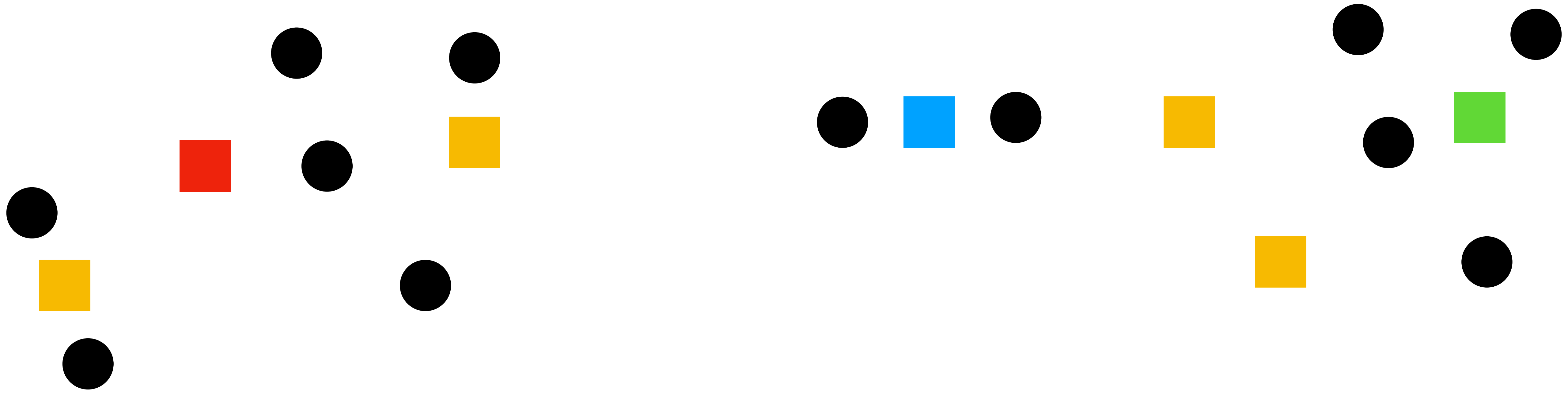
Center-based Clustering

Set of Objects — Points/Clients

Set of potential centers — Facilities

Want to choose centers and assign every point to a closest center

to minimize a clustering objective



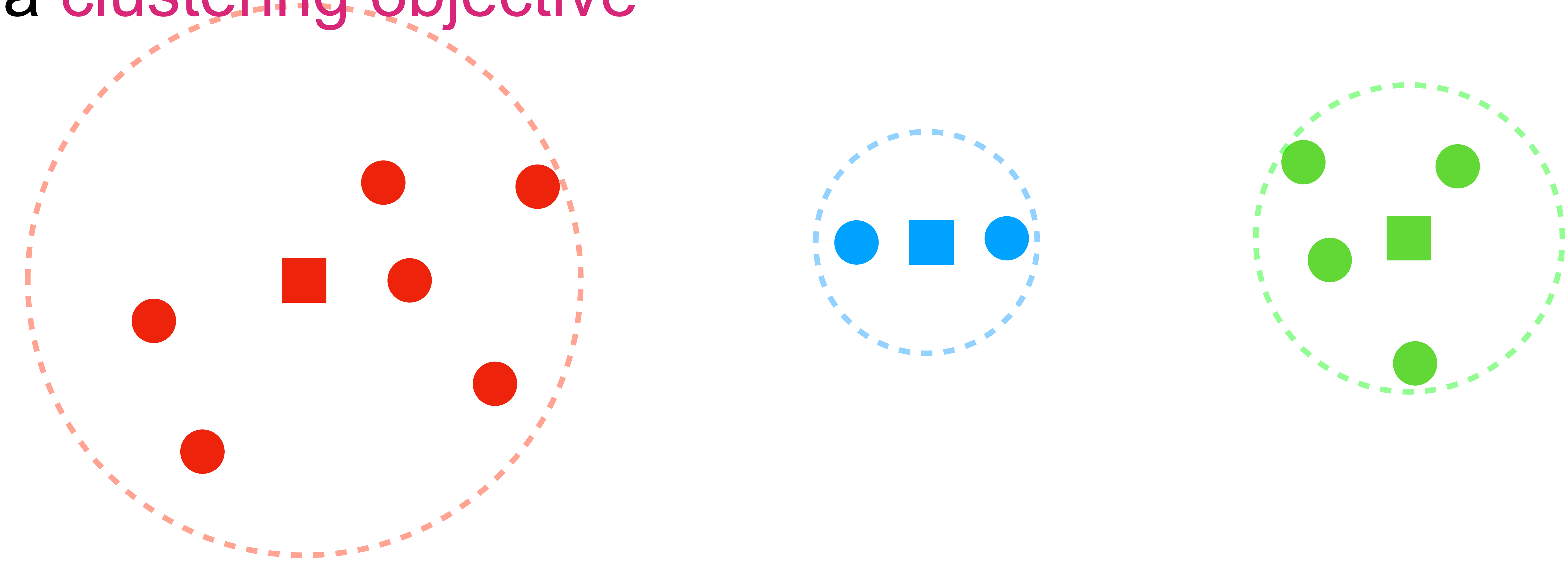
Center-based Clustering

Set of Objects — Points/Clients

Set of potential centers — Facilities

Want to choose centers and assign every point to a closest center

to minimize a clustering objective



Center-based Clustering

- Input

P: set of n points

F: set of facilities

d: distance function on $P \cup F$

k: positive integer

- Output

$X \subseteq F$: set of k centers

- Minimize an objective

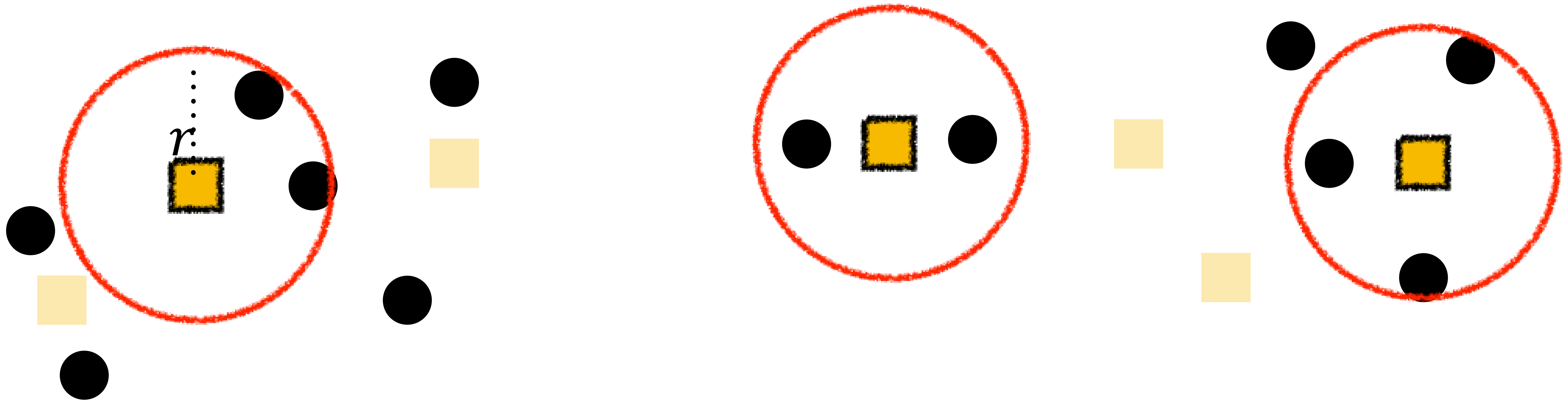
k-Median: $\sum_{p \in P} d(p, X)$

k-Center: $\max_{p \in P} d(p, X)$

k-Means: $\sum_{p \in P} d(p, X)^2$

Hybrid Clustering

- Interpolates between k-Median and k-Center
- Think of placing k WiFi routers, each with coverage radius r
- Clients within coverage, pay 0 (zero)
- Clients outside coverage, pay the distance to the nearest ball



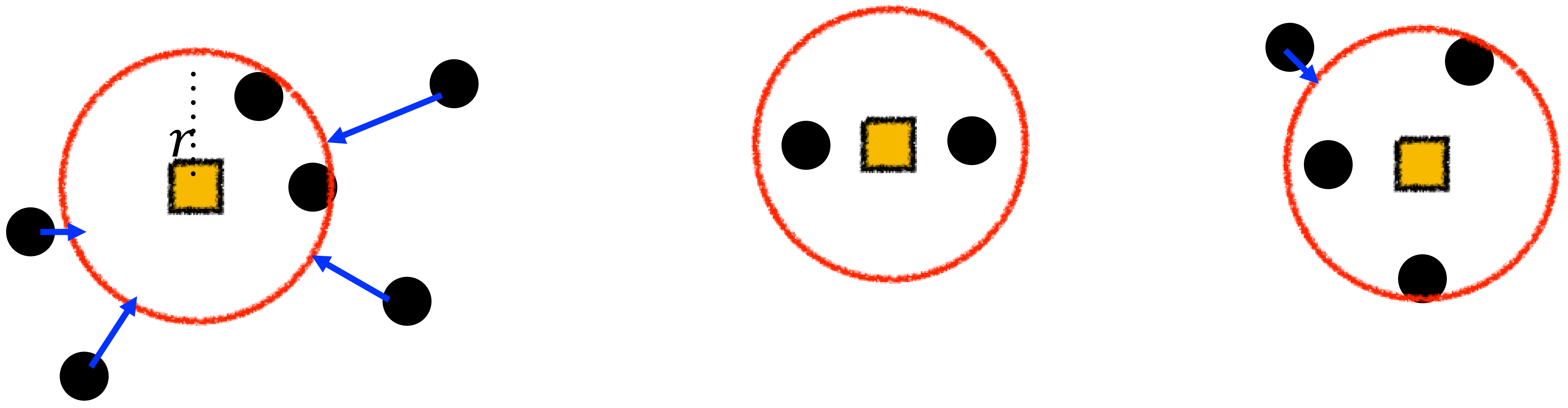
Hybrid Clustering

- Interpolates between k-Median and k-Center

- Think of placing k WiFi routers, each with coverage radius r

- Clients within coverage, pay 0 (zero)

- Clients outside coverage, pay the distance to the nearest ball



$d_r(p, X) := max\{d(p, X) - r, 0\}$ for $X \subseteq F$

Hybrid Clustering

- $d_r(p, X) := max\{d(p, X) - r, 0\}$ r-distance

- Input

  P: set of n points

  F: set of facilities

  k: positive integer

  d: distance function on $P \cup F$

  r: non-negative real

- Output

  $X \subseteq F$: set of k centers

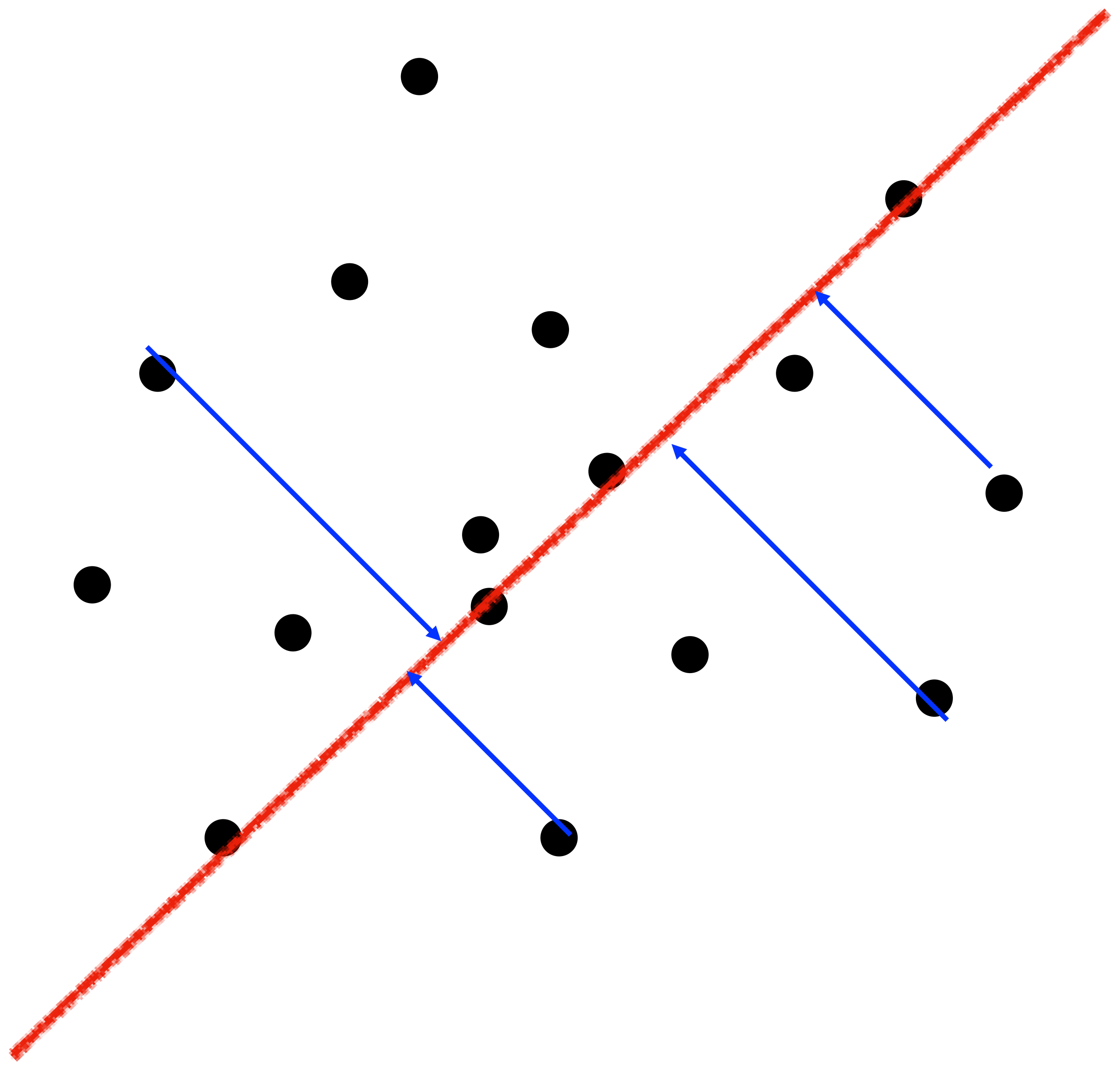
- Minimize

$$\sum_{p \in P} d_r(p, X)$$

Hybrid Clustering

Motivation

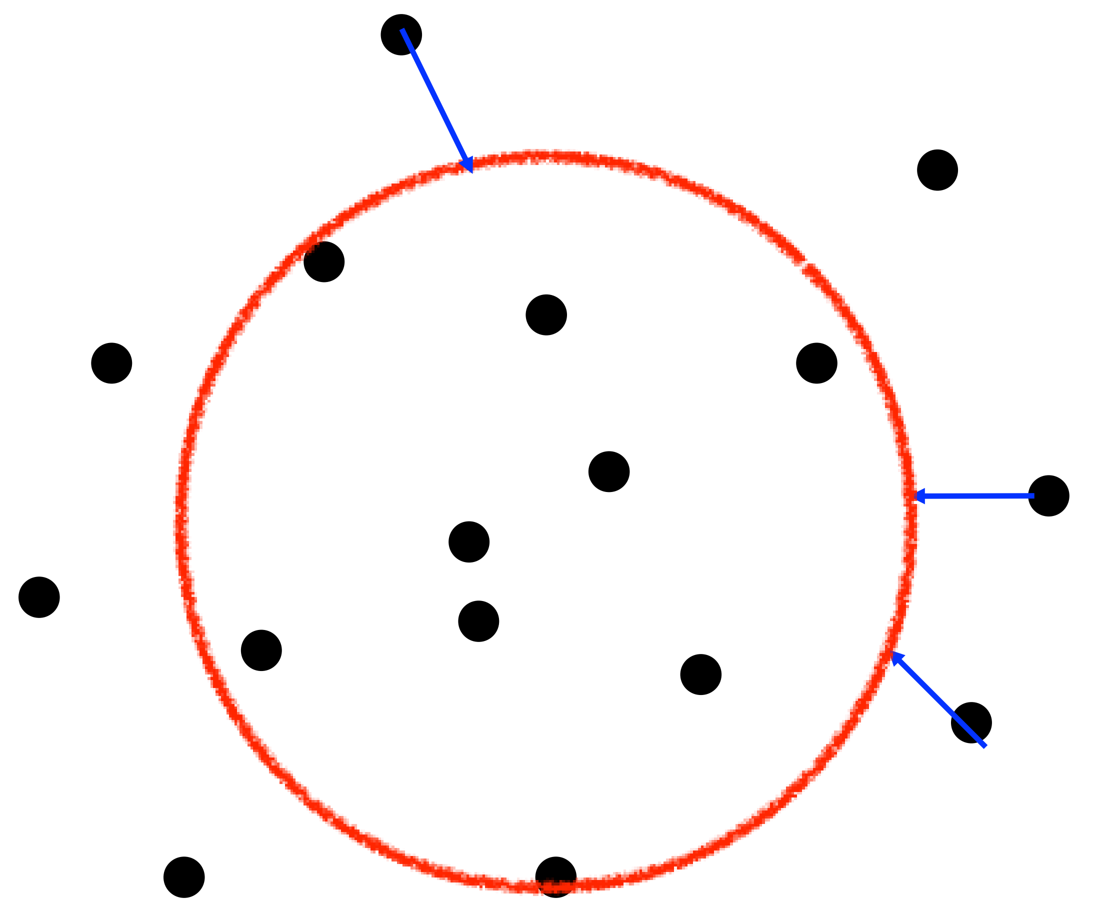
- Interpolates between k-Median and k-Center
- Shape Fitting
  - Extension of Linear regression: Fitting “best” lines
  - Projective Clustering: Fitting “best” affine spaces



Hybrid Clustering

Motivation

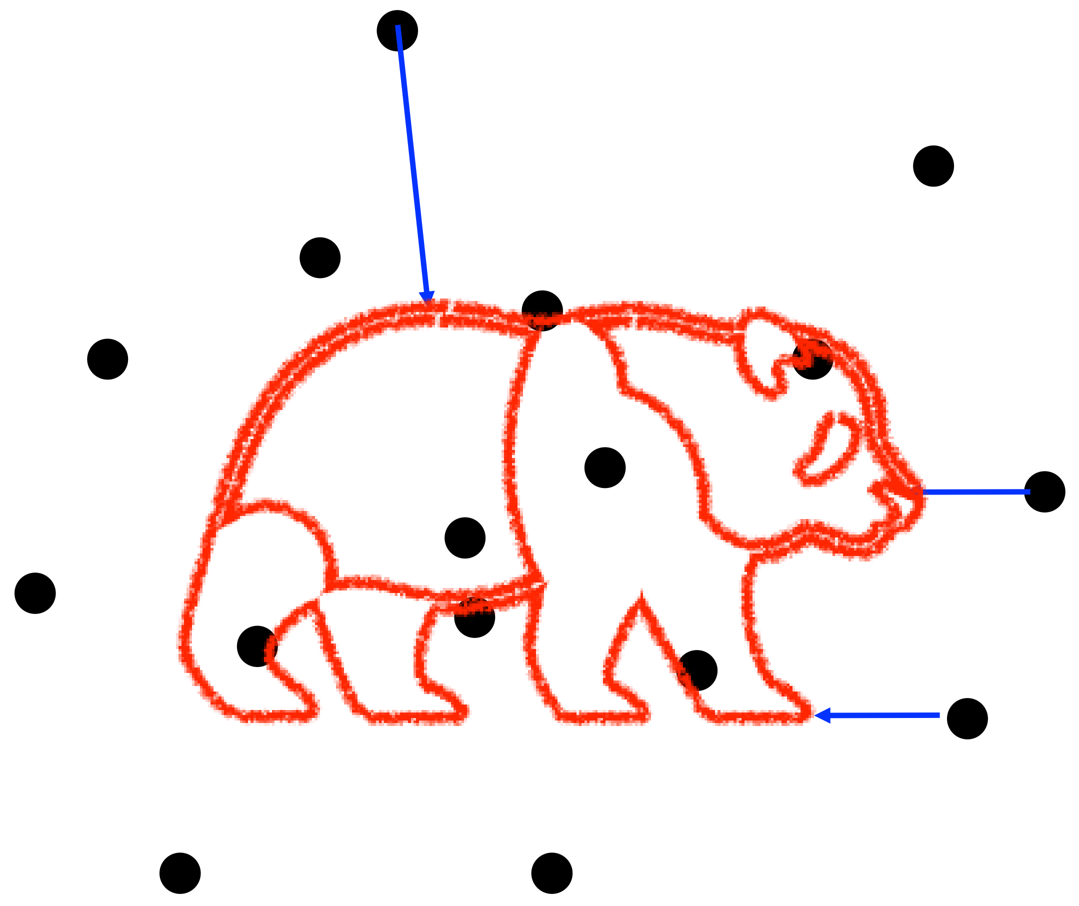
- Interpolates between k-Median and k-Center
- Shape Fitting
  - Extension of Linear regression: Fitting “best” lines
  - Projective Clustering: Fitting “best” affine spaces
  - Hybrid Clustering: Fitting “best” r-radius balls



Hybrid Clustering

Motivation

- Interpolates between k-Median and k-Center
- Shape Fitting
  - Extension of Linear regression: Fitting “best” lines
  - Projective Clustering: Fitting “best” affine spaces
  - Hybrid Clustering: Fitting “best” r-radius balls



Literature

- Recently introduced by Fomin, Golovach, Inamdar, Saurabh, Zehavi [Approx' 24]
- $r = 0: d_r(p, X) = d(p, X) \Longrightarrow k$-Median
- $r = OPT_{kc}: \sum d_r(p, X) = 0 \Longrightarrow k$-Center $\qquad OPT_{kc} = k$-Center OPT

- No Uni-criteria approximations: have to violate both—cost & radius

Literature

- Recently introduced by Fomin, Golovach, Inamdar, Saurabh, Zehavi [Approx’ 24]
- No Uni-criteria approximations: have to violate both—cost & radius

Literature

- Recently introduced by Fomin, Golovach, Inamdar, Saurabh, Zehavi [Approx’ 24]

- No Uni-criteria approximations: have to violate both—cost & radius

- Studied the problem in $\mathbb{R}^d$, where centers can be chosen anywhere

- For $\mathbb{R}^d$,designed $(1+\epsilon, 1+\epsilon)$-bicritera approximation

  OPT cost using r-radius balls

  - whose cost using $(1+\epsilon)r$-radius balls is at most $(1+\epsilon)\mathrm{OPT}_r$
  - in time $\mathrm{FPT}(k, d, \epsilon)$

Our Results

Theorem 1.

Substantially improve and generalize the results of Fomin at al.

Our Results

no d here

Theorem 1.

For $\mathbb{R}^d$,design $(1+\epsilon, 1+\epsilon)$-bicritera approximation in time $FPT(k,\epsilon)$

Fedor et al [Approx'24].

For $\mathbb{R}^d$,design $(1+\epsilon, 1+\epsilon)$-bicritera approximation in time $2^{(kd/\epsilon)^{O(1)}}n^{O(1)}$

Our Results

Theorem 1.

no d here

For $\mathbb{R}^d$,design $(1+\epsilon, 1+\epsilon)$-bicritera approximation in time $2^{\tilde{O}(k/\epsilon^5)}n^{O(1)}$

Fedor et al [Approx'24].

For $\mathbb{R}^d$,design $(1+\epsilon, 1+\epsilon)$-bicritera approximation in time $2^{(kd/\epsilon)^{O(1)}}n^{O(1)}$

Our Results

Theorem 1.

For $\mathbb{R}^d$,design $(1+\epsilon, 1+\epsilon)$-bicritera approximation in time $FPT(k,\epsilon)$

no d here

Works for metric spaces with bounded (algorithmic) scatter dimension

Bounded Doubling

Bounded Treewidth

Planar

Minor-closed

Works even when the objective is a monotone norm of r-distances

Generalizes the FOCS'23 framework of Abbasi* et al. to r-distances

* *Parameterized Approximation Schemes for Clustering with General Norm Objectives*
Abbasi, Banerjee, Byrka, Chalermsook, G., Khodamoradi, Marx, Sharma, Spoerhase

Our Results

no d here

Theorem 1.

For $\mathbb{R}^d$,design $(1+\epsilon, 1+\epsilon)$-bicritera approximation in time $FPT(k,\epsilon)$

Generalizes the FOCS'23 framework of Abbasi* et al. to r-distances

Theorem 2.

Design coresets of size $2^{O(d\log(1/\epsilon))}k\log n$ in doubling metrics of dimension d

* *Parameterized Approximation Schemes for Clustering with General Norm Objectives*
Abbasi, Banerjee, Byrka, Chalermsook, G., Khodamoradi, Marx, Sharma, Spoerhase

This talk

Theorem 1.

For $\mathbb{R}^d$,design $(1+\epsilon, 1+\epsilon)$-bicritera approximation in time $FPT(k,\epsilon)$

This talk

Theorem 1.

For $\mathbb{R}^d$,design $(1+\epsilon, 1+\epsilon)$-bicritera approximation in time $\mathit{FPT}(k,\epsilon)$

- Idea based on EPAS framework of Abbasi et al. [FOCS'23], Unified-EPAS
  - $(1+\epsilon)$-approximation running in time $\mathit{FPT}(k,\epsilon)$
  - for many clustering problems
  - under any metric space that has bounded (algorithmic) scatter dimension
  - in a unified manner

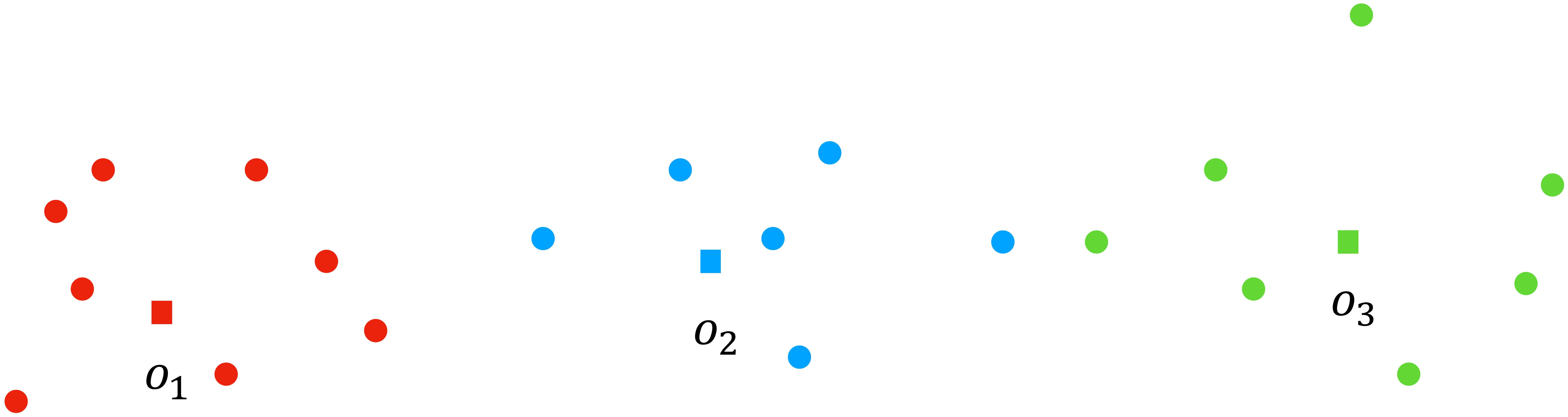
EPAS: Efficient Parameterized Approximation Schemes
FPT-AS

Unified-EPAS: Basic Idea

Consider the clustering corresponding to an optimal solution O

For each cluster $j \in [k]$, we maintain a cluster constraint Q_j

Each Q_j is a sequence of pairs (p, r_p), where p ∈Cluster j and $r_p \leq d(p, O)$

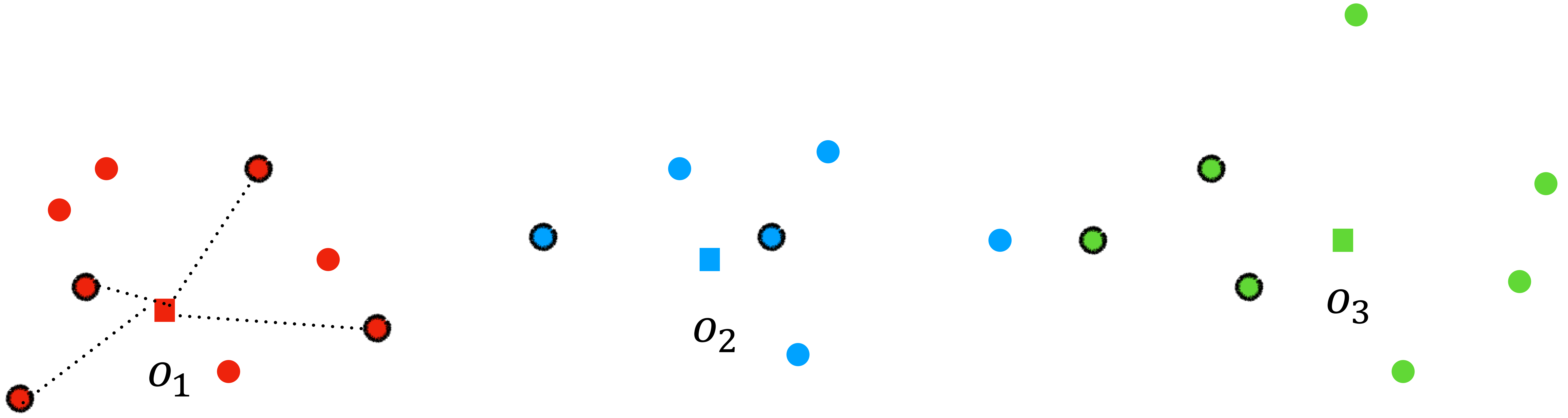


Unified-EPAS: Basic Idea

Consider the clustering corresponding to an optimal solution O

For each cluster $j \in [k]$, we maintain a cluster constraint Q_j

Each Q_j is a sequence of pairs (p, r_p), where p ∈Cluster j and $r_p \leq d(p, O)$



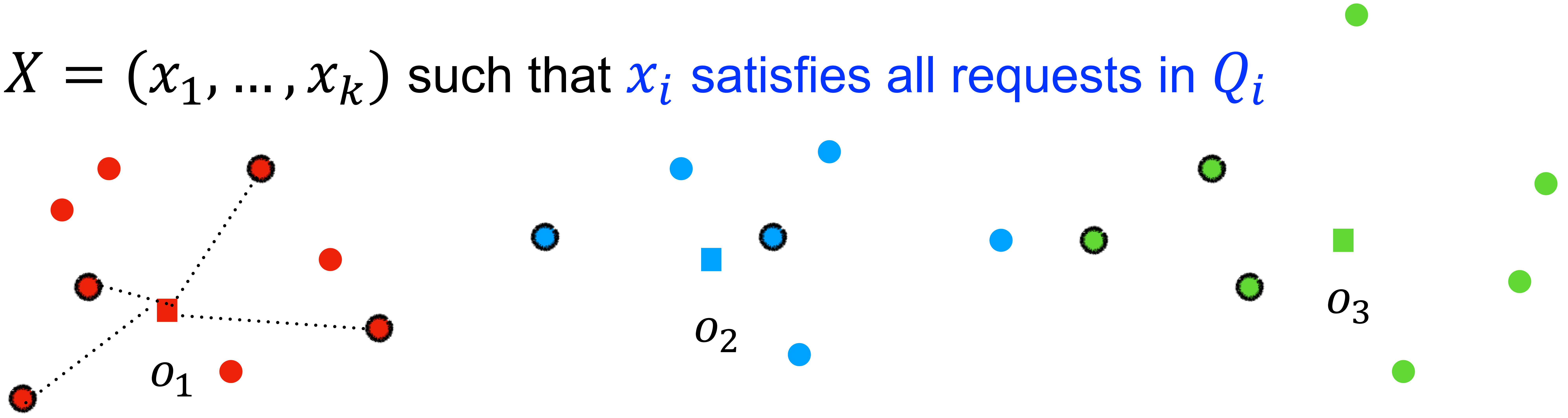
Unified-EPAS: Basic Idea

Consider the clustering corresponding to an optimal solution O

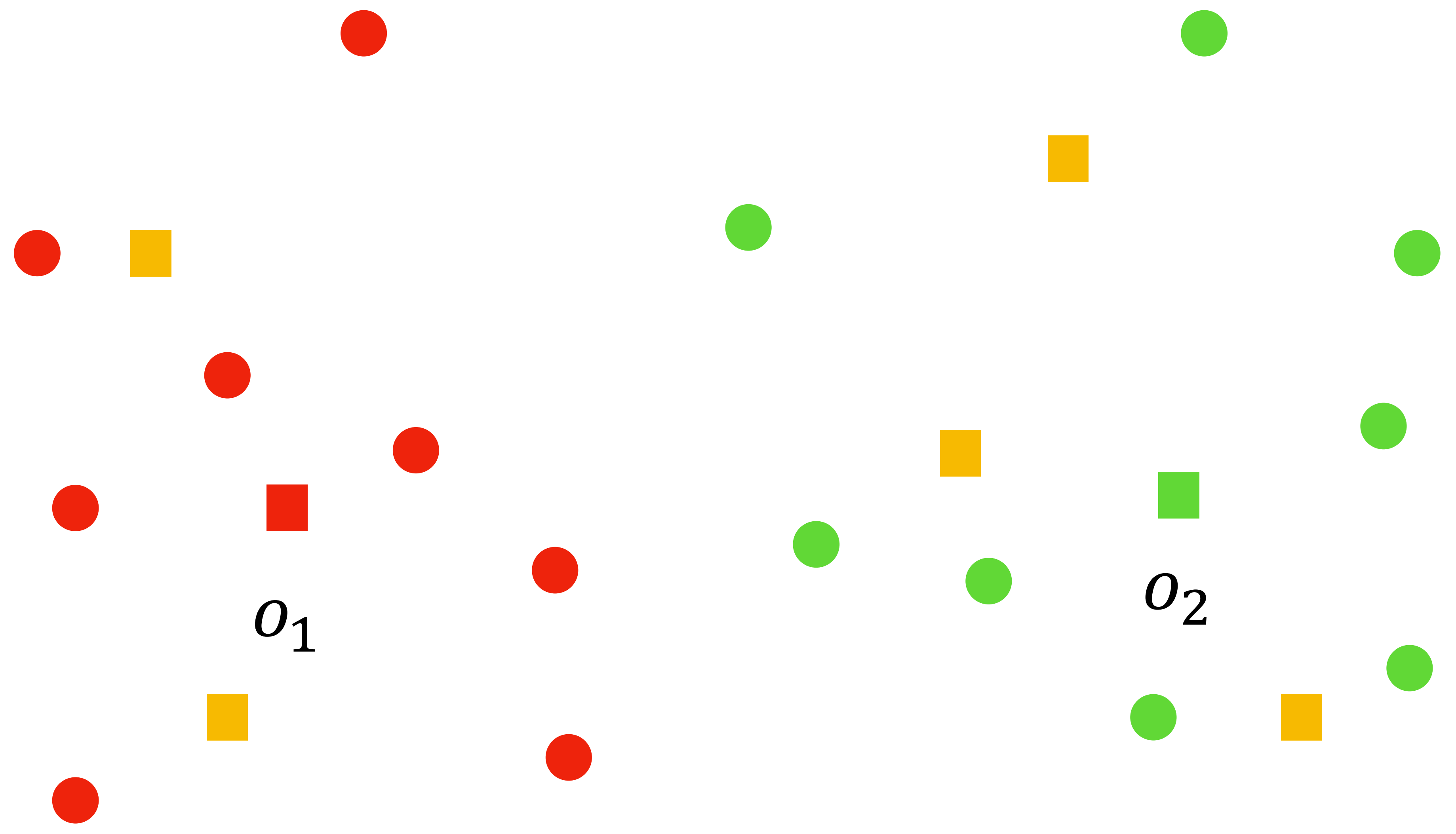
For each cluster $j \in [k]$, we maintain a cluster constraint Q_j

Each Q_j is a sequence of pairs (p, r_p), where p ∈Cluster j and $r_p \leq d(p, O)$

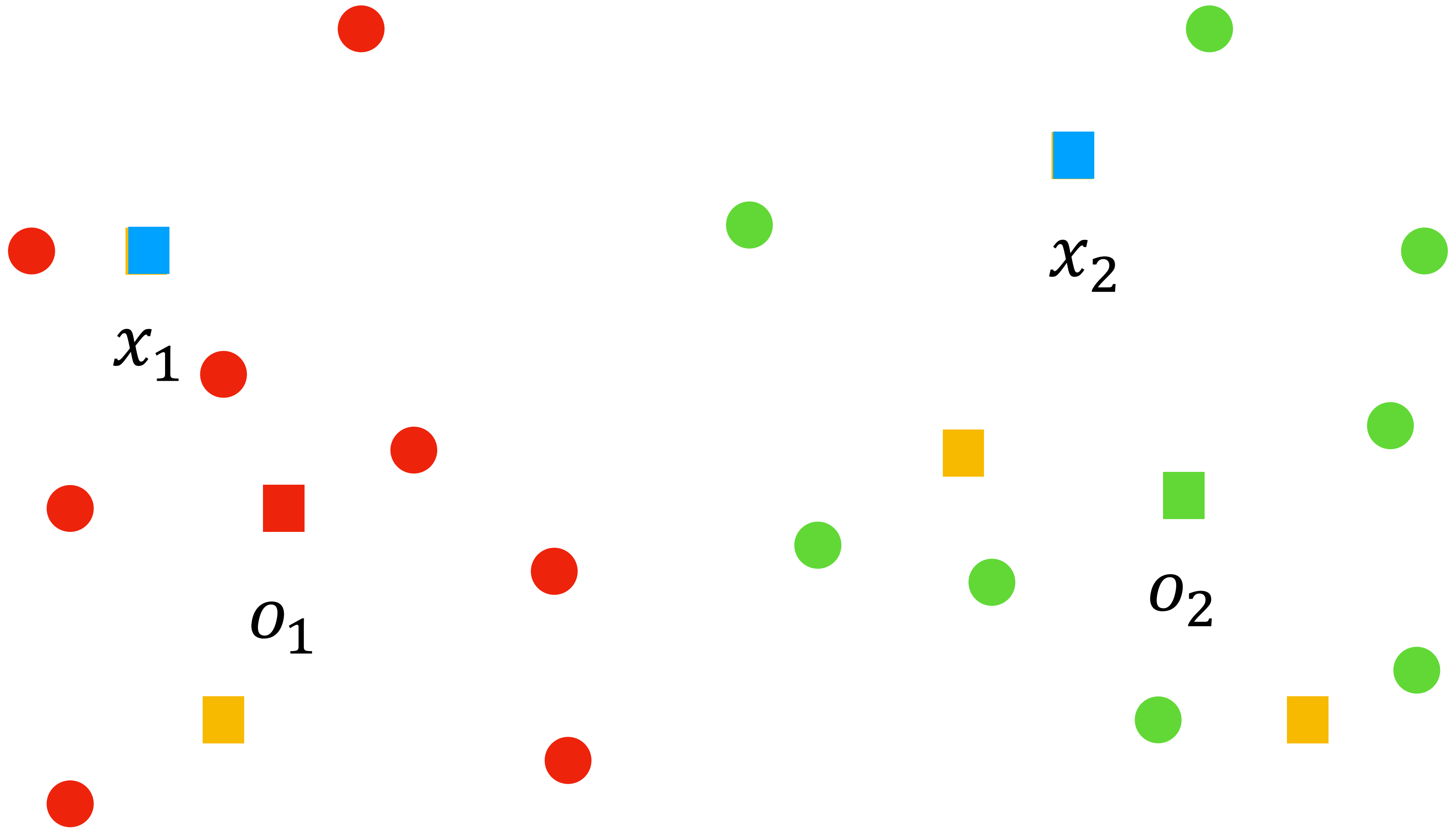
Find $X = (x_1, \ldots, x_k)$ such that x_i satisfies all requests in Q_i

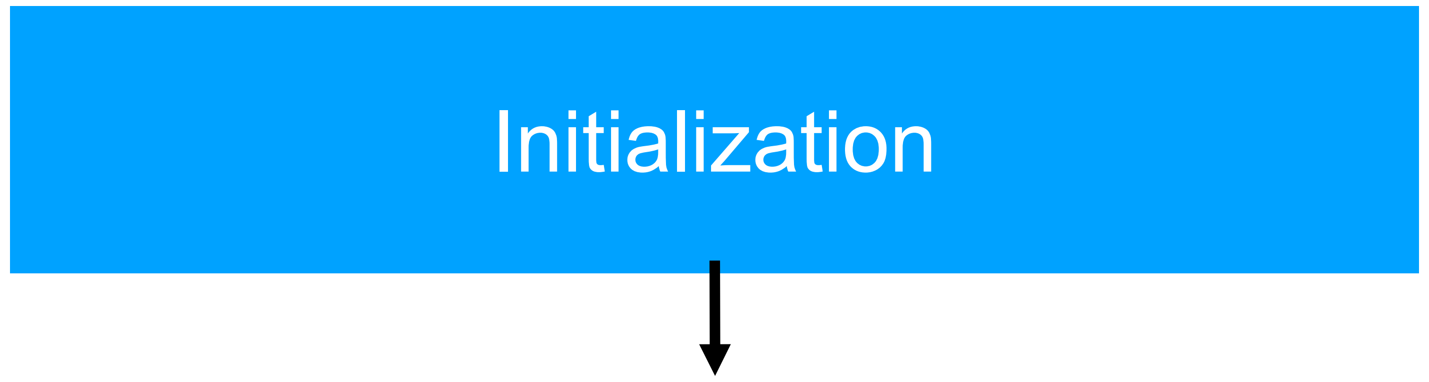


Unified-EPAS

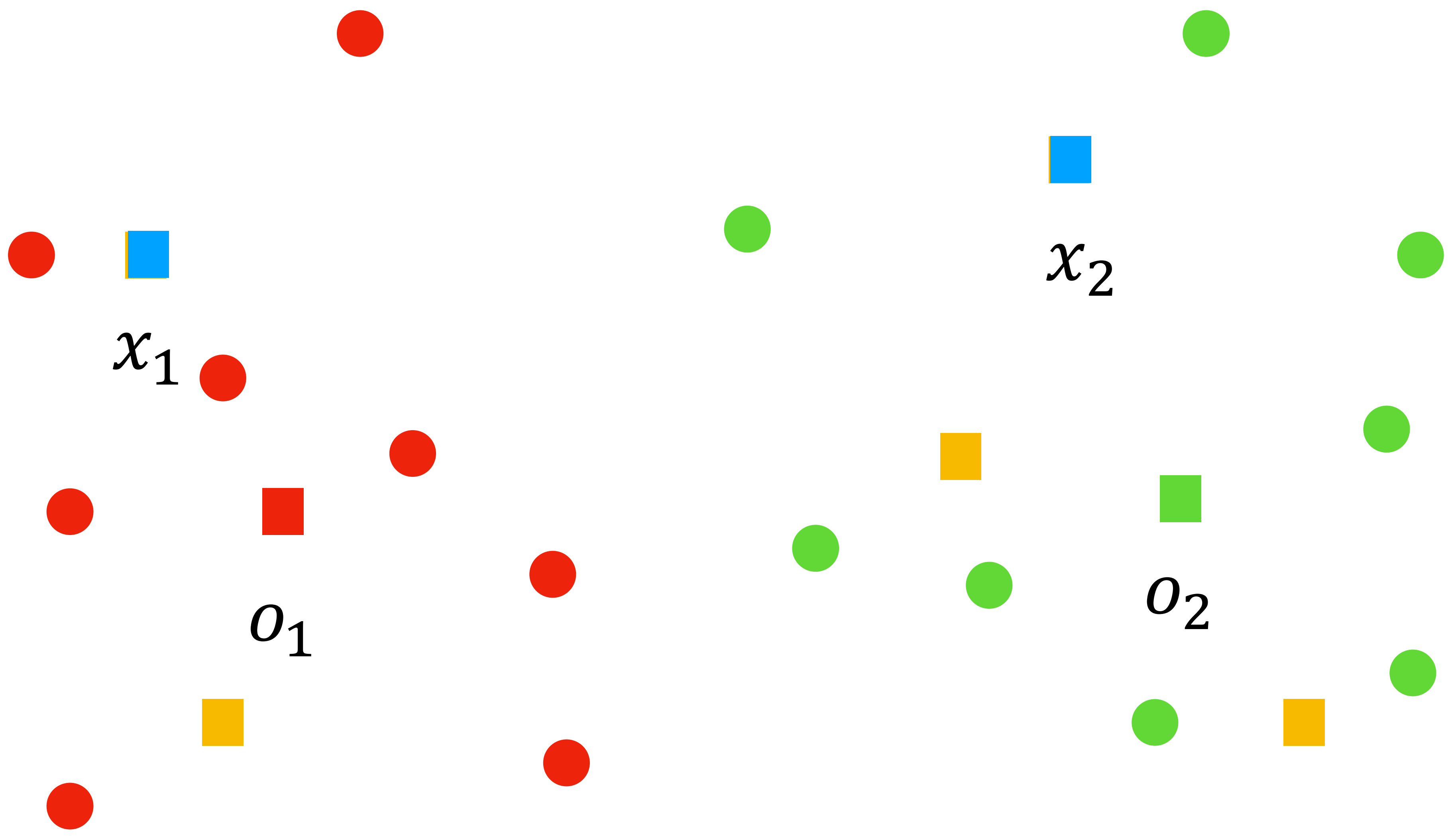


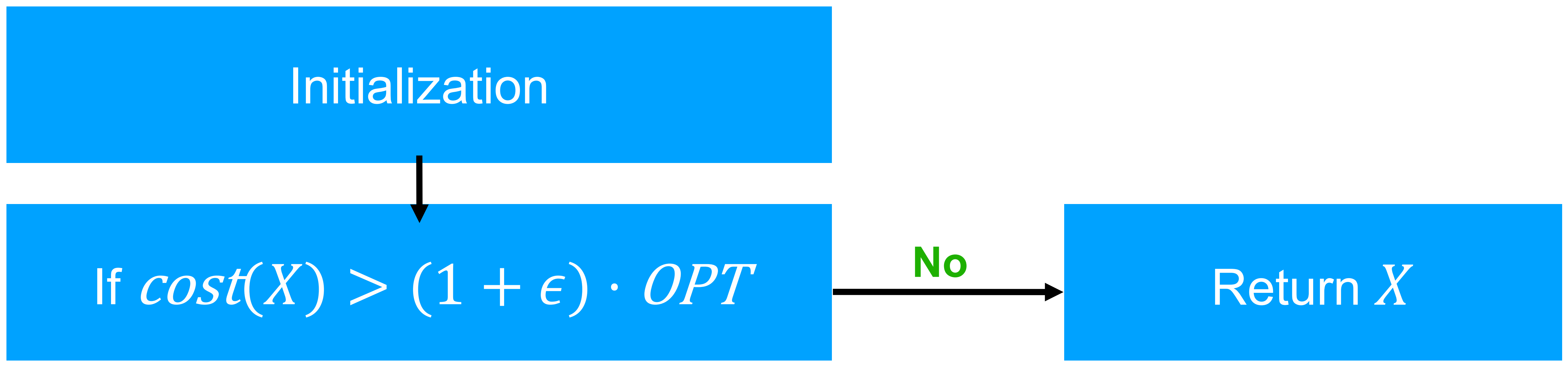
Unified-EPAS





Unified-EPAS





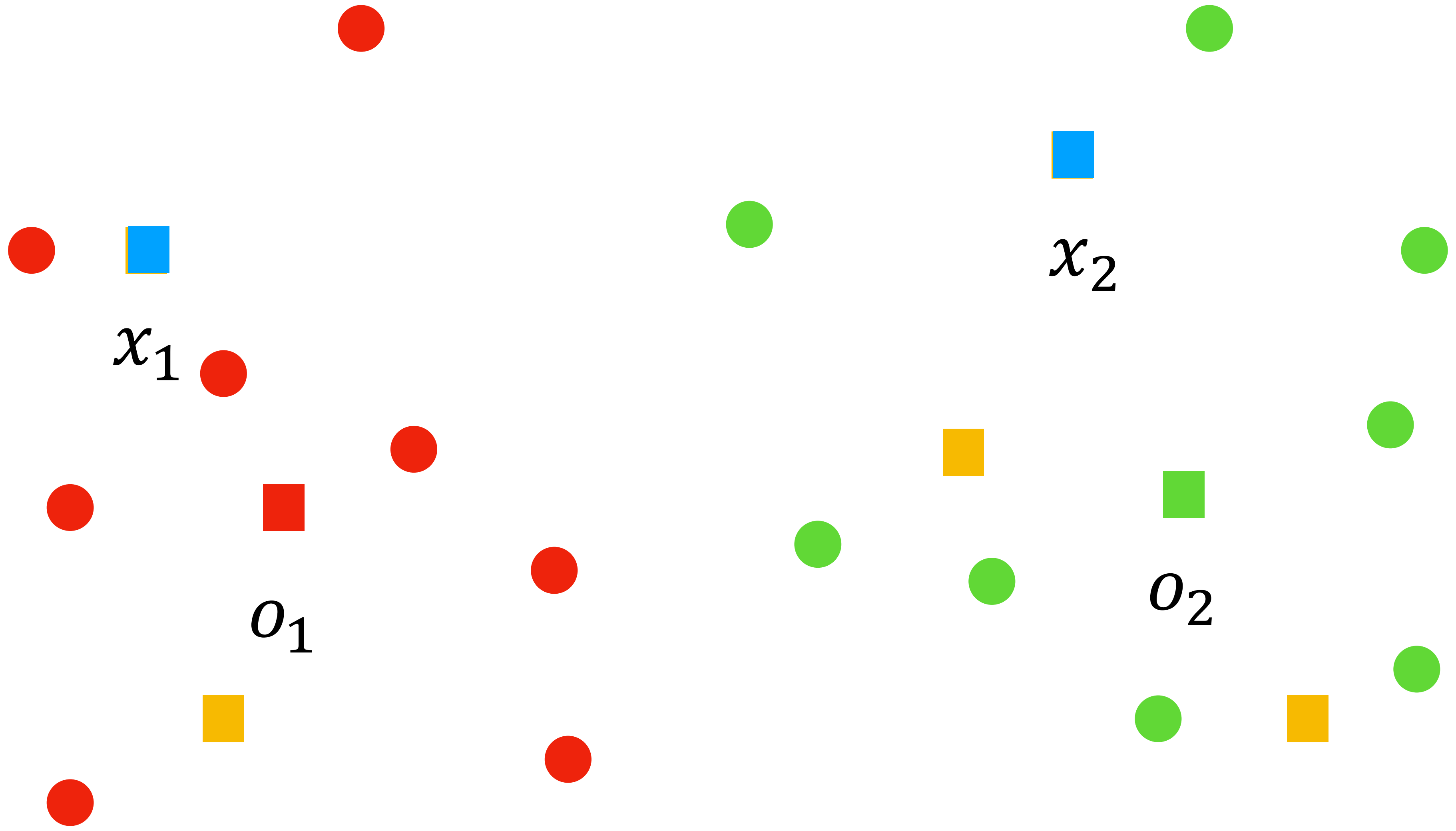
Initialization

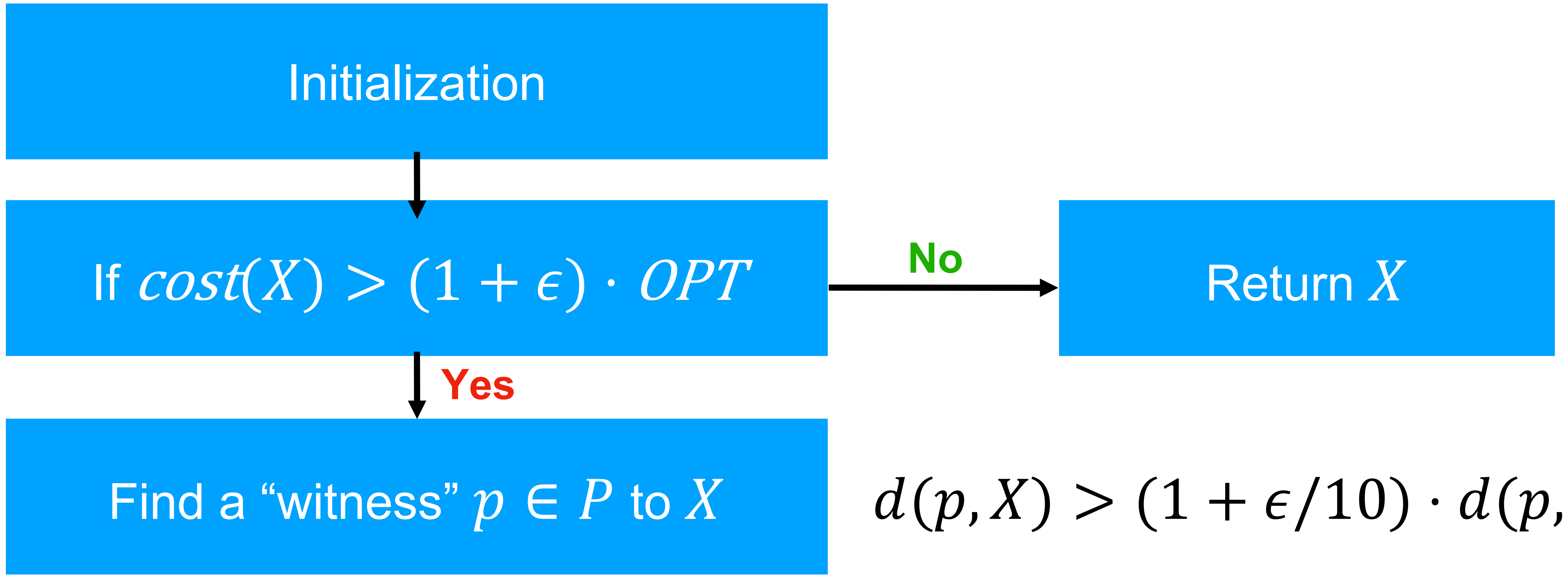
If $cost(X) > (1+\epsilon)\cdot OPT$

No

Return X

Unified-EPAS





Initialization

If $cost(X) > (1+\epsilon) \cdot OPT$

No

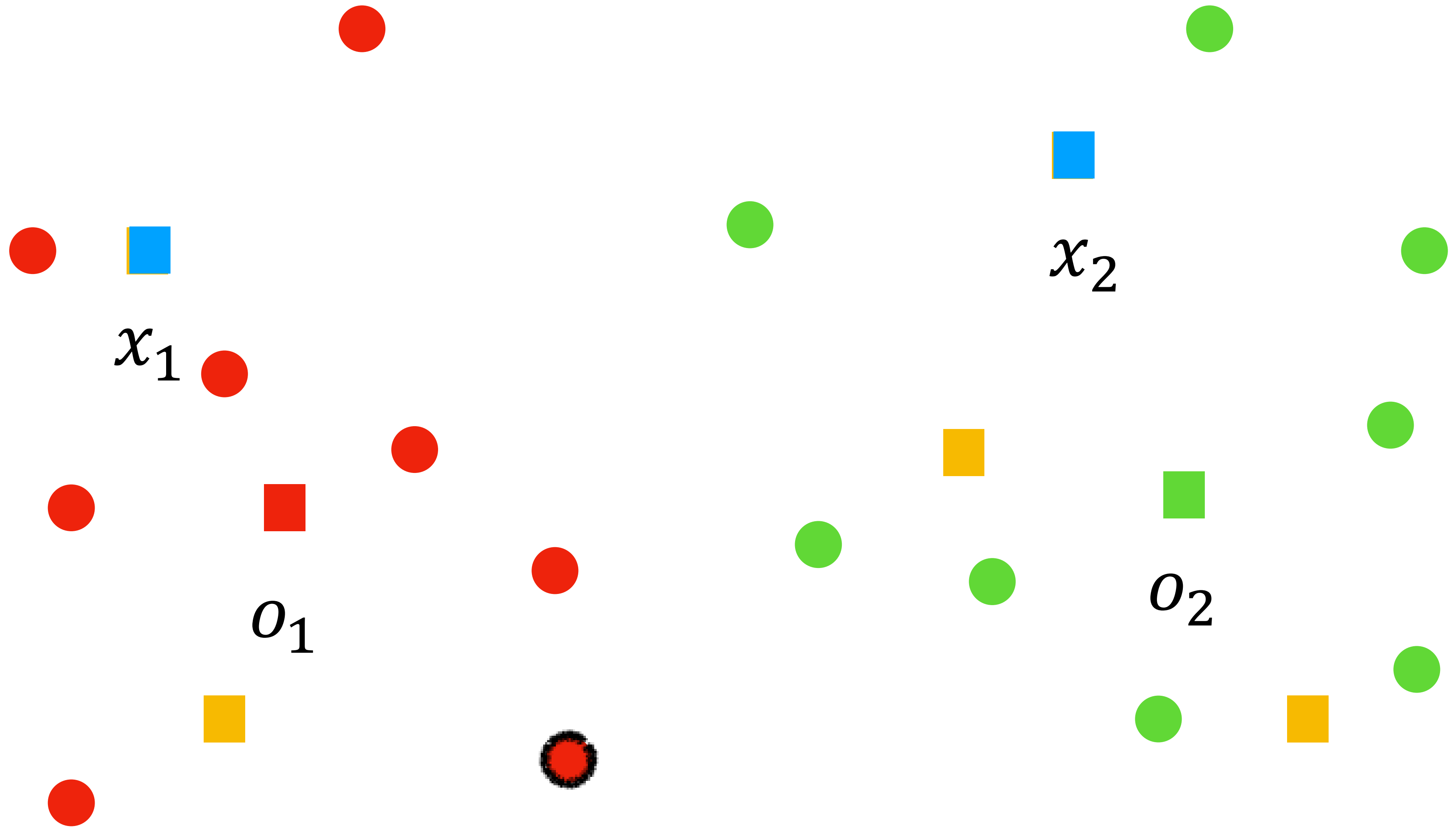
Return X

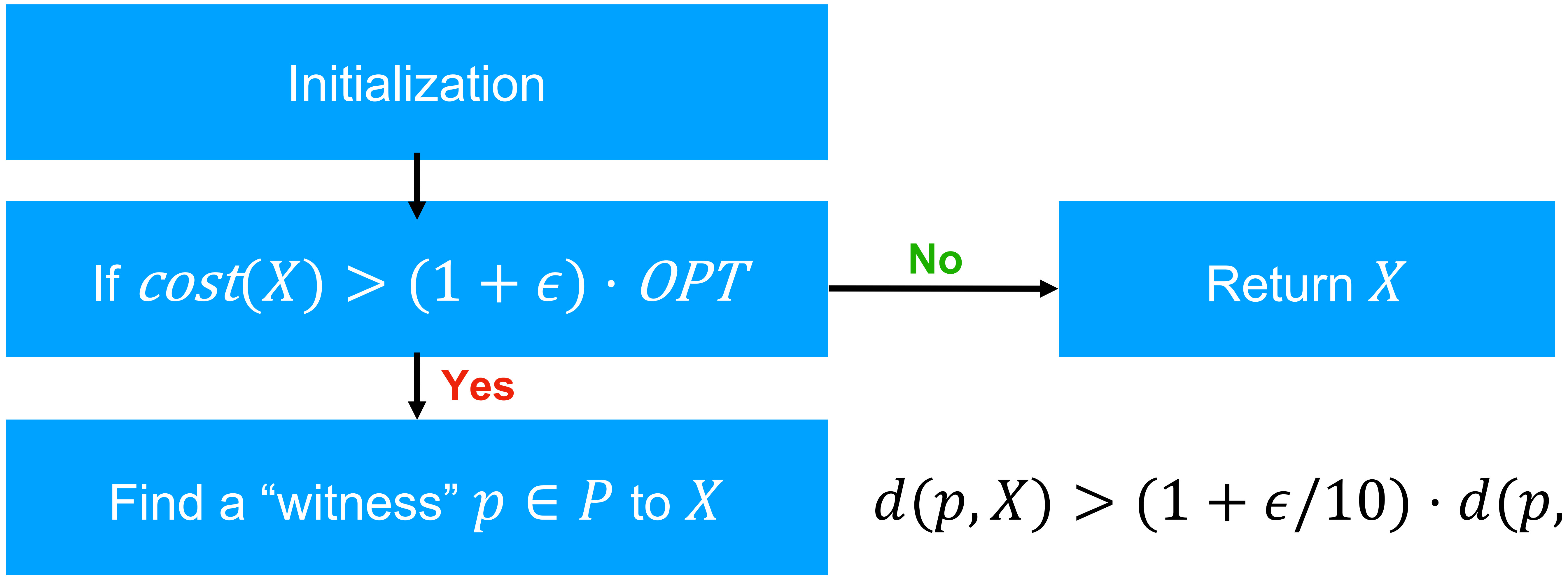
Yes

Find a “witness” $p \in P$ to X

$d(p, X) > (1+\epsilon/10) \cdot d(p, O)$

Unified-EPAS





Initialization

If $cost(X) > (1+\epsilon)\cdot OPT$

No

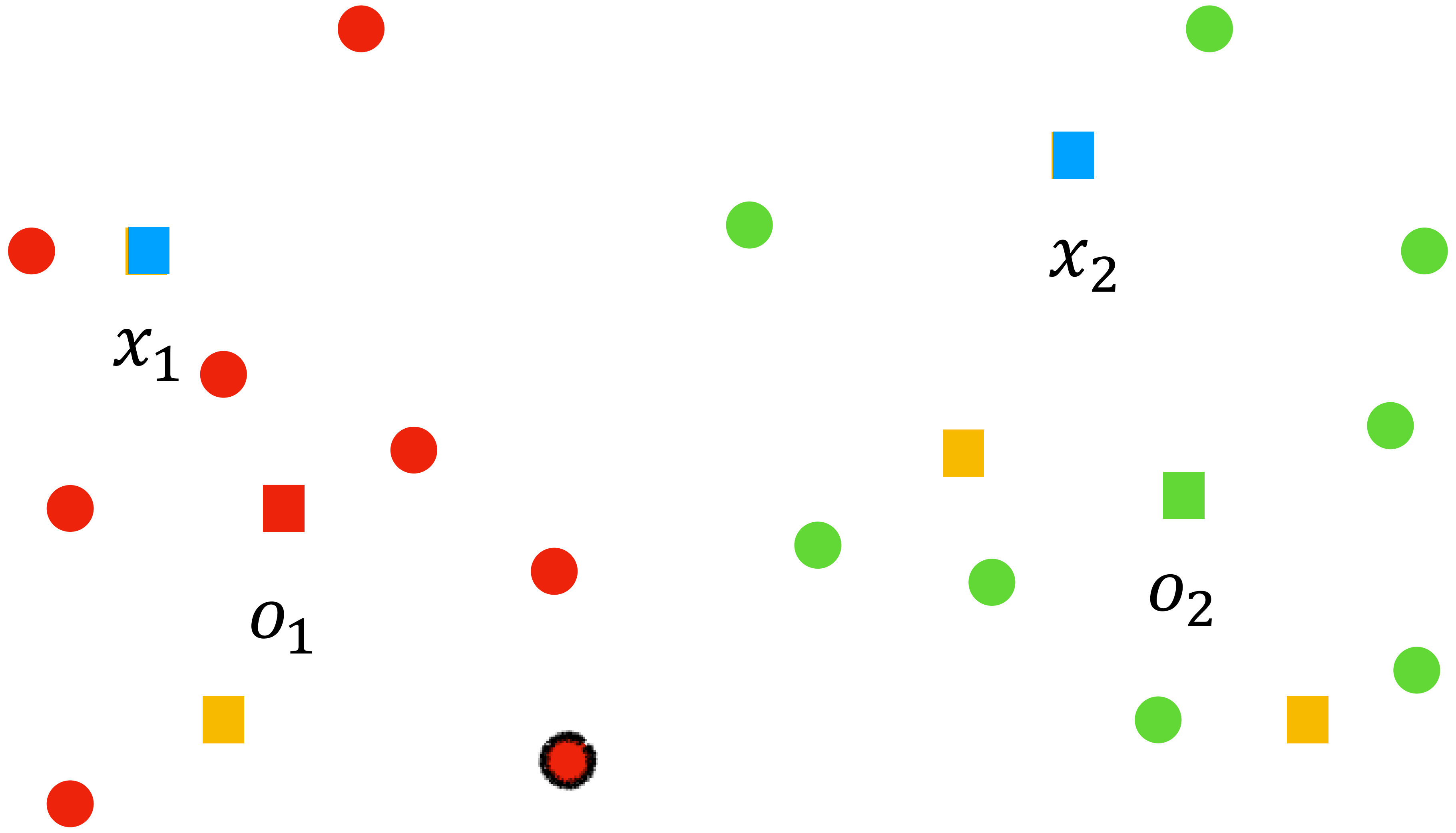
Return X

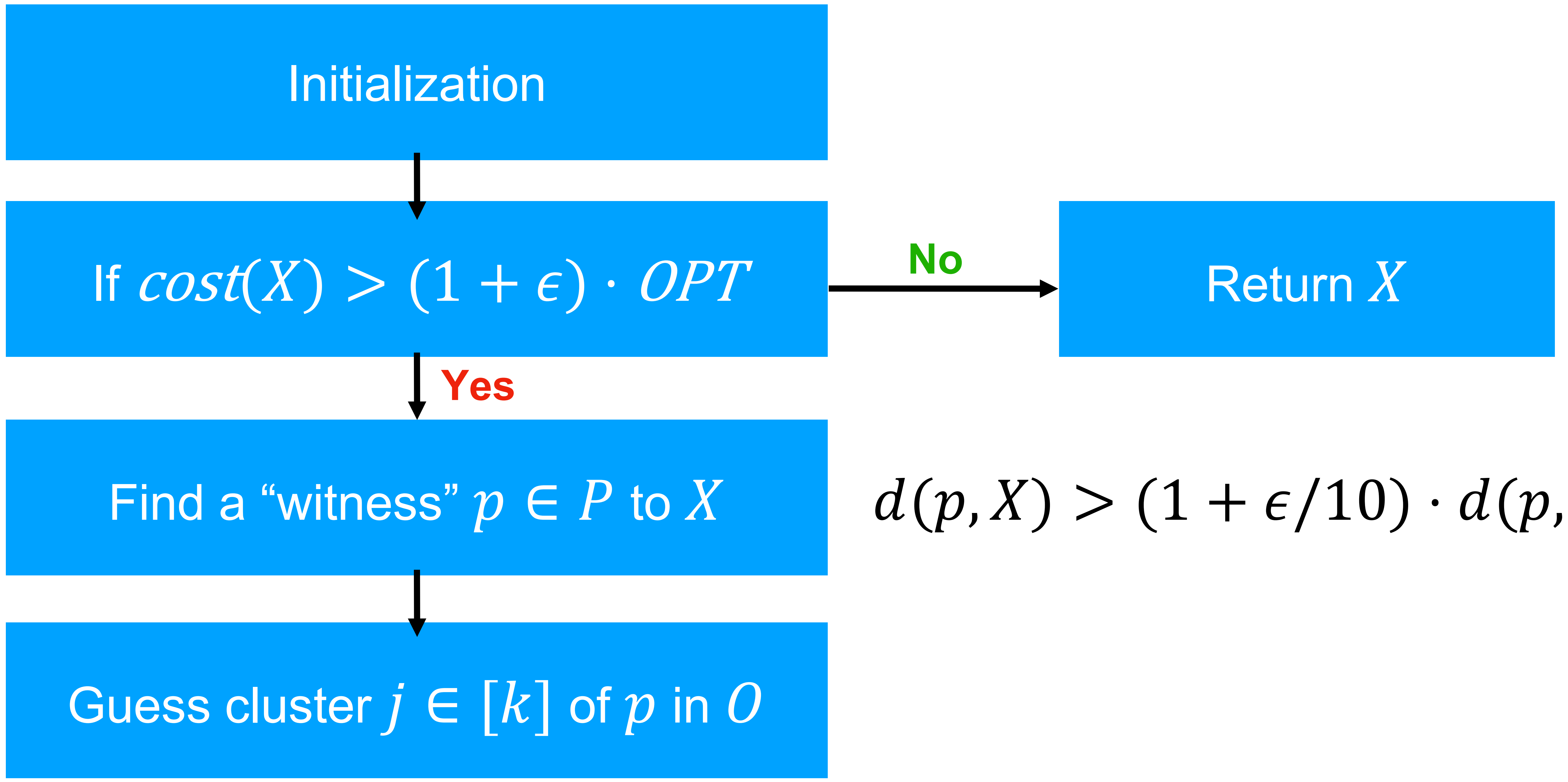
Yes

Find a “witness” $p \in P$ to X

$$d(p,X) > (1+\epsilon/10)\cdot d(p,O)$$

Unified-EPAS





Initialization

If $cost(X) > (1+\epsilon) \cdot OPT$

No

Return X

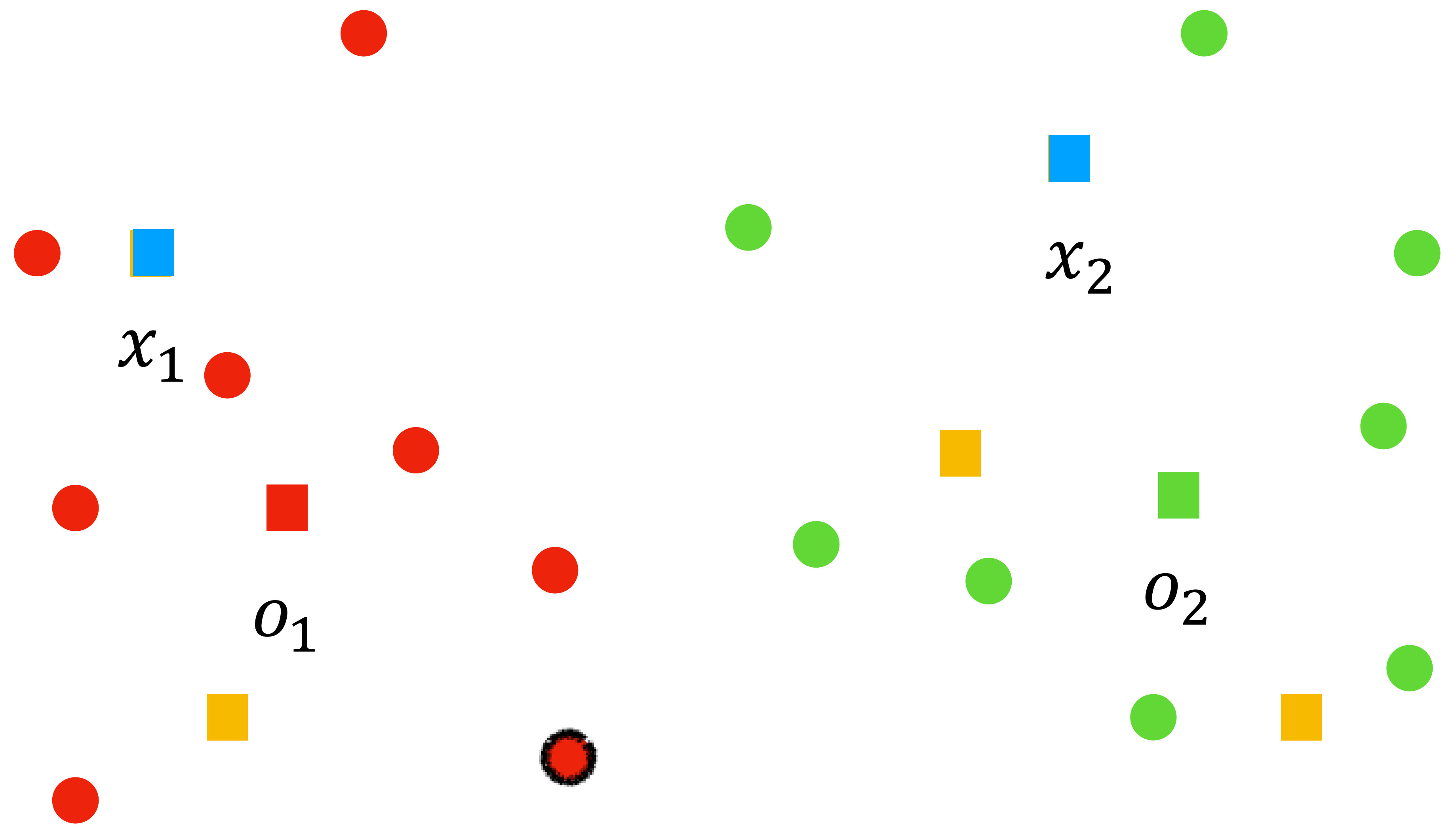
Yes

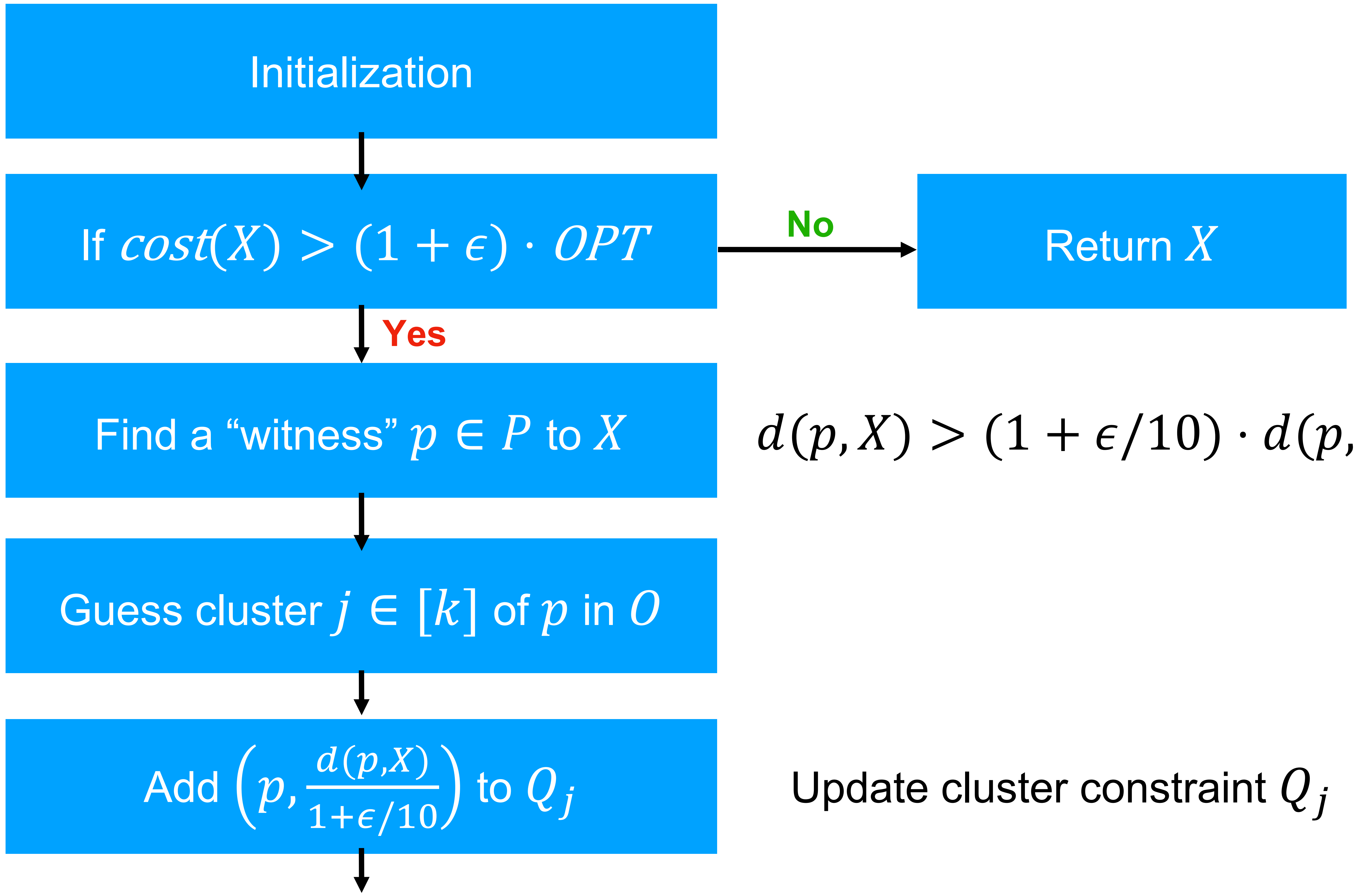
Find a “witness” $p \in P$ to X

$d(p, X) > (1 + \epsilon/10) \cdot d(p, O)$

Guess cluster $j \in [k]$ of p in O

Unified-EPAS





Initialization

If $cost(X) > (1+\epsilon)\cdot OPT$

No

Return X

Yes

Find a "witness" $p \in P$ to X

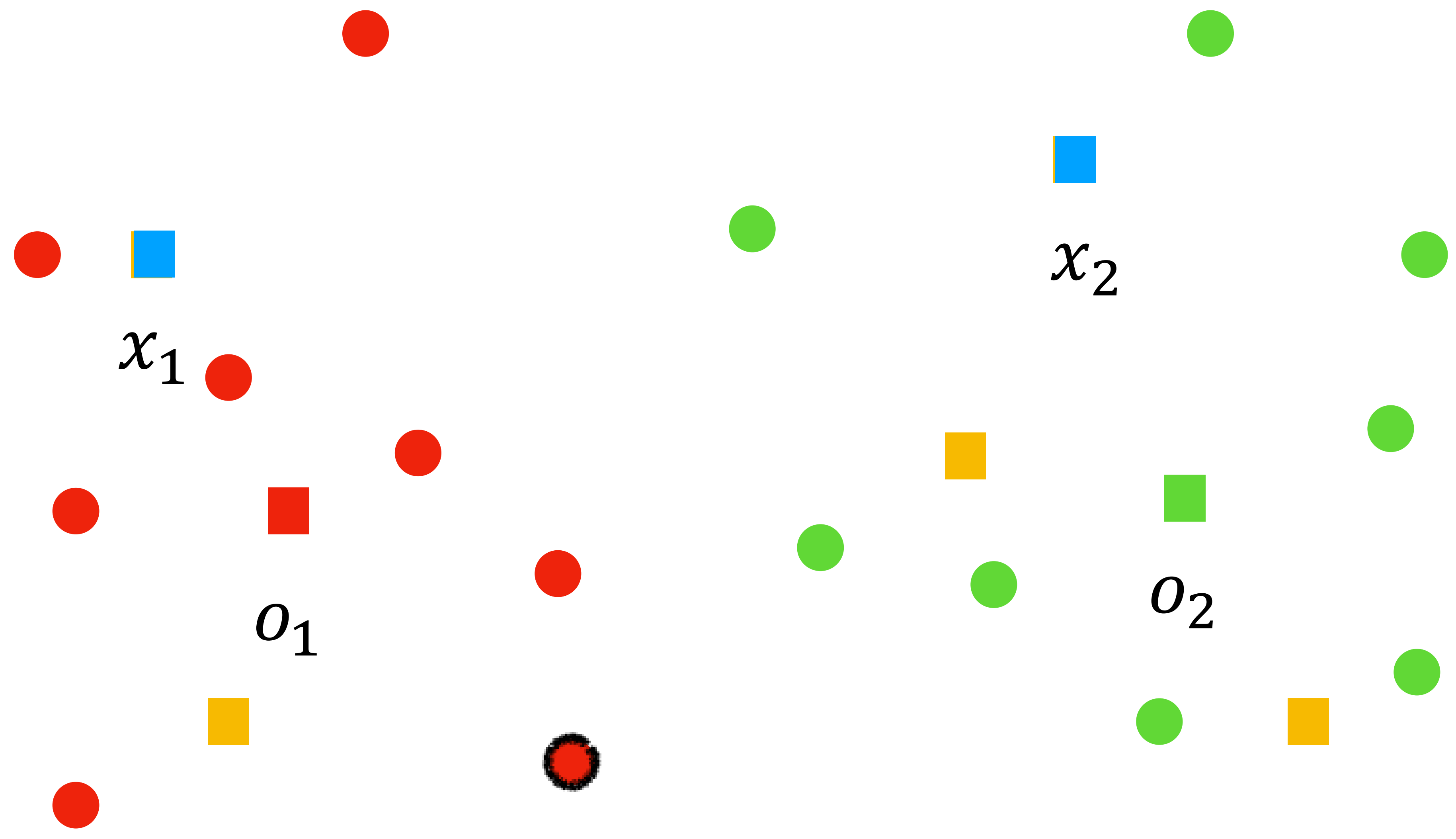
Guess cluster $j \in [k]$ of p in O

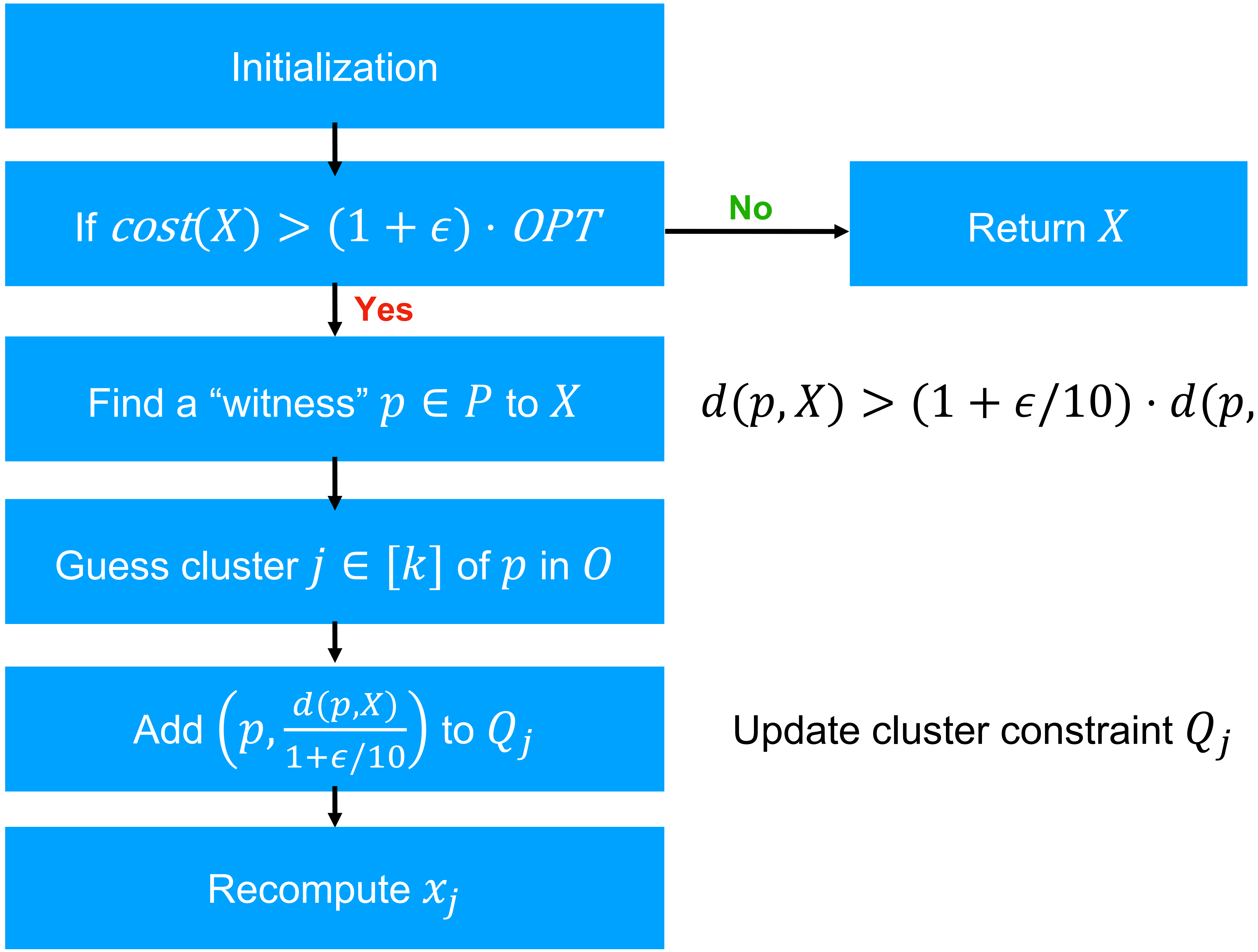
Add $\left(p, \frac{d(p,X)}{1+\epsilon/10}\right)$ to Q_j

$$d(p,X) > (1+\epsilon/10)\cdot d(p,O)$$

Update cluster constraint Q_j

Unified-EPAS





Initialization

If $cost(X) > (1+\epsilon) \cdot OPT$

No

Return X

Yes

Find a "witness" $p \in P$ to X

$d(p, X) > (1+\epsilon/10) \cdot d(p, O)$

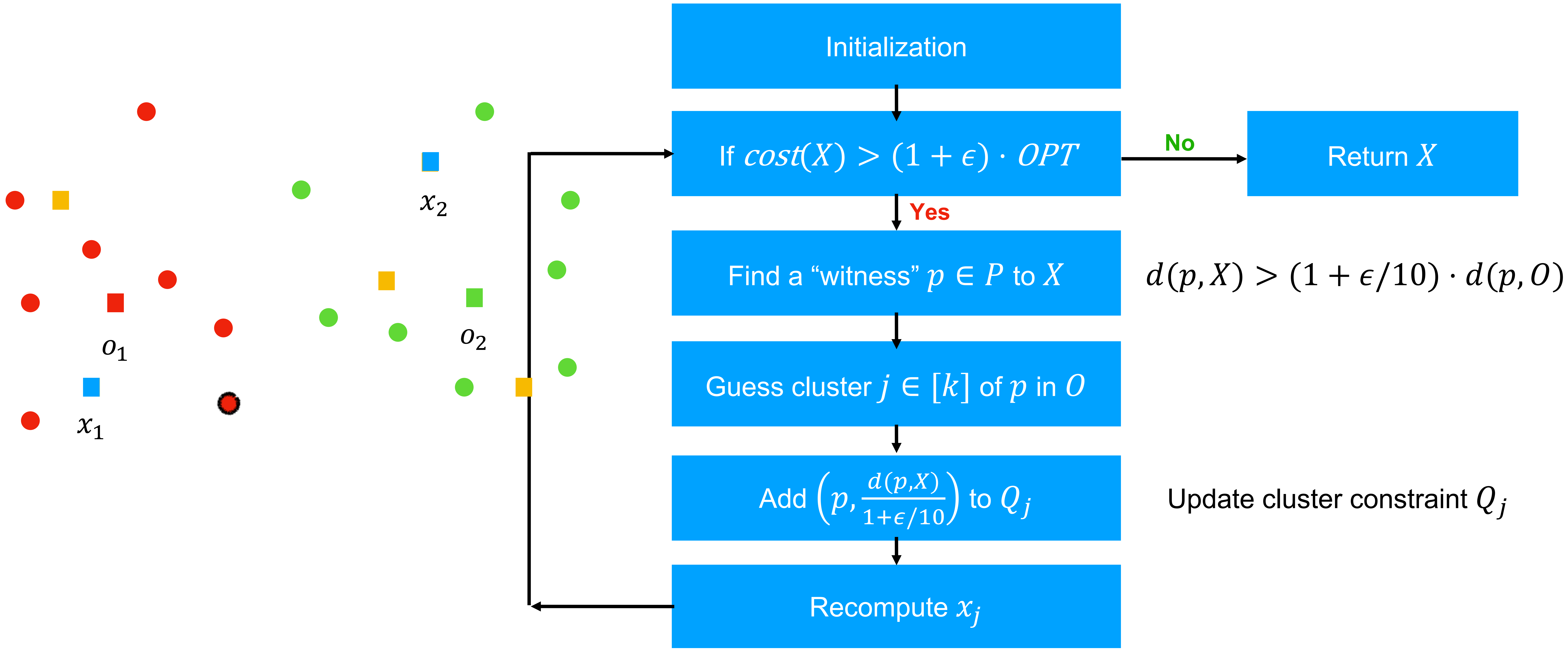
Guess cluster $j \in [k]$ of p in O

Add $\left(p, \frac{d(p,X)}{1+\epsilon/10}\right)$ to Q_j

Update cluster constraint Q_j

Recompute x_j

Unified-EPAS



Initialization

If $cost(X) > (1+\epsilon)\cdot OPT$

No

Return X

Yes

Find a “witness” $p \in P$ to X

$d(p,X) > (1+\epsilon/10)\cdot d(p,O)$

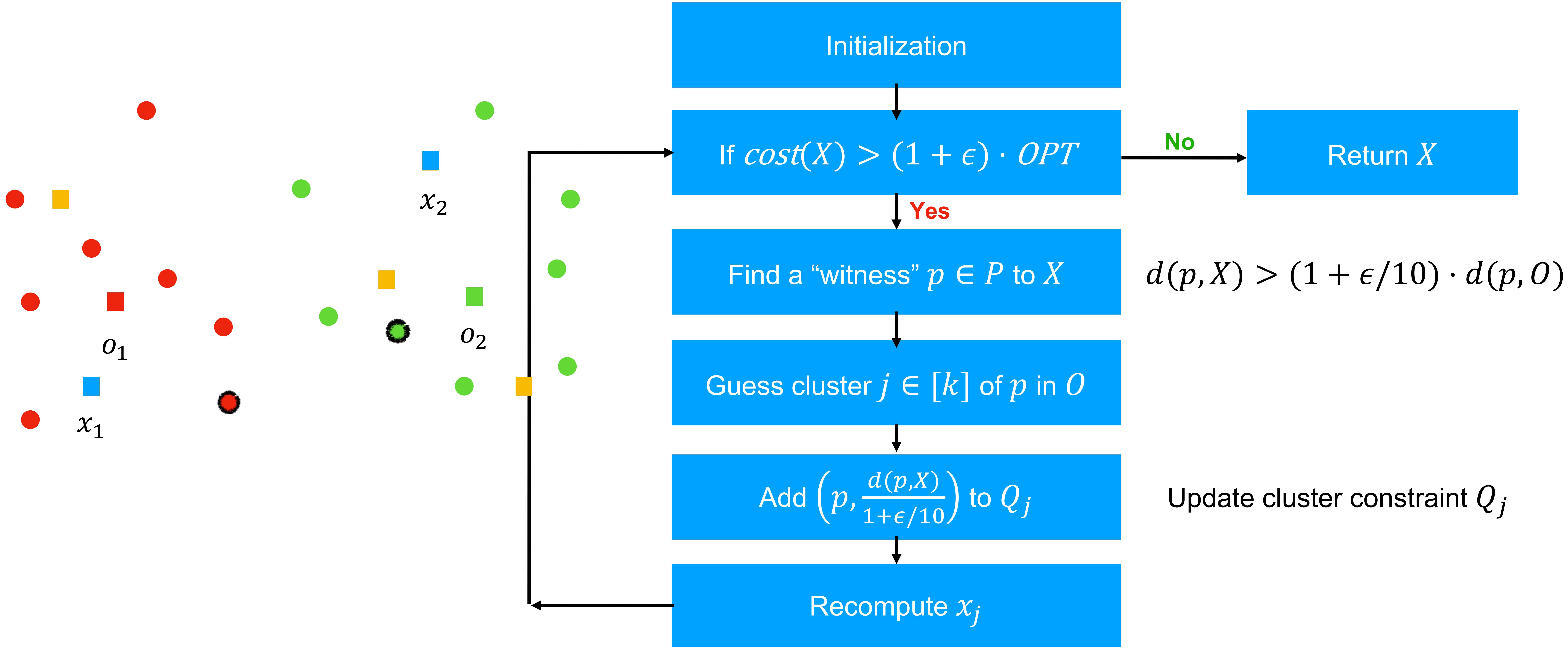
Guess cluster $j \in [k]$ of p in O

Add $\left(p, \frac{d(p,X)}{1+\epsilon/10}\right)$ to Q_j

Update cluster constraint Q_j

Recompute x_j

Unified-EPAS



Initialization

If $cost(X) > (1+\epsilon) \cdot OPT$

Return X

Find a "witness" $p \in P$ to X

$d(p,X) > (1+\epsilon/10) \cdot d(p,O)$

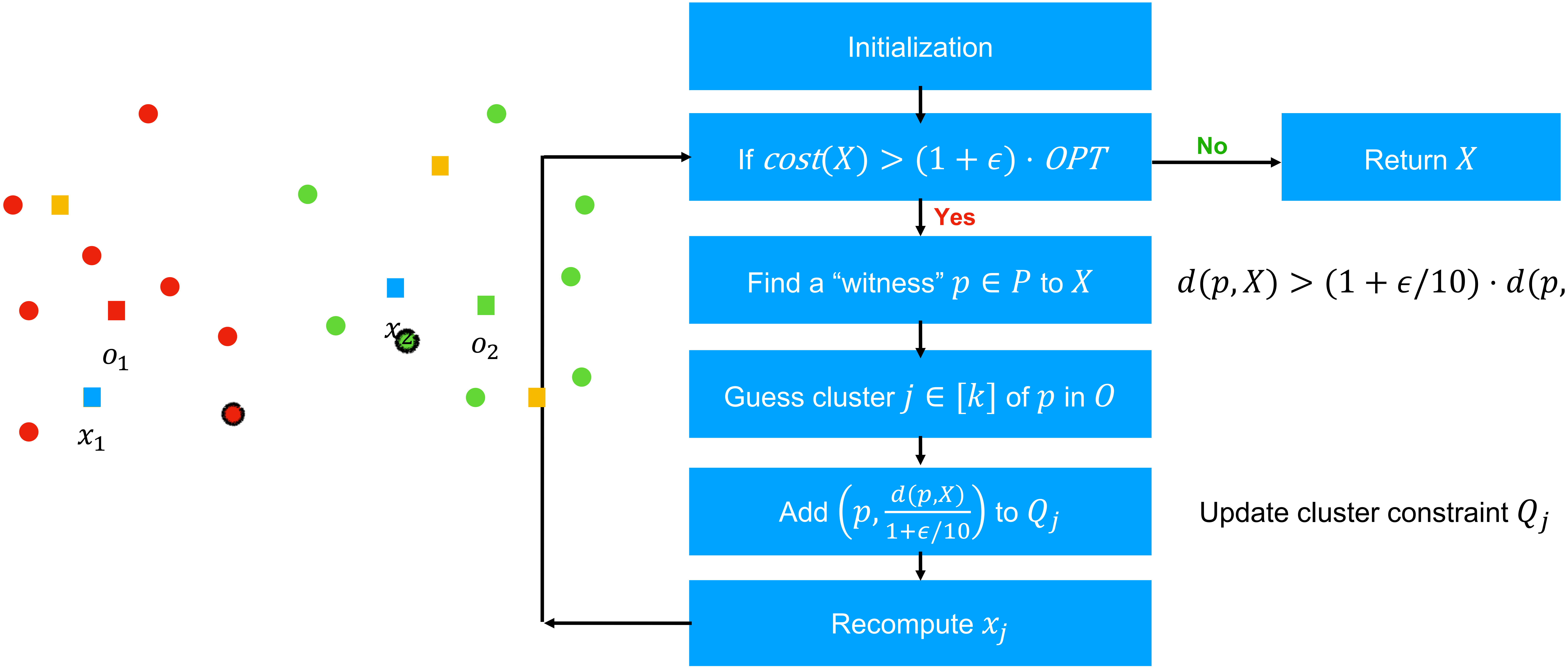
Guess cluster $j \in [k]$ of p in O

Add $\left(p, \frac{d(p,X)}{1+\epsilon/10}\right)$ to Q_j

Update cluster constraint Q_j

Recompute x_j

Unified-EPAS



Initialization

If $cost(X) > (1+\epsilon) \cdot OPT$

Return X

Find a “witness” $p \in P$ to X

Guess cluster $j \in [k]$ of p in O

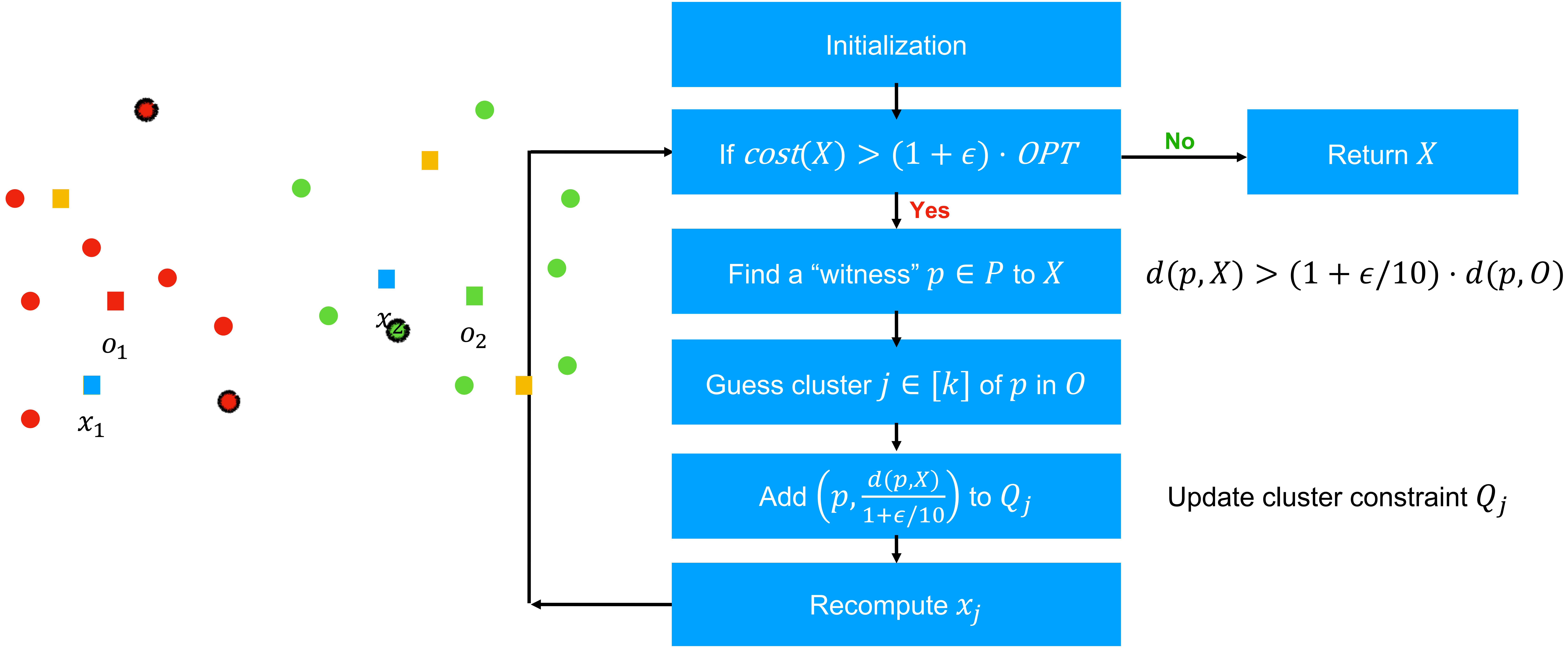
Add $\left(p, \frac{d(p,X)}{1+\epsilon/10}\right)$ to Q_j

Recompute x_j

$d(p,X) > (1+\epsilon/10) \cdot d(p,O)$

Update cluster constraint Q_j

Unified-EPAS



Initialization

If $cost(X) > (1+\epsilon) \cdot OPT$

Return X

Find a “witness” $p \in P$ to X

$d(p, X) > (1+\epsilon/10) \cdot d(p, O)$

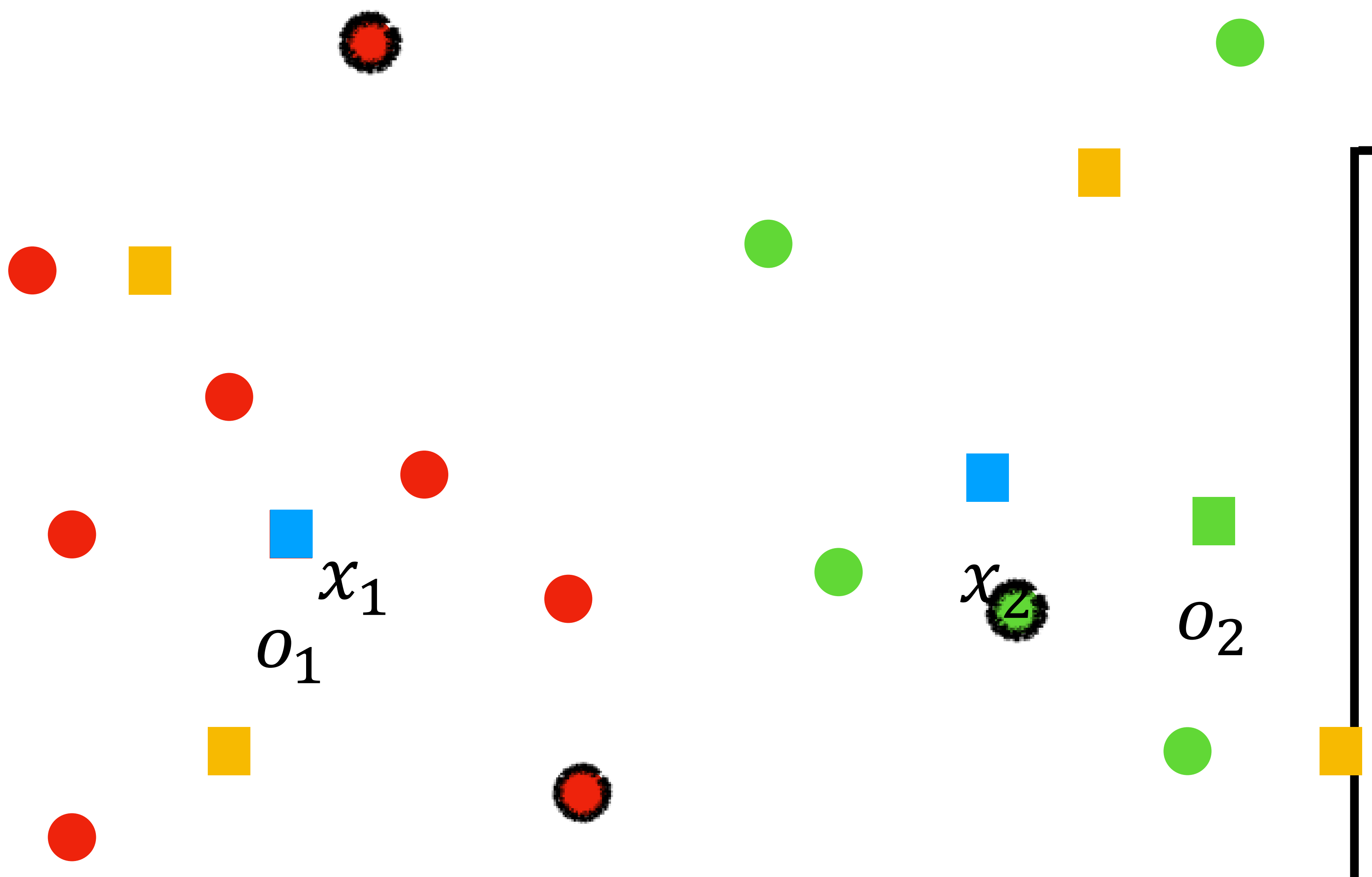
Guess cluster $j \in [k]$ of p in O

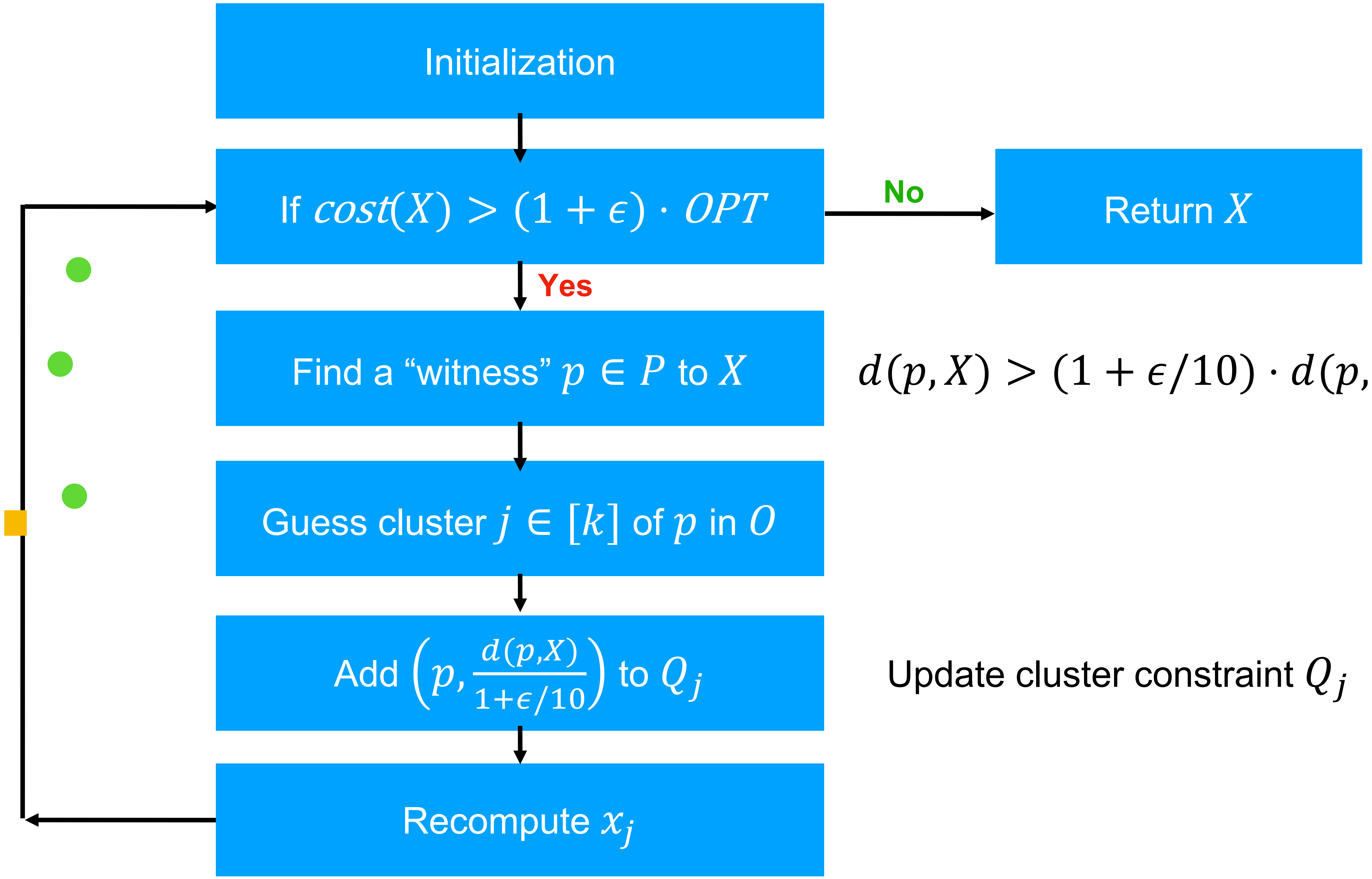
Add $\left(p, \frac{d(p,X)}{1+\epsilon/10}\right)$ to Q_j

Update cluster constraint Q_j

Recompute x_j

Unified-EPAS





Initialization

If $cost(X) > (1+\epsilon) \cdot OPT$

No

Return X

Yes

Find a "witness" $p \in P$ to X

$d(p,X) > (1+\epsilon/10) \cdot d(p,O)$

Guess cluster $j \in [k]$ of p in O

Add $\left(p, \frac{d(p,X)}{1+\epsilon/10}\right)$ to Q_j

Update cluster constraint Q_j

Recompute x_j

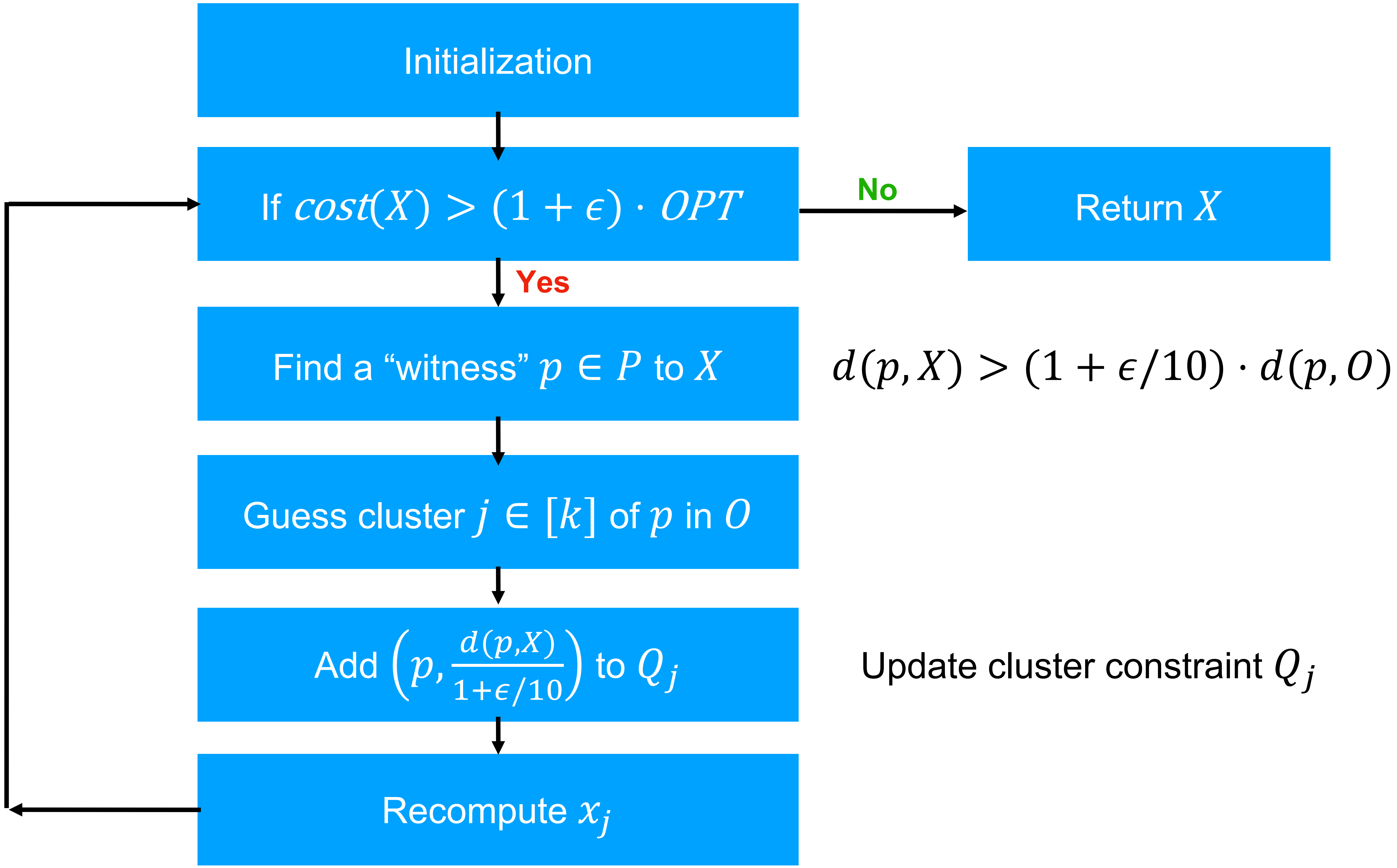
Unified-EPAS

Lemma 1

If $cost(X) > (1+\epsilon) \cdot OPT$, then we can find a witness to X w.h.p.

Question:

Bound #iterations?



Initialization
If $cost(X) > (1+\epsilon) \cdot OPT$
No
Return X
Yes
Find a “witness” $p \in P$ to X
Guess cluster $j \in [k]$ of p in O
Add $\left(p, \frac{d(p,X)}{1+\epsilon/10}\right)$ to Q_j
Recompute x_j

$d(p,X) > (1+\epsilon/10) \cdot d(p,O)$

Update cluster constraint Q_j

Unified-EPAS

Lemma 1

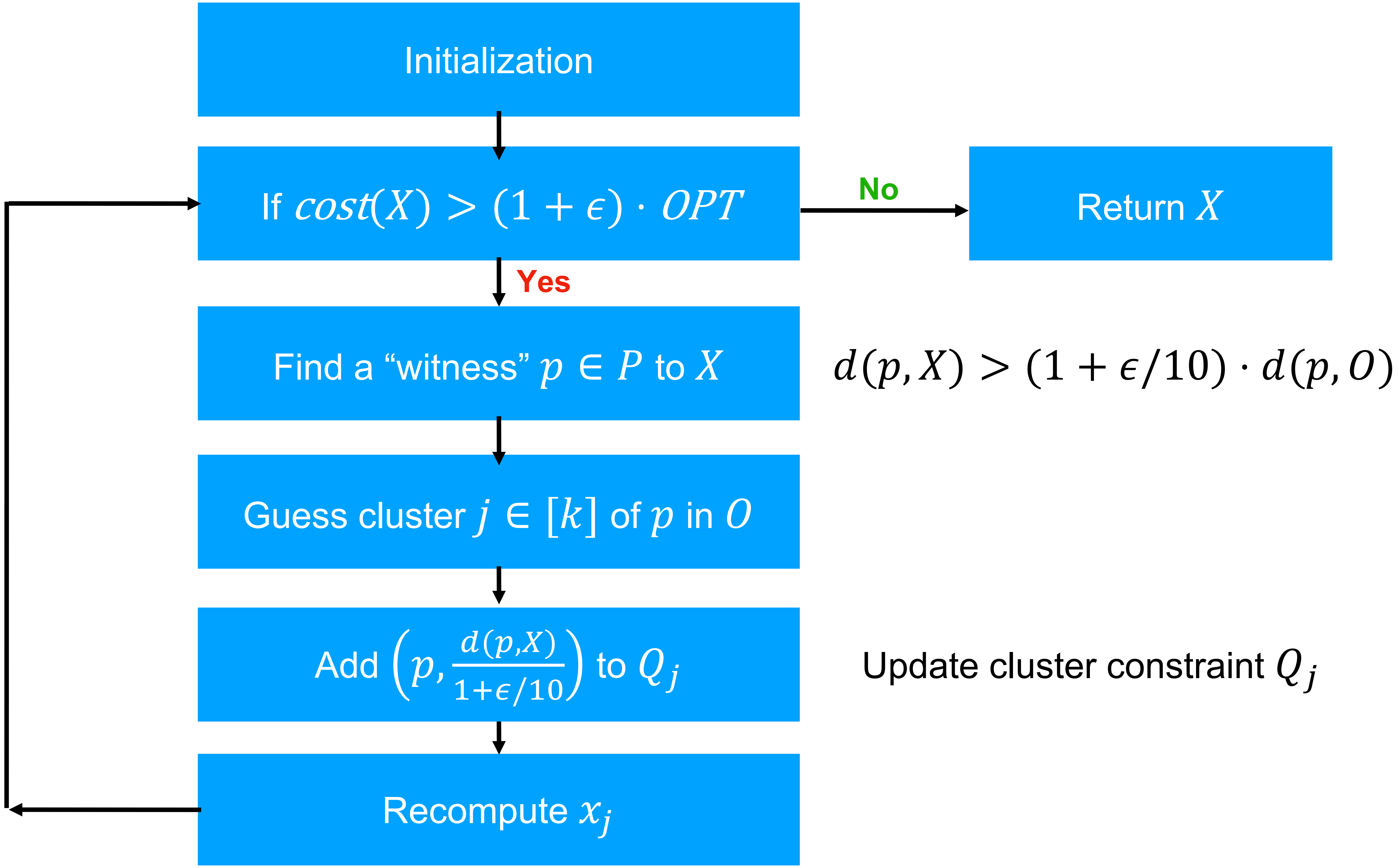
If $cost(X) > (1+\epsilon)\cdot OPT$, then we can find a witness to X w.h.p.

Question:

Bound #iterations?

ϵ-scatter dimension

Upper Bounds



Initialization

If $cost(X) > (1+\epsilon)\cdot OPT$

No

Return X

Yes

Find a “witness” $p \in P$ to X

$d(p,X) > (1+\epsilon/10)\cdot d(p,O)$

Guess cluster $j \in [k]$ of p in O

Add $\left(p, \frac{d(p,X)}{1+\epsilon/10}\right)$ to Q_j

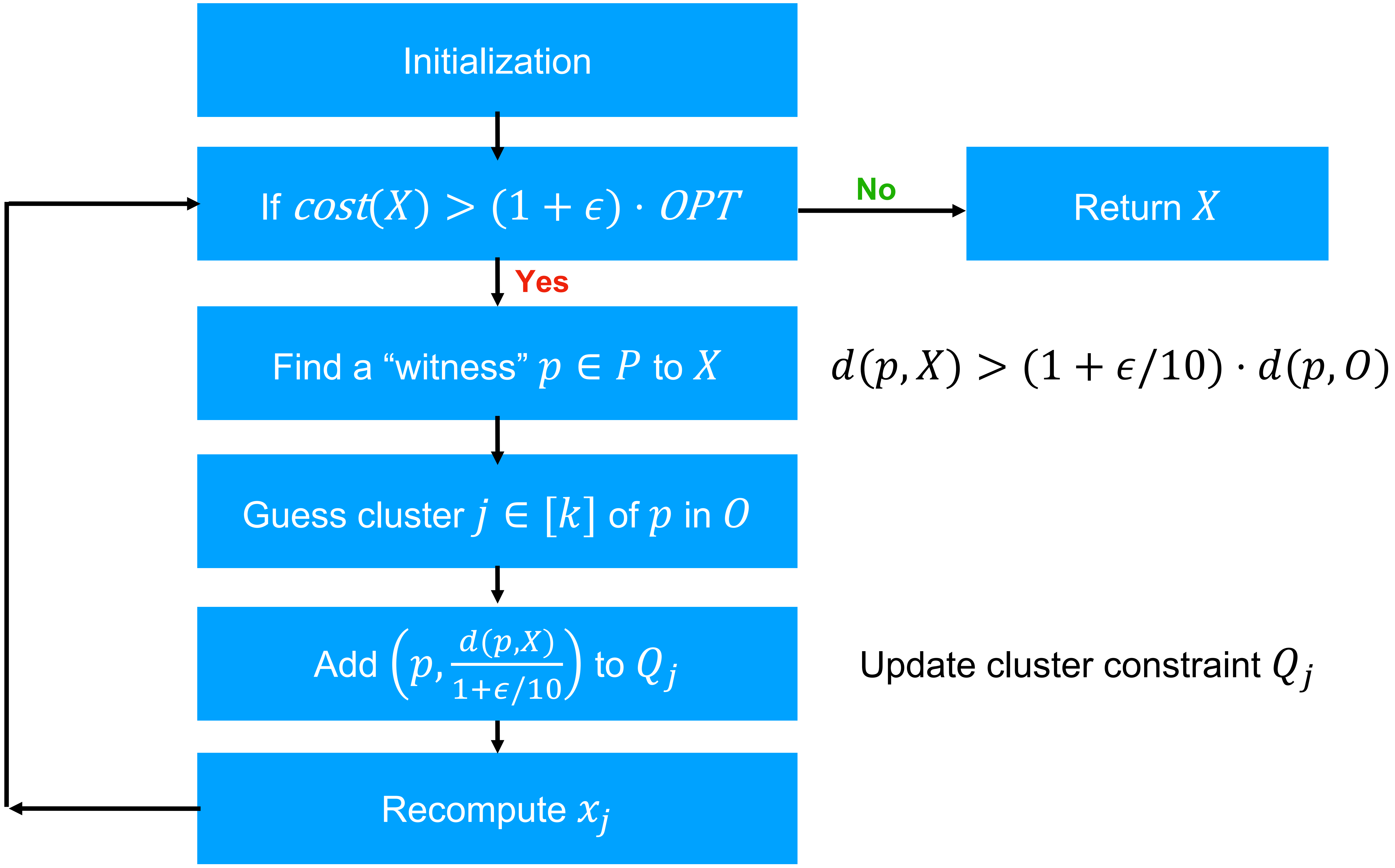
Update cluster constraint Q_j

Recompute x_j

Unified-EPAS

Bound #iterations?

ϵ-scatter dimension



Initialization

If $cost(X) > (1+\epsilon) \cdot OPT$

No

Return X

Yes

Find a "witness" $p \in P$ to X

$d(p,X) > (1+\epsilon/10) \cdot d(p,O)$

Guess cluster $j \in [k]$ of p in O

Add $\left(p, \frac{d(p,X)}{1+\epsilon/10}\right)$ to Q_j

Update cluster constraint Q_j

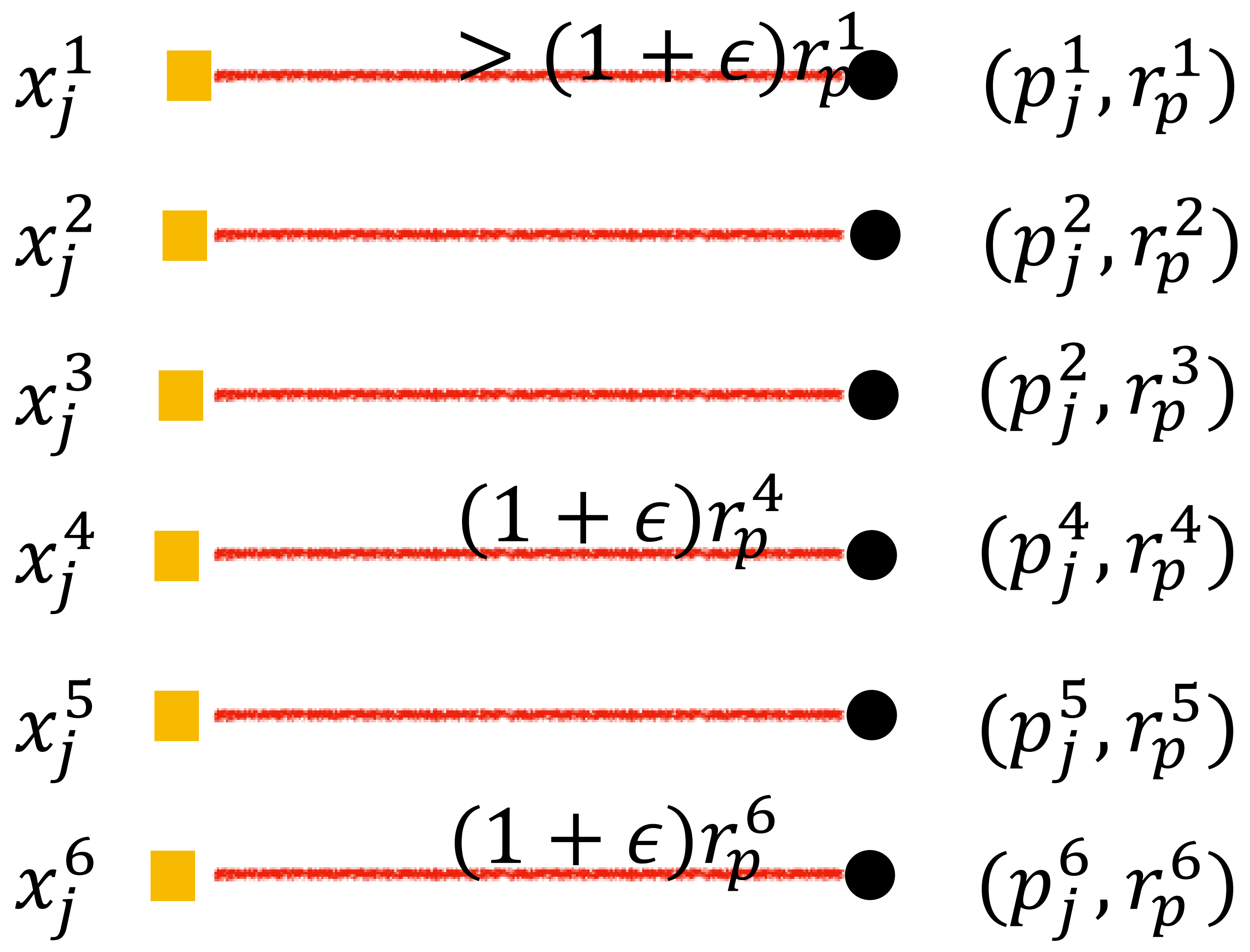
Recompute x_j

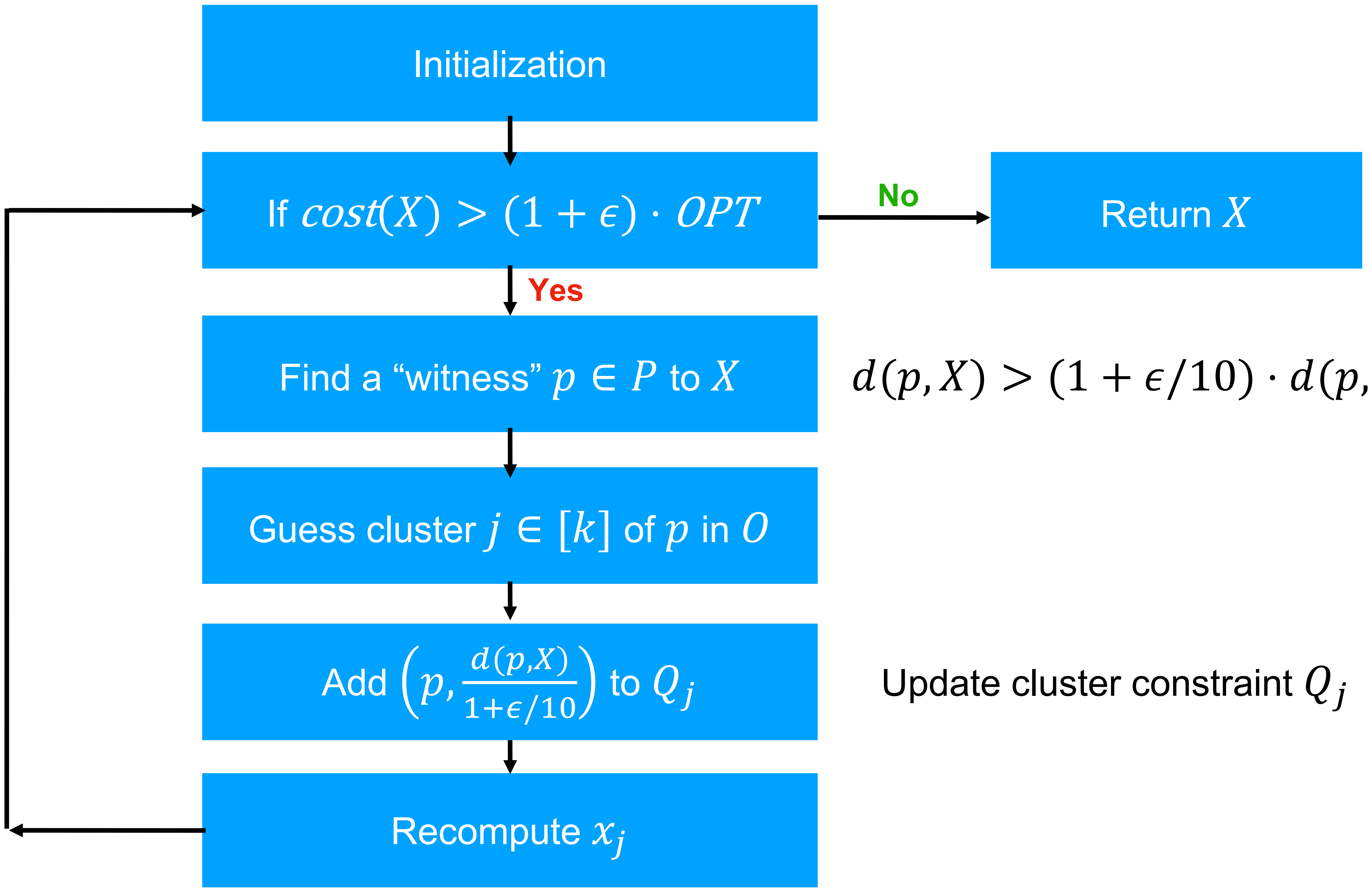
Unified-EPAS

Bound #iterations?

ϵ-scatter dimension

Fix Q_j




Initialization
If $cost(X) > (1+\epsilon)\cdot OPT$
No
Return X
Yes
Find a "witness" $p \in P$ to X
Guess cluster $j \in [k]$ of p in O
Add $\left(p, \frac{d(p,X)}{1+\epsilon/10}\right)$ to Q_j
Recompute x_j

$d(p,X) > (1+\epsilon/10)\cdot d(p,O)$

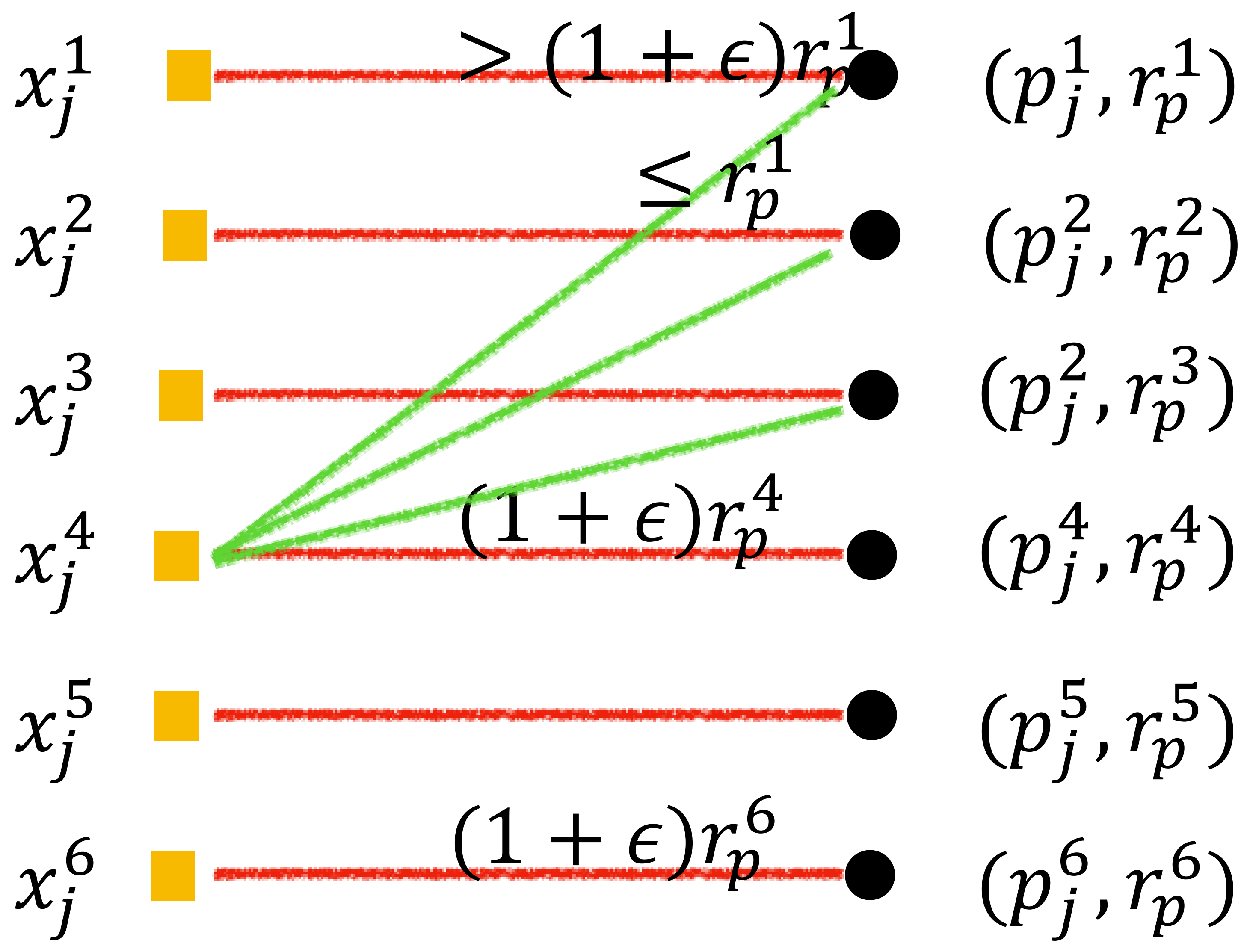
Update cluster constraint Q_j

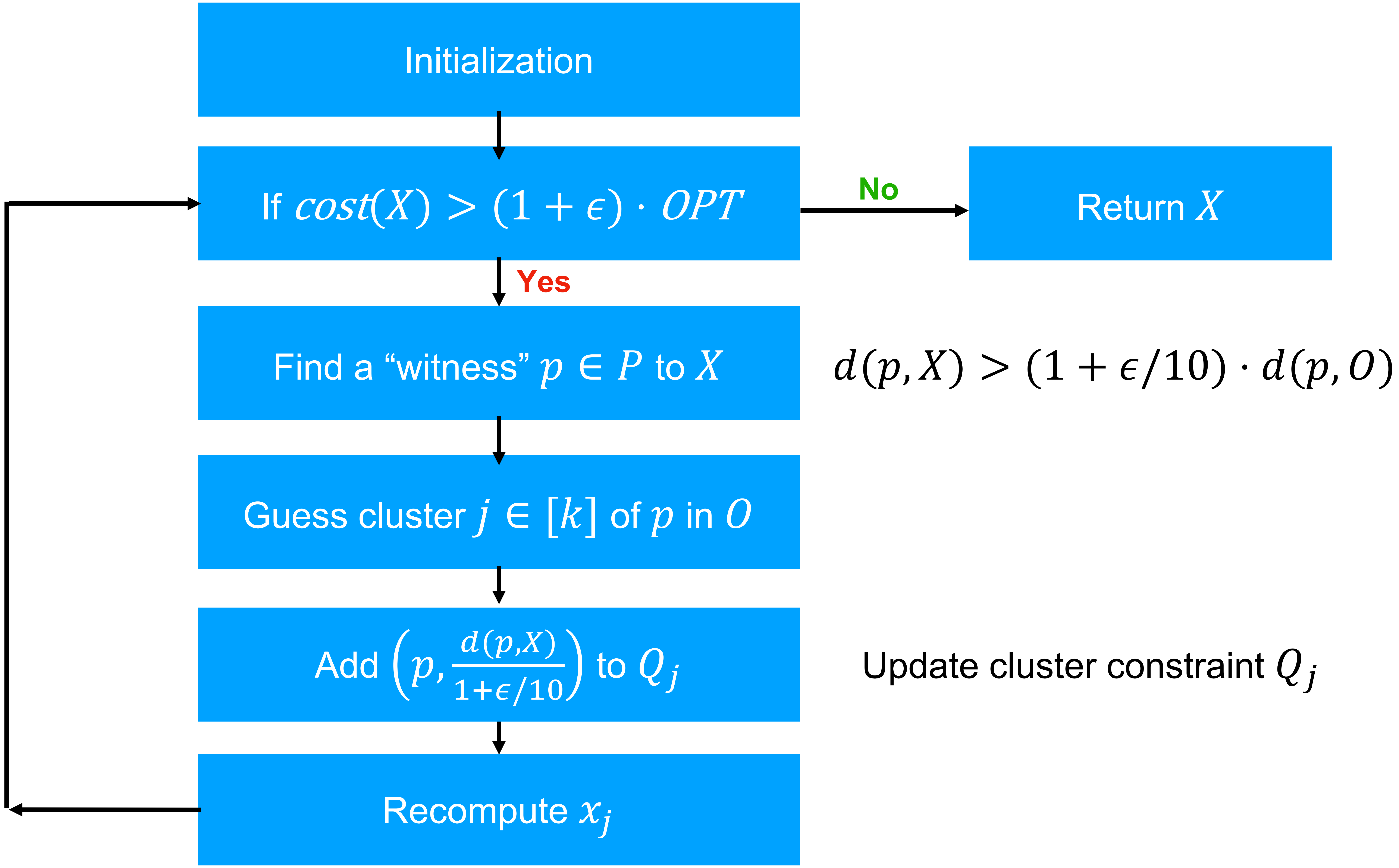
Unified-EPAS

Bound #iterations?

ϵ-scatter dimension

Fix Q_j





Initialization

If $cost(X) > (1+\epsilon) \cdot OPT$

No

Return X

Yes

Find a “witness” $p \in P$ to X

$d(p, X) > (1 + \epsilon/10) \cdot d(p, O)$

Guess cluster $j \in [k]$ of p in O

Add $\left(p, \frac{d(p,X)}{1+\epsilon/10}\right)$ to Q_j

Update cluster constraint Q_j

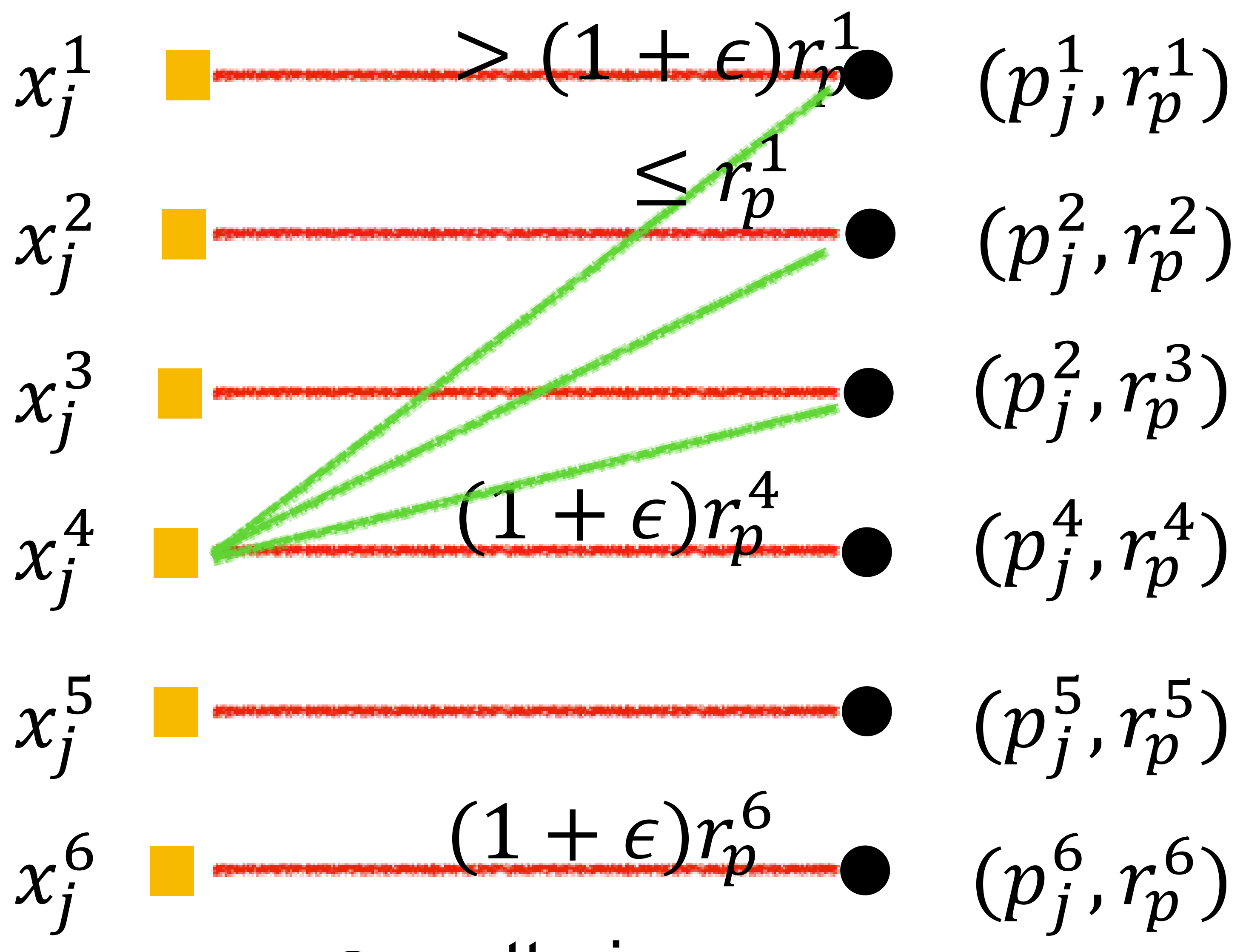
Recompute x_j

Unified-EPAS

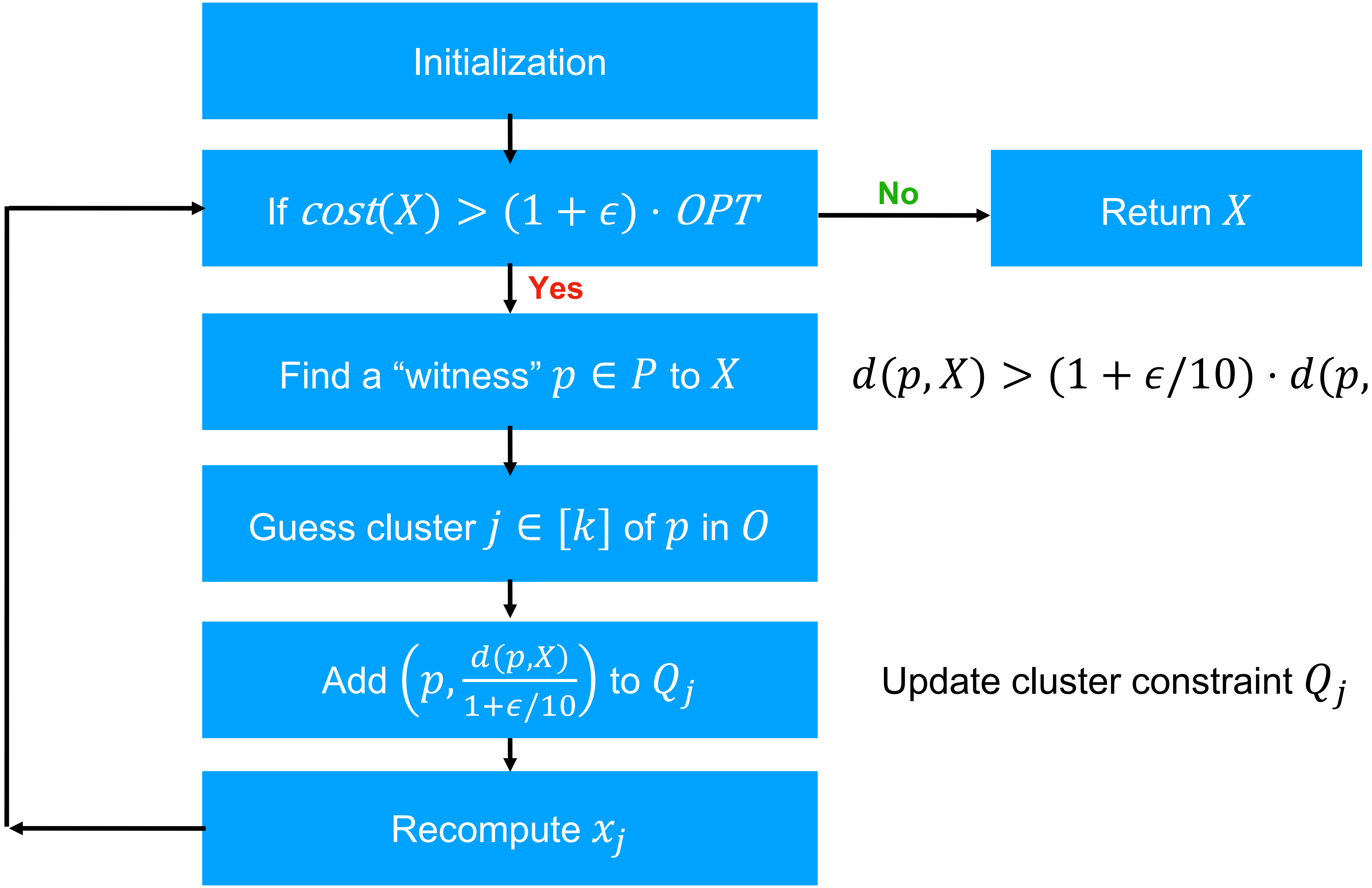
Bound #iterations?

ϵ-scatter dimension

Fix Q_j



ϵ-scattering



Initialization

If $cost(X) > (1+\epsilon) \cdot OPT$

No

Return X

Yes

Find a "witness" $p \in P$ to X

Guess cluster $j \in [k]$ of p in O

Add $\left(p, \frac{d(p,X)}{1+\epsilon/10}\right)$ to Q_j

Recompute x_j

$d(p,X) > (1+\epsilon/10) \cdot d(p,O)$

Update cluster constraint Q_j

ϵ-scatter dimension of a metric space is λ

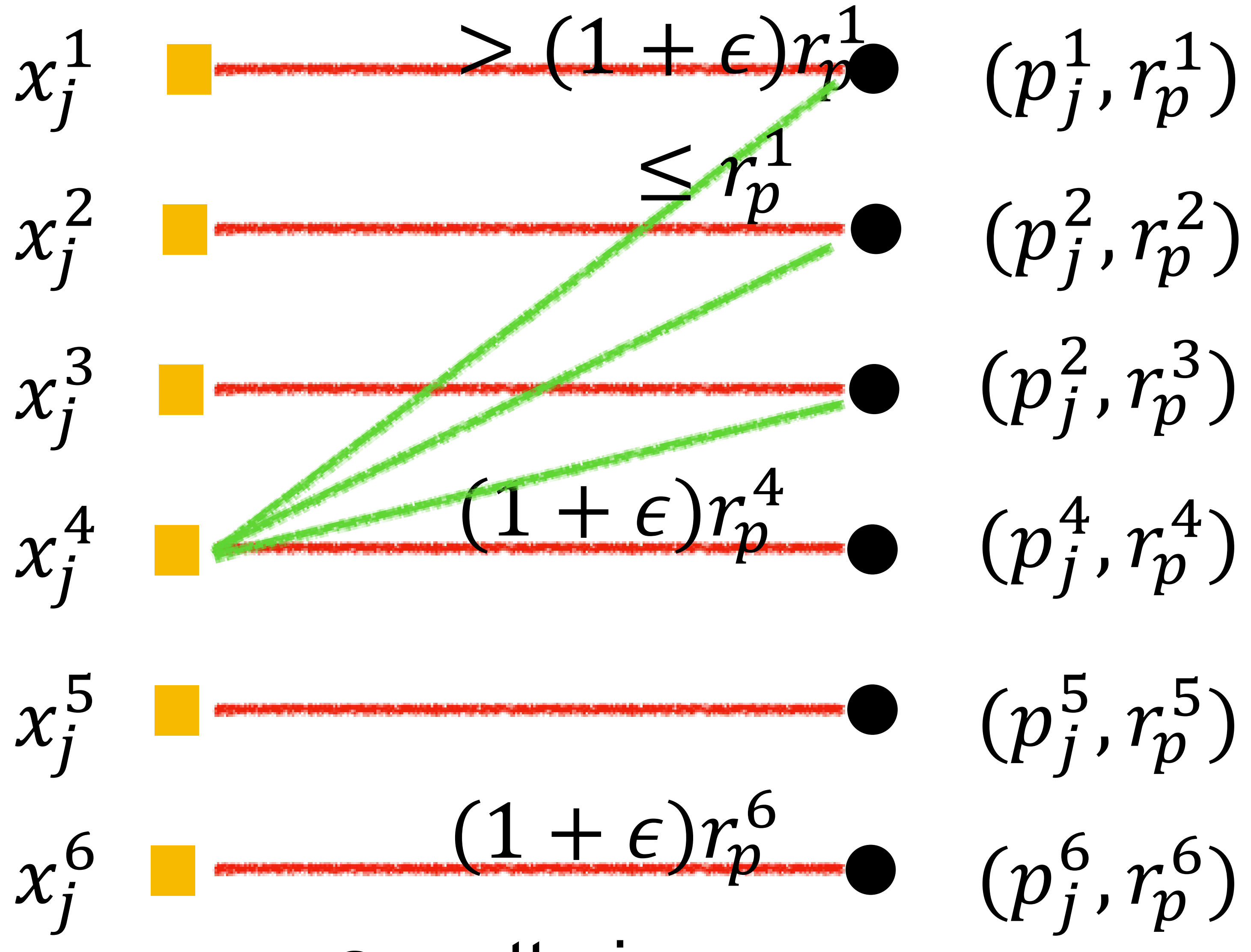
if any ϵ-scattering contains at most λ many triples with same radius

Unified-EPAS

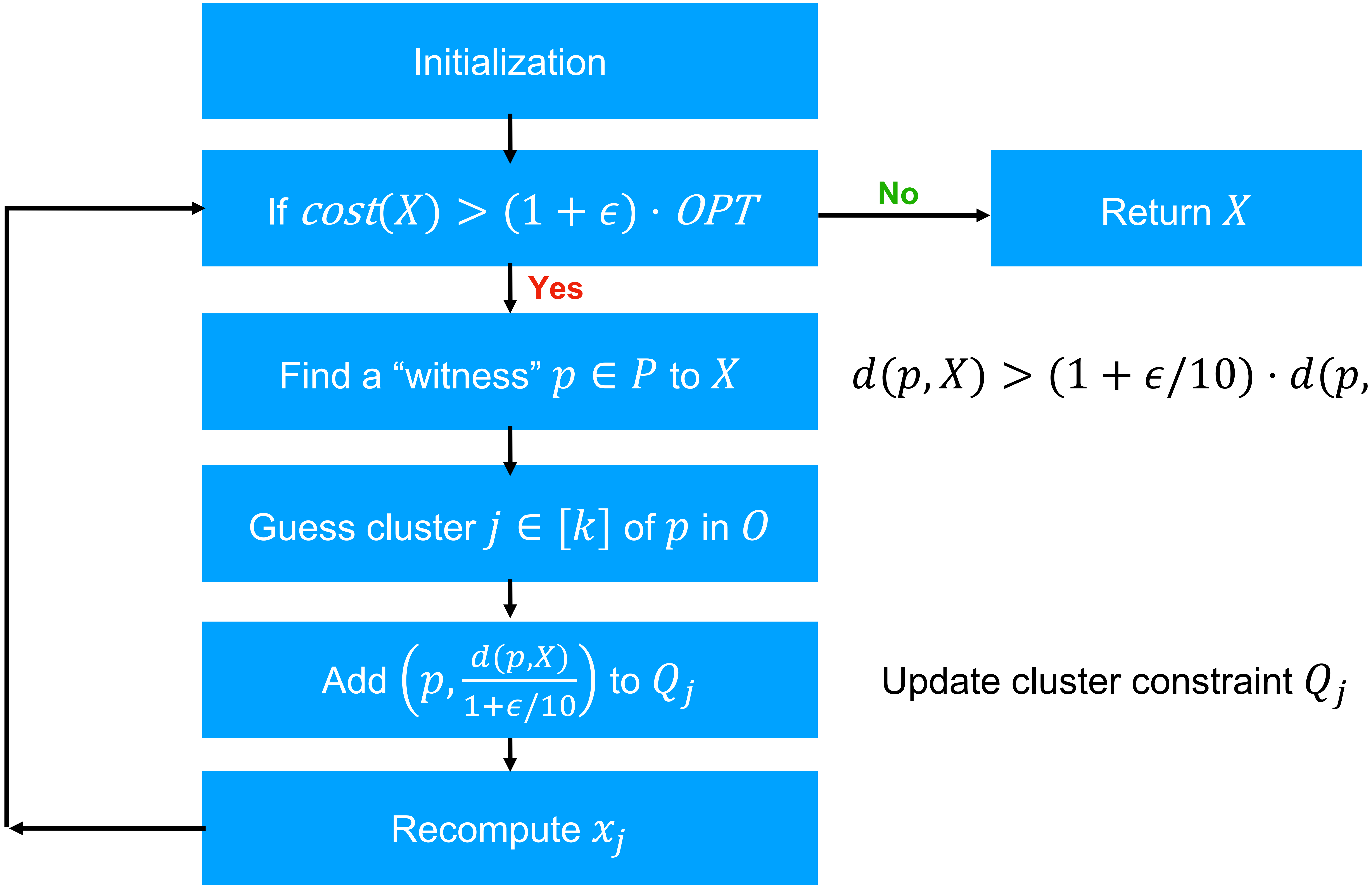
Bound #iterations?

ϵ-scatter dimension

Fix Q_j



ϵ-scattering



Initialization
If $cost(X) > (1+\epsilon) \cdot OPT$
No
Return X
Yes
Find a “witness” $p \in P$ to X
Guess cluster $j \in [k]$ of p in O
Add $\left(p, \frac{d(p,X)}{1+\epsilon/10}\right)$ to Q_j
Recompute x_j

$d(p, X) > (1 + \epsilon/10) \cdot d(p, O)$

Update cluster constraint Q_j

ϵ-scatter dimension of a metric space is λ ⇨ if radius aspect ratio is bounded, then the length is bounded

Unified-EPAS

Bound #iterations?

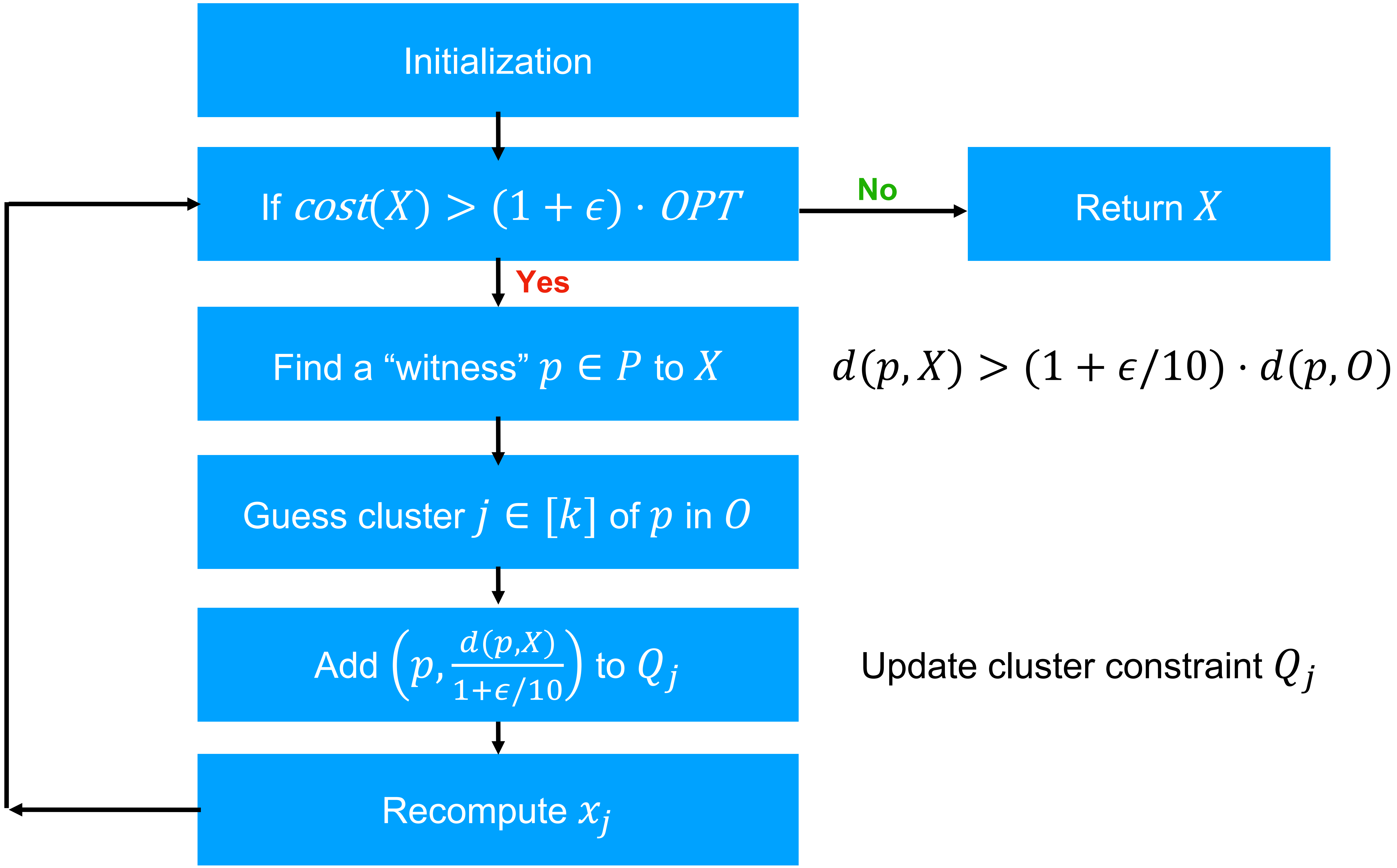
Upper Bounds

Initialization

Compute Upper bounds

Initialize Cluster constraints $Q_1, \ldots, Q_k$ using upper bounds

Initialize solution $X = (x_1, \ldots, x_k)$ using $Q_1, \ldots, Q_k$



Initialization

If $cost(X) > (1 + \epsilon) \cdot OPT$

No

Return X

Yes

Find a “witness” $p \in P$ to X

$d(p, X) > (1 + \epsilon/10) \cdot d(p, O)$

Guess cluster $j \in [k]$ of p in O

Add $\left(p, \frac{d(p,X)}{1+\epsilon/10}\right)$ to Q_j

Update cluster constraint Q_j

Recompute x_j

Lemma 2

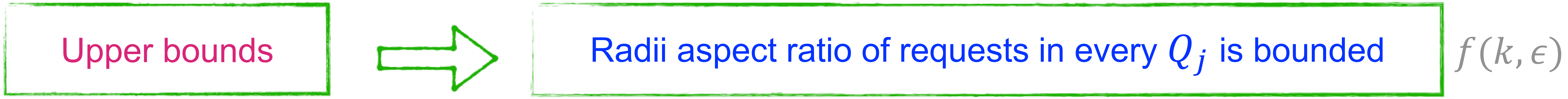
Upper bounds ⇨ Radii aspect ratio of requests in every Q_j is bounded

Unified-EPAS

Lemma 1

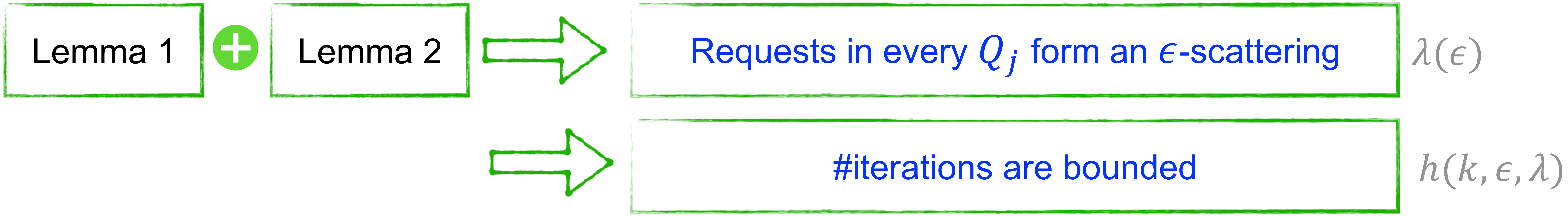
If $cost(X) > (1+\epsilon)\cdot OPT$, then we can find a witness to X w.h.p. $g(k,\epsilon)$

Lemma 2



Upper bounds ⇒ Radii aspect ratio of requests in every Q_j is bounded $f(k,\epsilon)$

Lemma 3 (Theorem)



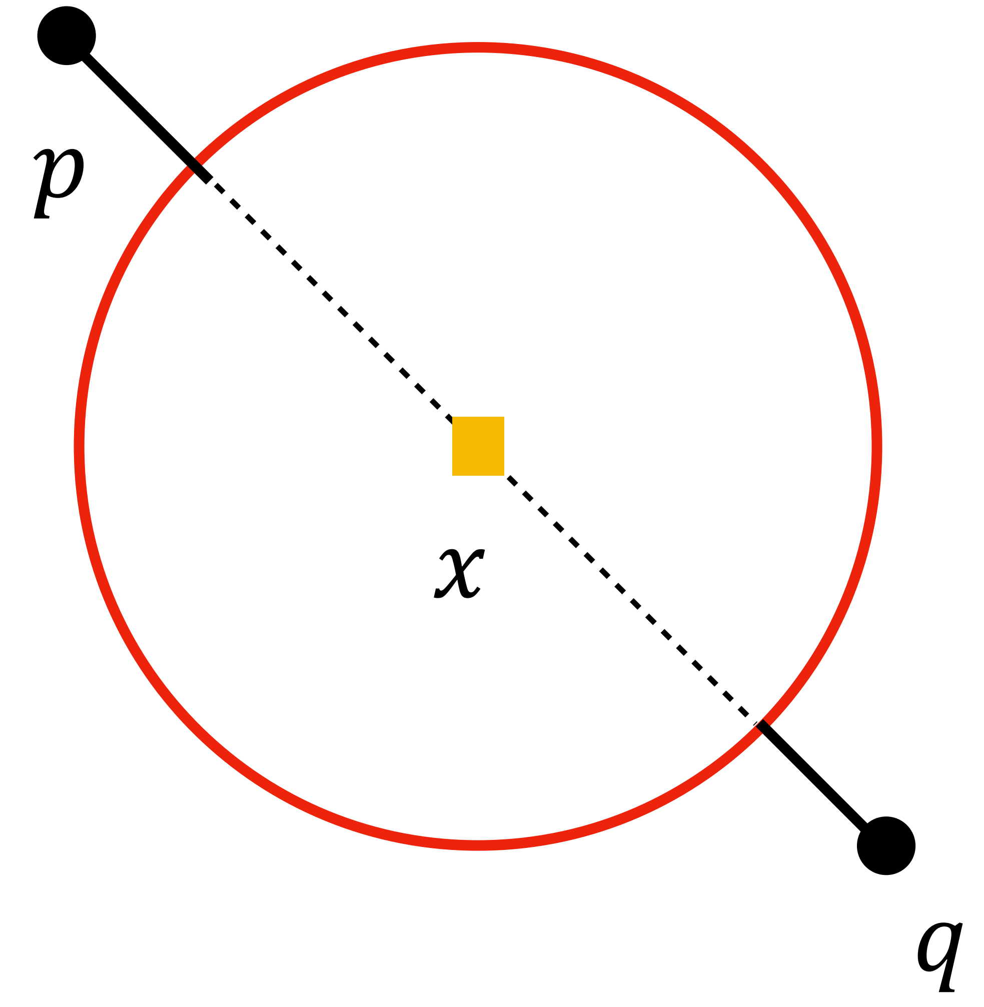
Lemma 1 + Lemma 2 ⇒ Requests in every Q_j form an ϵ-scattering $\lambda(\epsilon)$

⇒ #iterations are bounded $h(k,\epsilon,\lambda)$

■

Hybrid Clustering

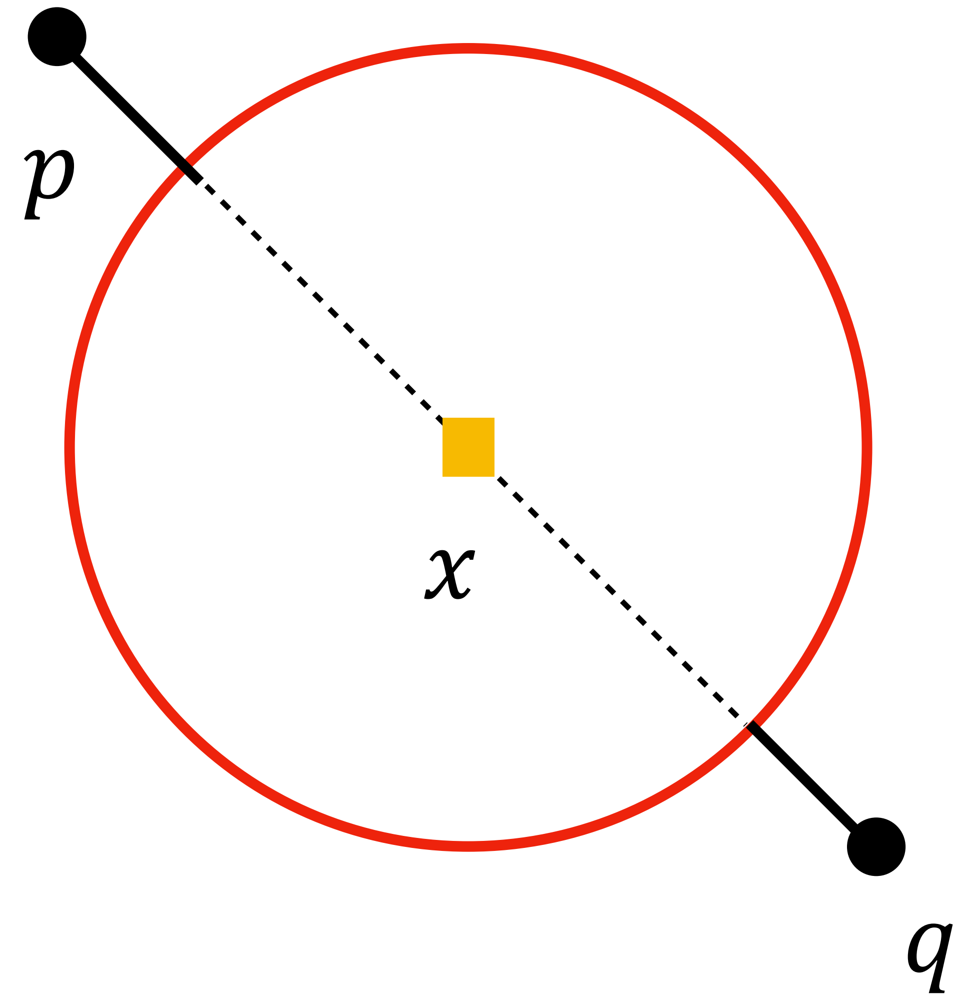
d_r does not satisfy triangle inequality



- Computing Upper bounds fails!
- Sampling lemma (Lemma 1) does not work!
- Radii Aspect Ratio lemma (Lemma 2) fails!
- Iteration lemma (Lemma 3) does not apply since the new requests may not be feasible!

Hybrid Clustering

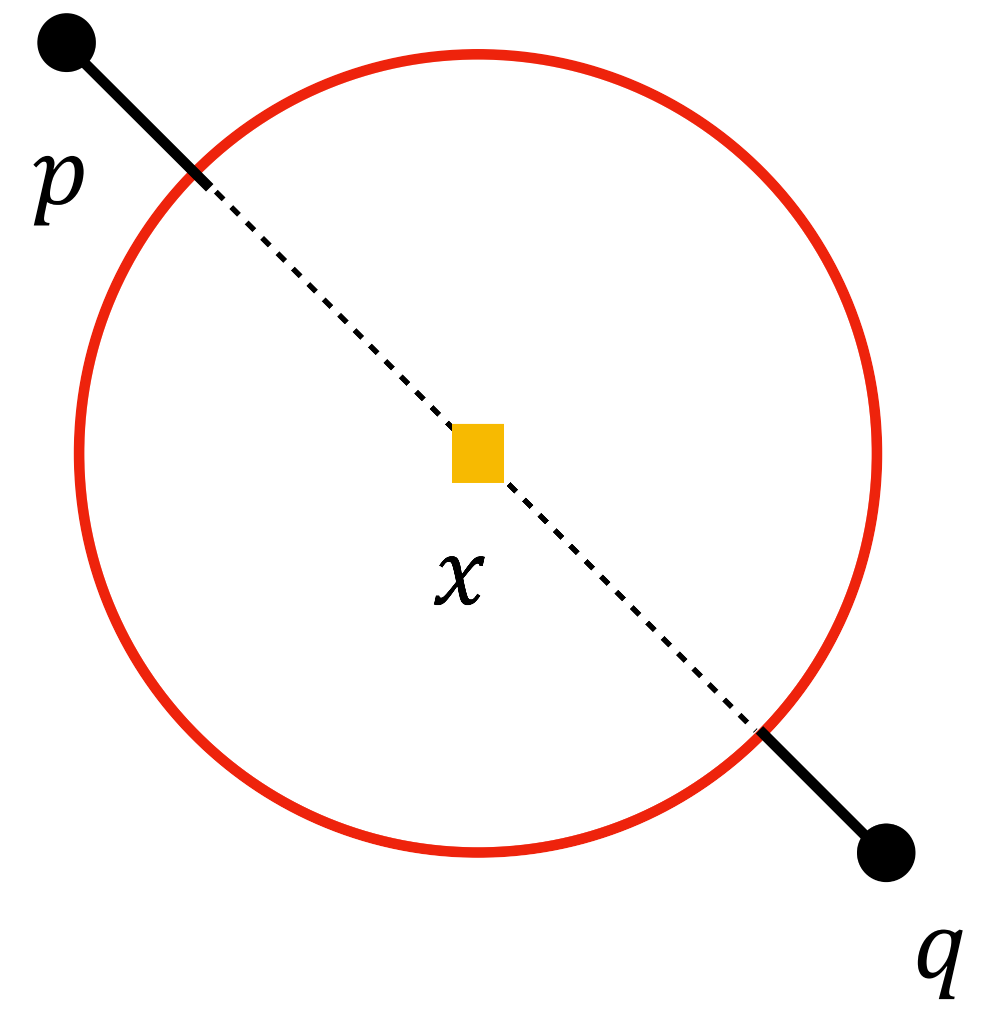
d_r does not satisfy triangle inequality



- Computing Upper bounds fails!
- Sampling lemma (Lemma 1) does not work!
- Radii Aspect Ratio lemma (Lemma 2) fails!
- Iteration lemma (Lemma 3) does not apply since the new requests may not be feasible!

Hybrid Clustering

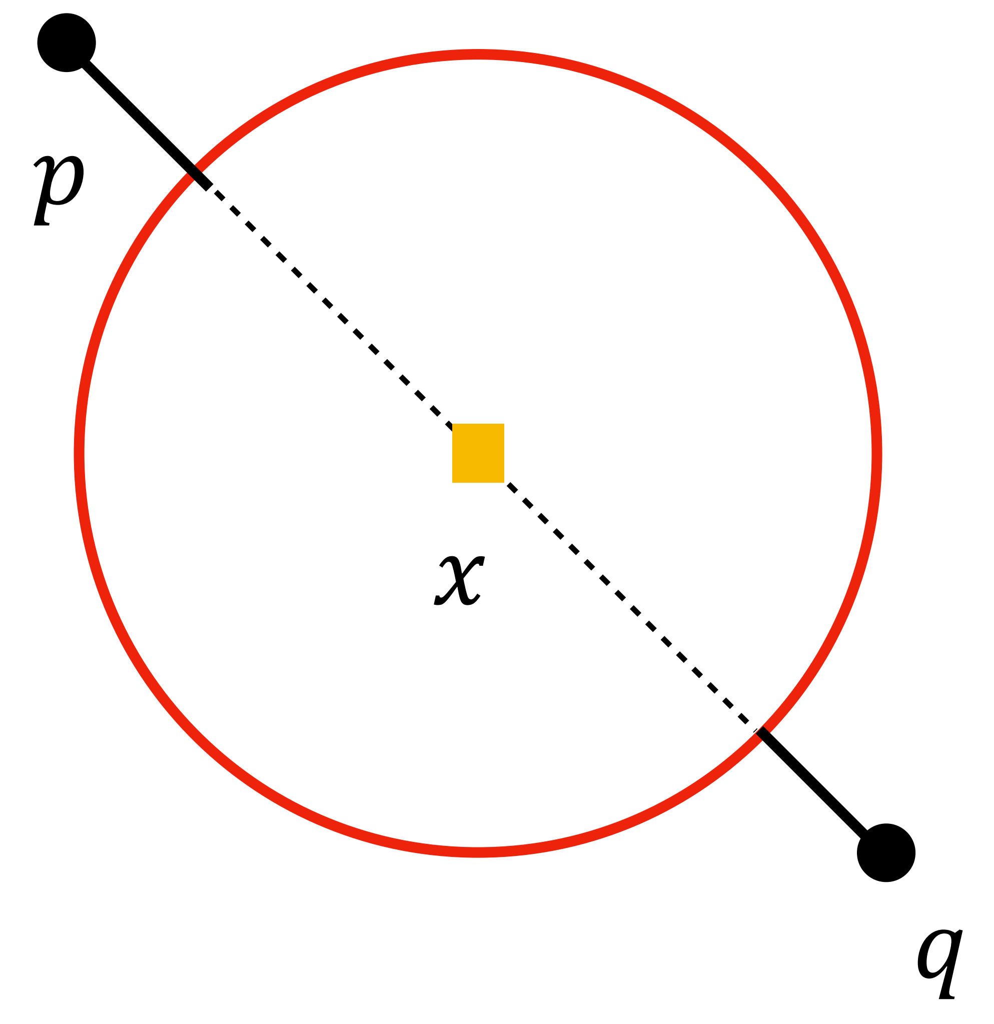
d_r does not satisfy triangle inequality



- Computing Upper bounds fails!
- Sampling lemma (Lemma 1) does not work!
- Radii Aspect Ratio lemma (Lemma 2) fails!
- Iteration lemma (Lemma 3) does not apply since the new requests may not be feasible!

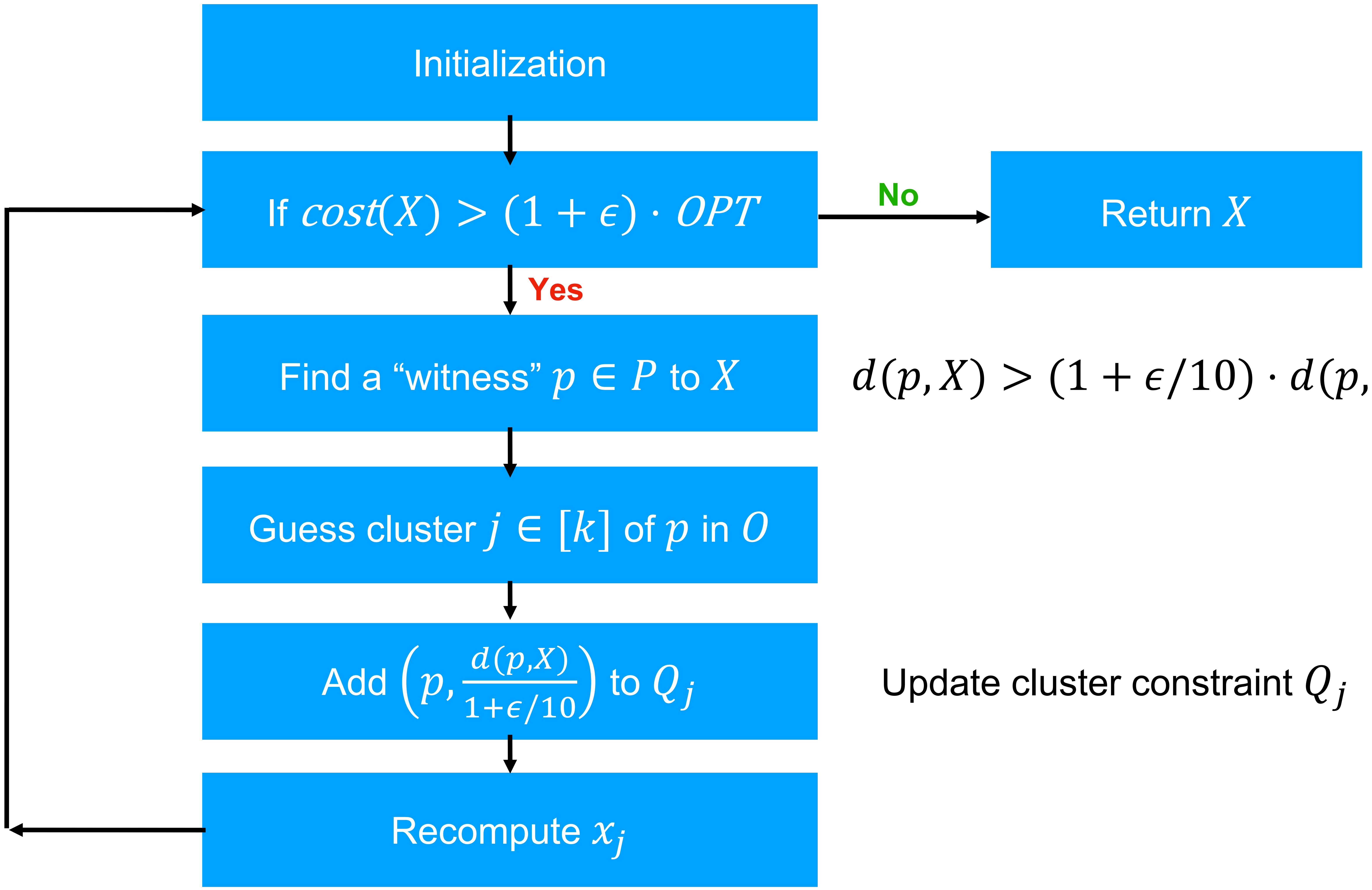
Hybrid Clustering

d_r does not satisfy triangle inequality



- Computing Upper bounds fails!
- Sampling lemma (Lemma 1) does not work!
- Radii Aspect Ratio lemma (Lemma 2) fails!
- Iteration lemma (Lemma 3) does not apply since the new requests may not be feasible!

Unified-EPAS



Initialization

If $cost(X) > (1 + \epsilon) \cdot OPT$

No

Return X

Yes

Find a “witness” $p \in P$ to X

$d(p, X) > (1 + \epsilon/10) \cdot d(p, O)$

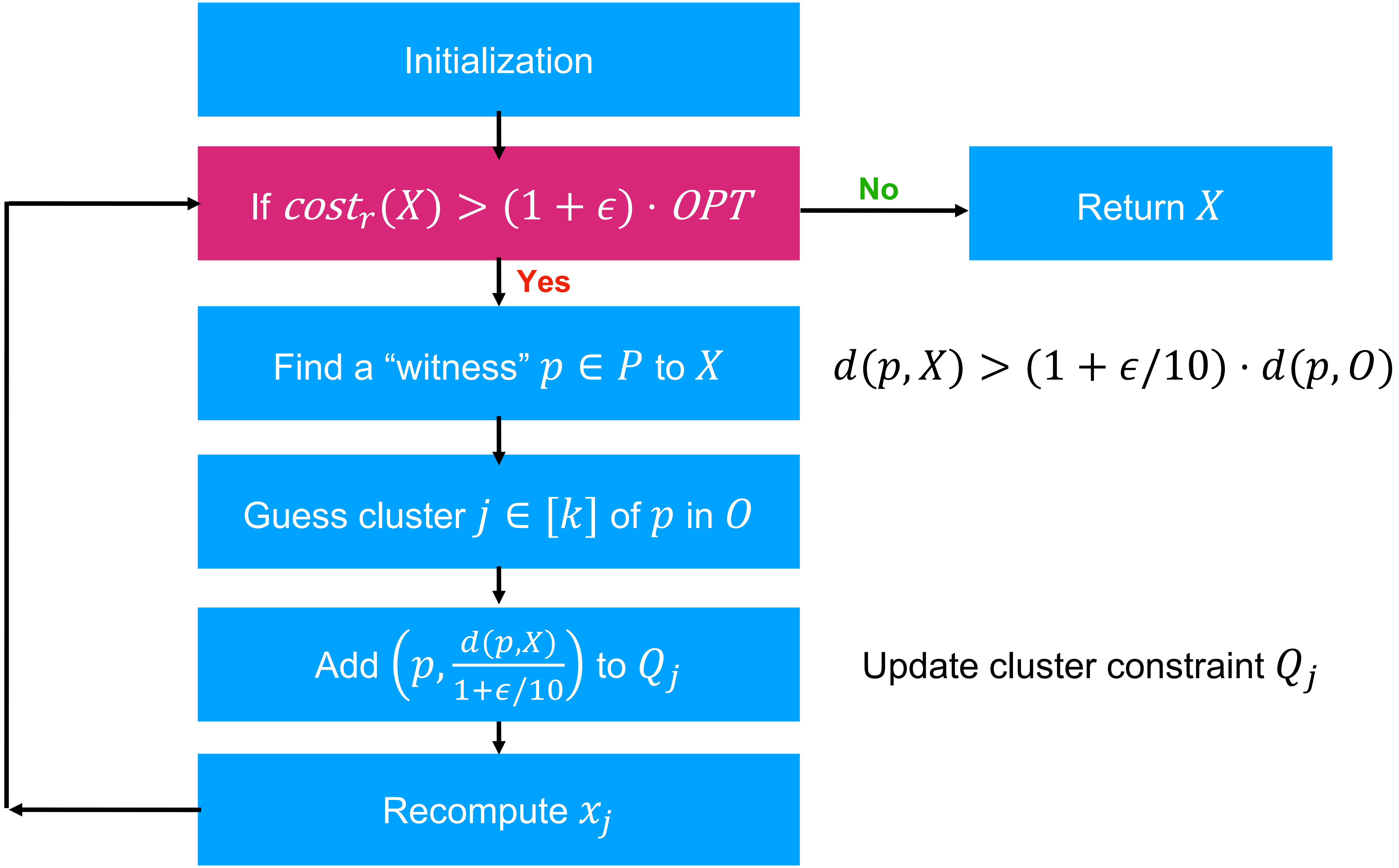
Guess cluster $j \in [k]$ of p in O

Add $\left(p, \frac{d(p,X)}{1+\epsilon/10}\right)$ to Q_j

Update cluster constraint Q_j

Recompute x_j

Attempt 1



Initialization

If $cost_r(X) > (1+\epsilon)\cdot OPT$

No

Return X

Yes

Find a “witness” $p \in P$ to X

$d(p,X) > (1+\epsilon/10)\cdot d(p,O)$

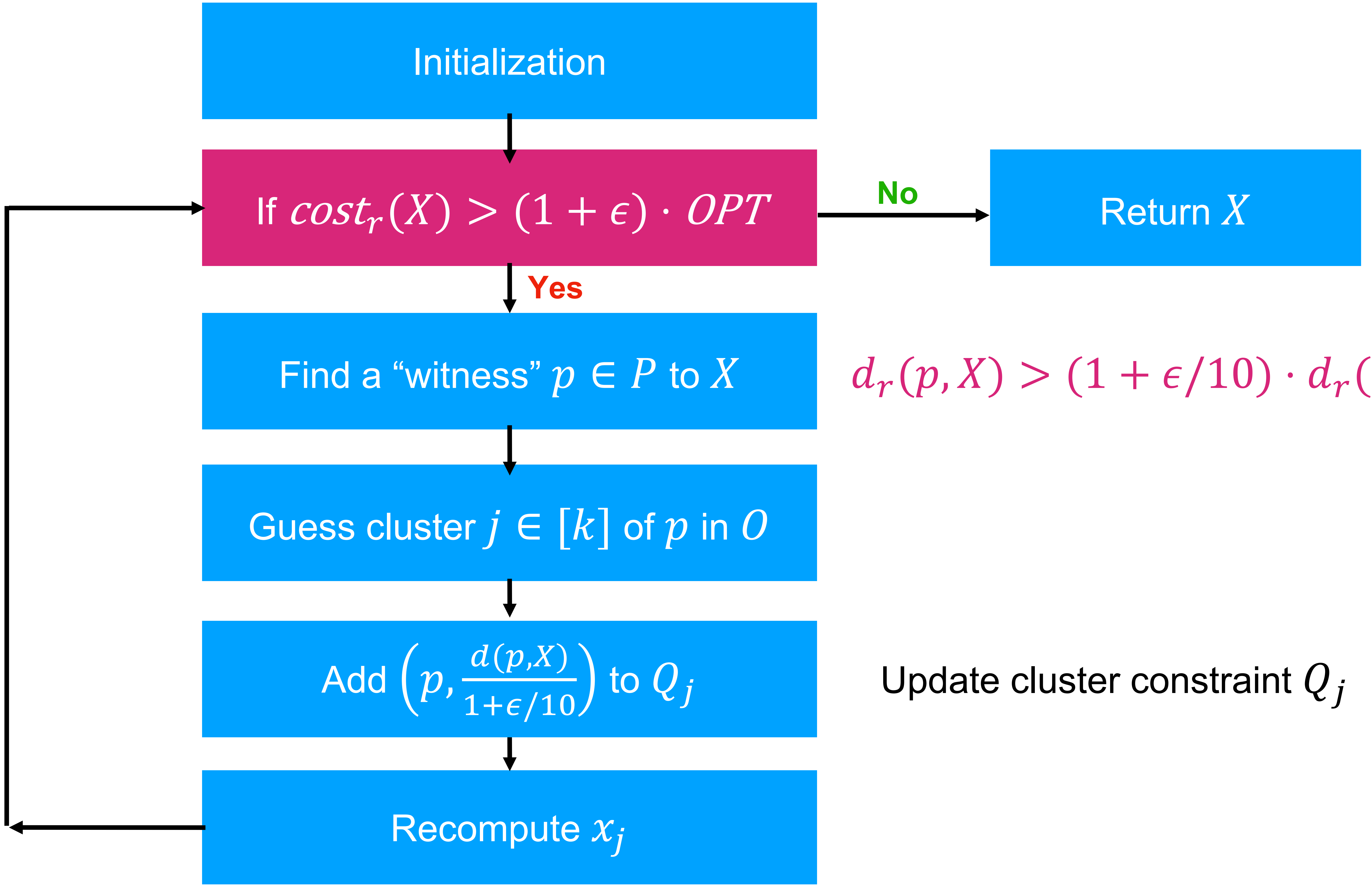
Guess cluster $j \in [k]$ of p in O

Add $\left(p, \frac{d(p,X)}{1+\epsilon/10}\right)$ to Q_j

Update cluster constraint Q_j

Recompute x_j

Attempt 1



Initialization

If $cost_r(X) > (1+\epsilon) \cdot OPT$

No

Return X

Yes

Find a “witness” $p \in P$ to X

Guess cluster $j \in [k]$ of p in O

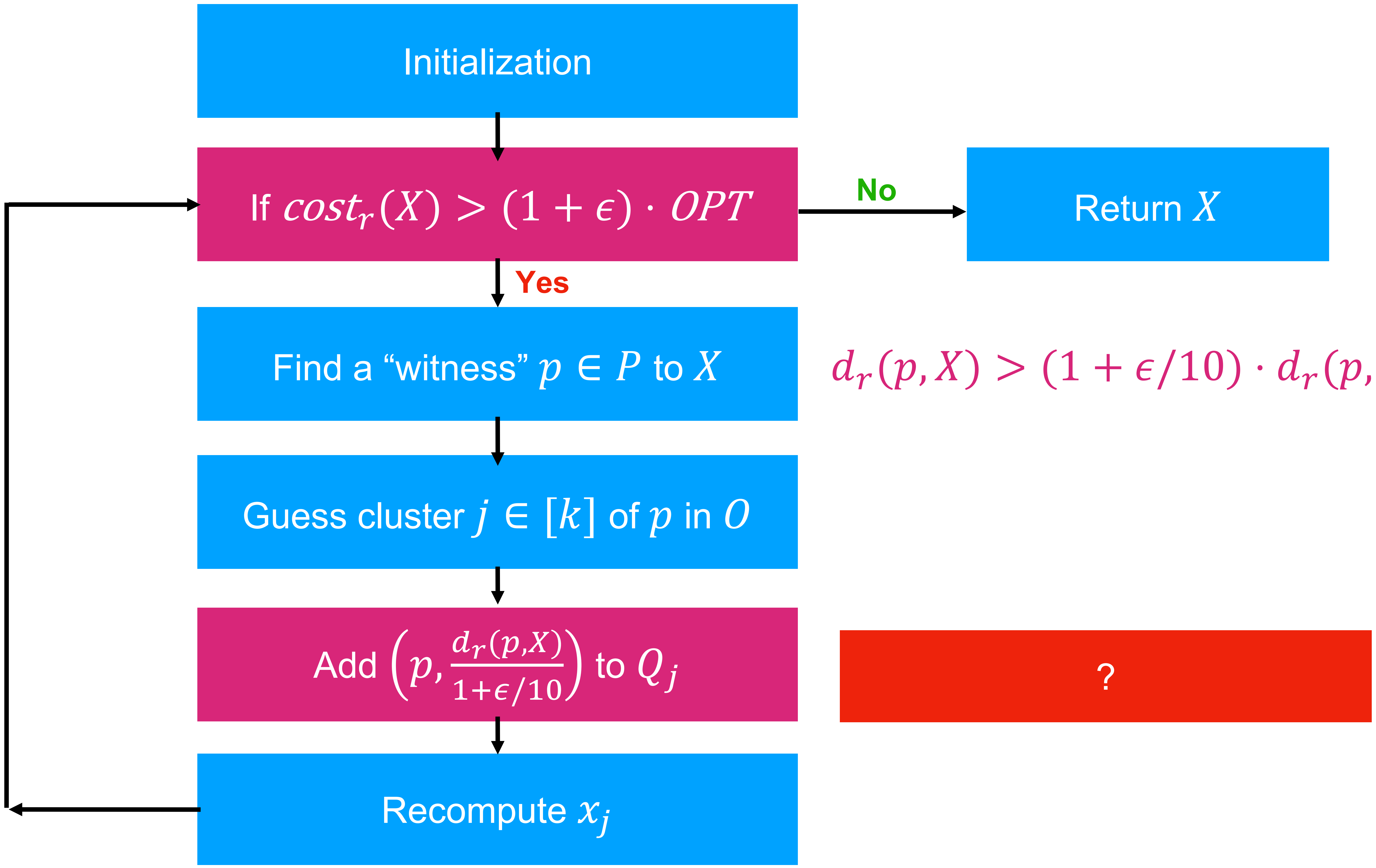
Add $\left(p, \frac{d(p,X)}{1+\epsilon/10}\right)$ to Q_j

Recompute x_j

$d_r(p,X) > (1+\epsilon/10) \cdot d_r(p,O)$

Update cluster constraint Q_j

Attempt 1



Initialization

If $cost_r(X) > (1+\epsilon) \cdot OPT$

No

Return X

Yes

Find a “witness” $p \in P$ to X

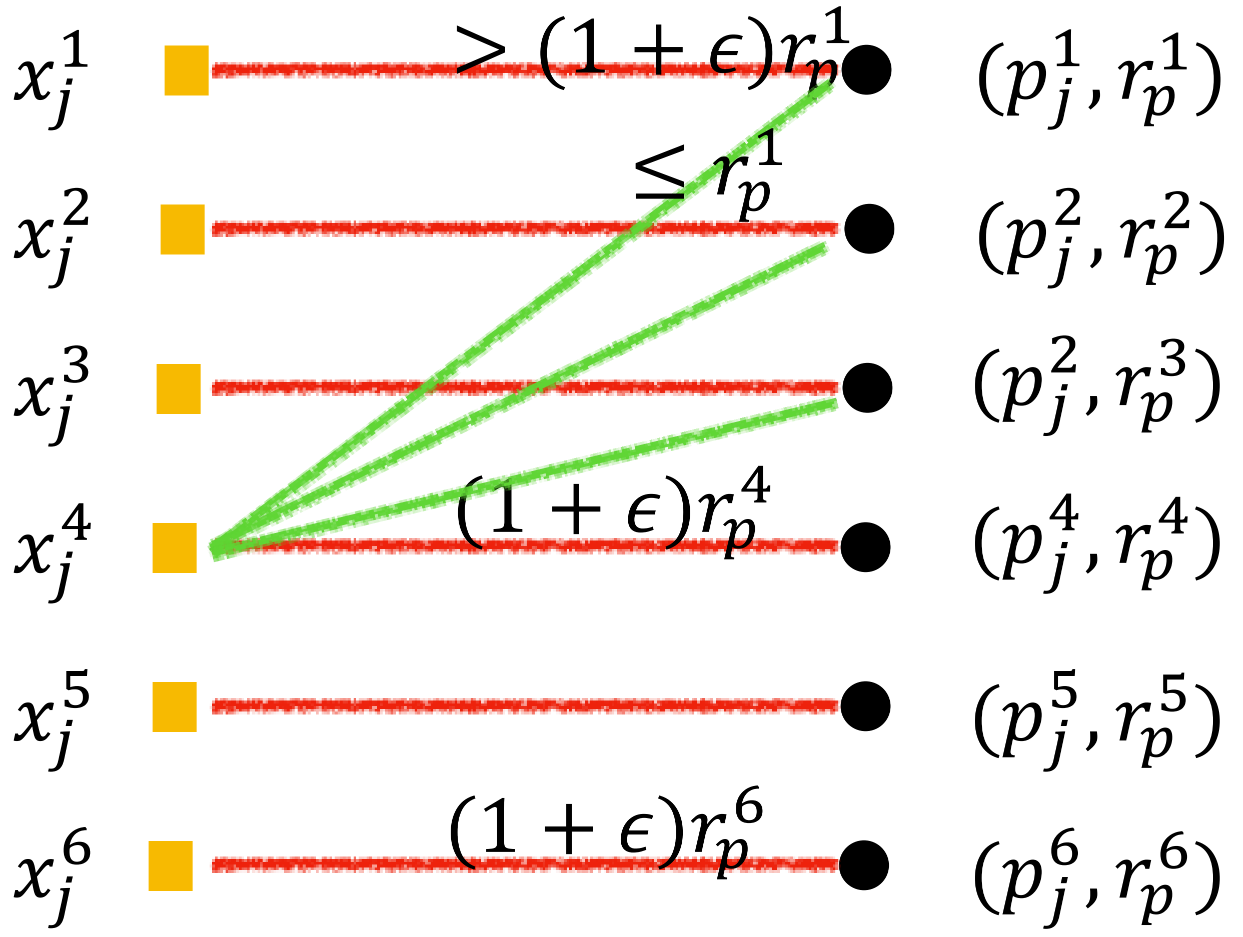
$d_r(p, X) > (1 + \epsilon/10) \cdot d_r(p, O)$

Guess cluster $j \in [k]$ of p in O

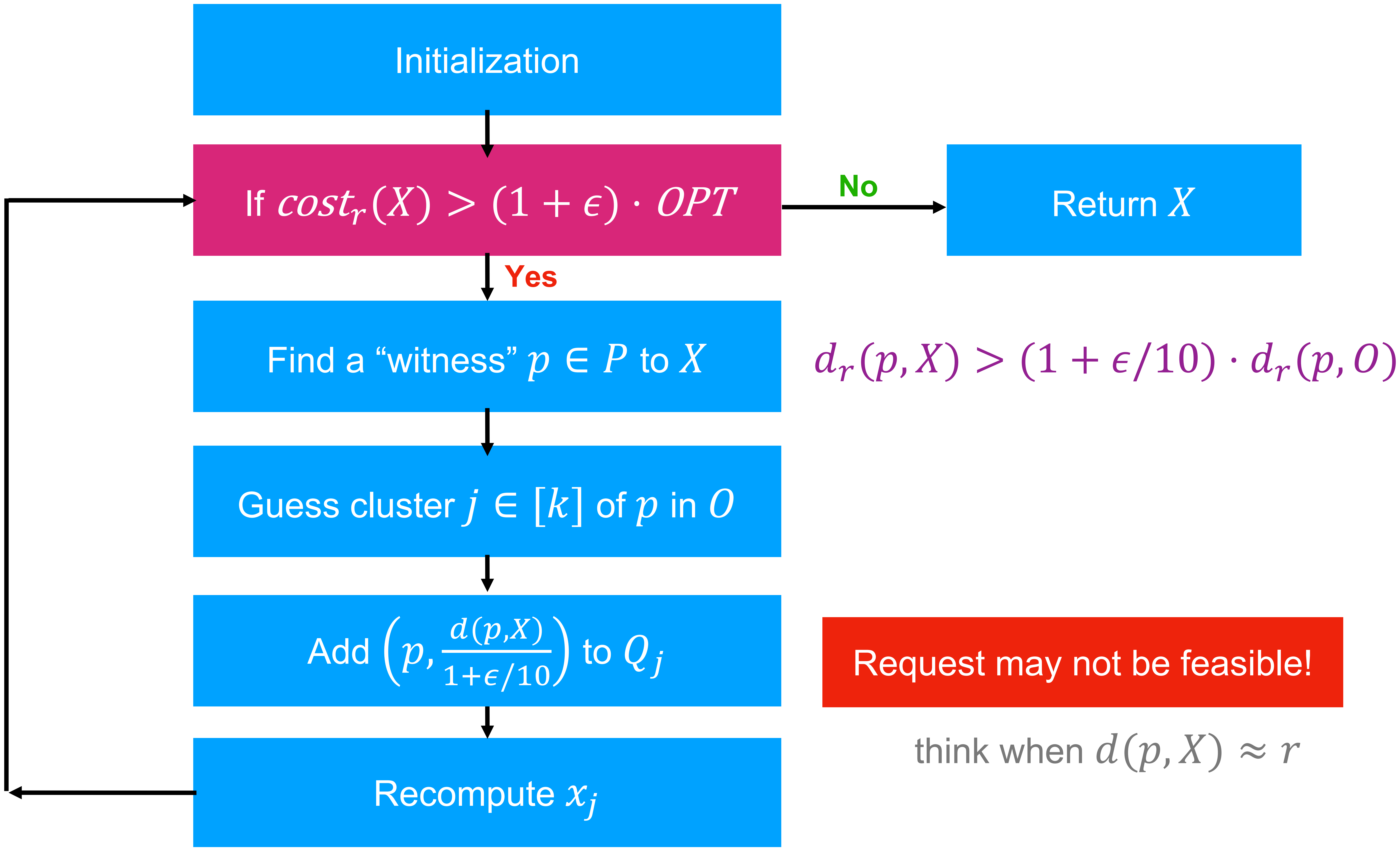
Add $\left(p, \frac{d_r(p,X)}{1+\epsilon/10}\right)$ to Q_j

?

Recompute x_j



Attempt 1



Initialization

If $cost_r(X) > (1+\epsilon) \cdot OPT$

No

Return X

Yes

Find a "witness" $p \in P$ to X

$d_r(p, X) > (1 + \epsilon/10) \cdot d_r(p, O)$

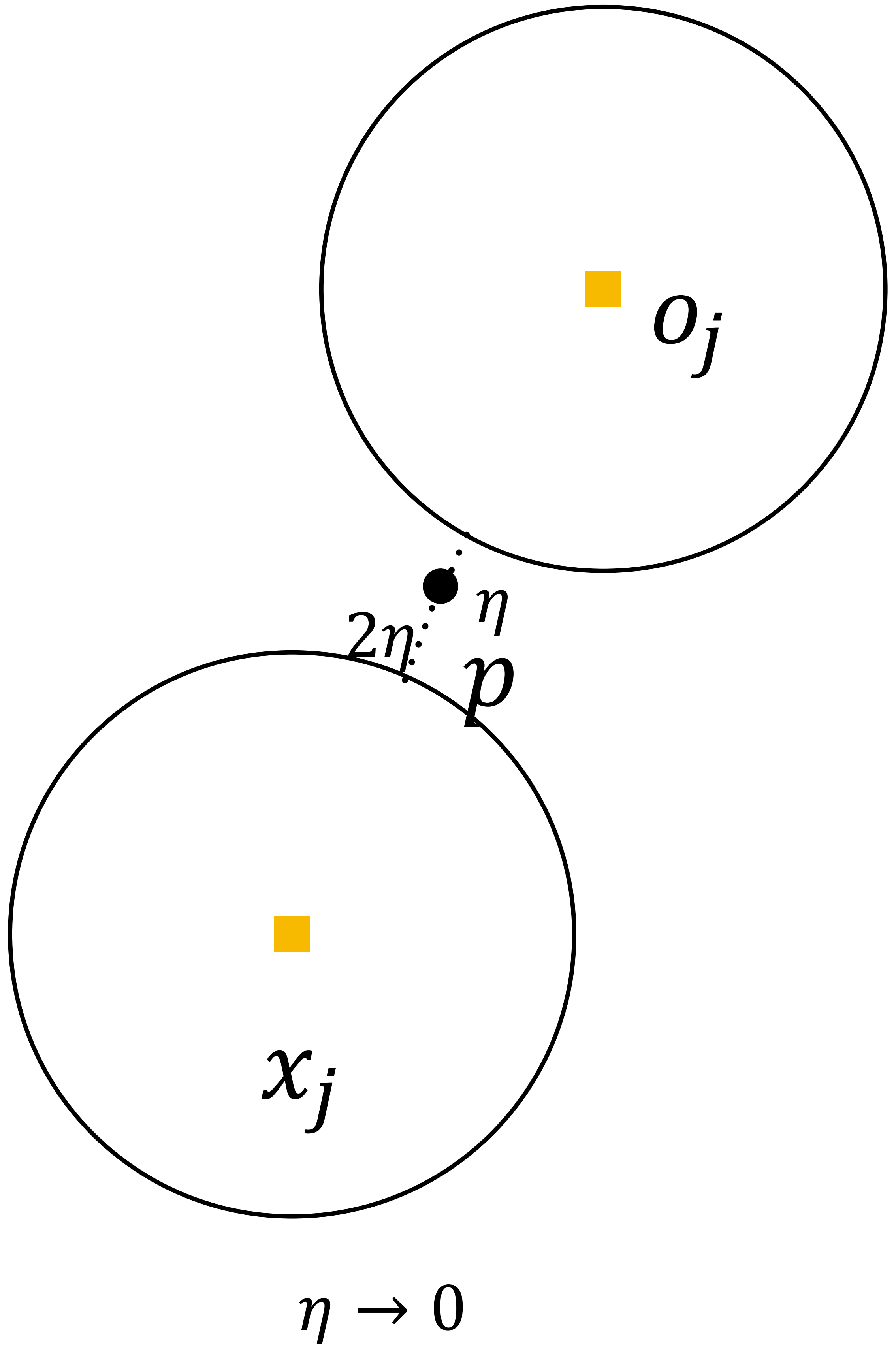
Guess cluster $j \in [k]$ of p in O

Add $\left(p, \frac{d(p,X)}{1+\epsilon/10}\right)$ to Q_j

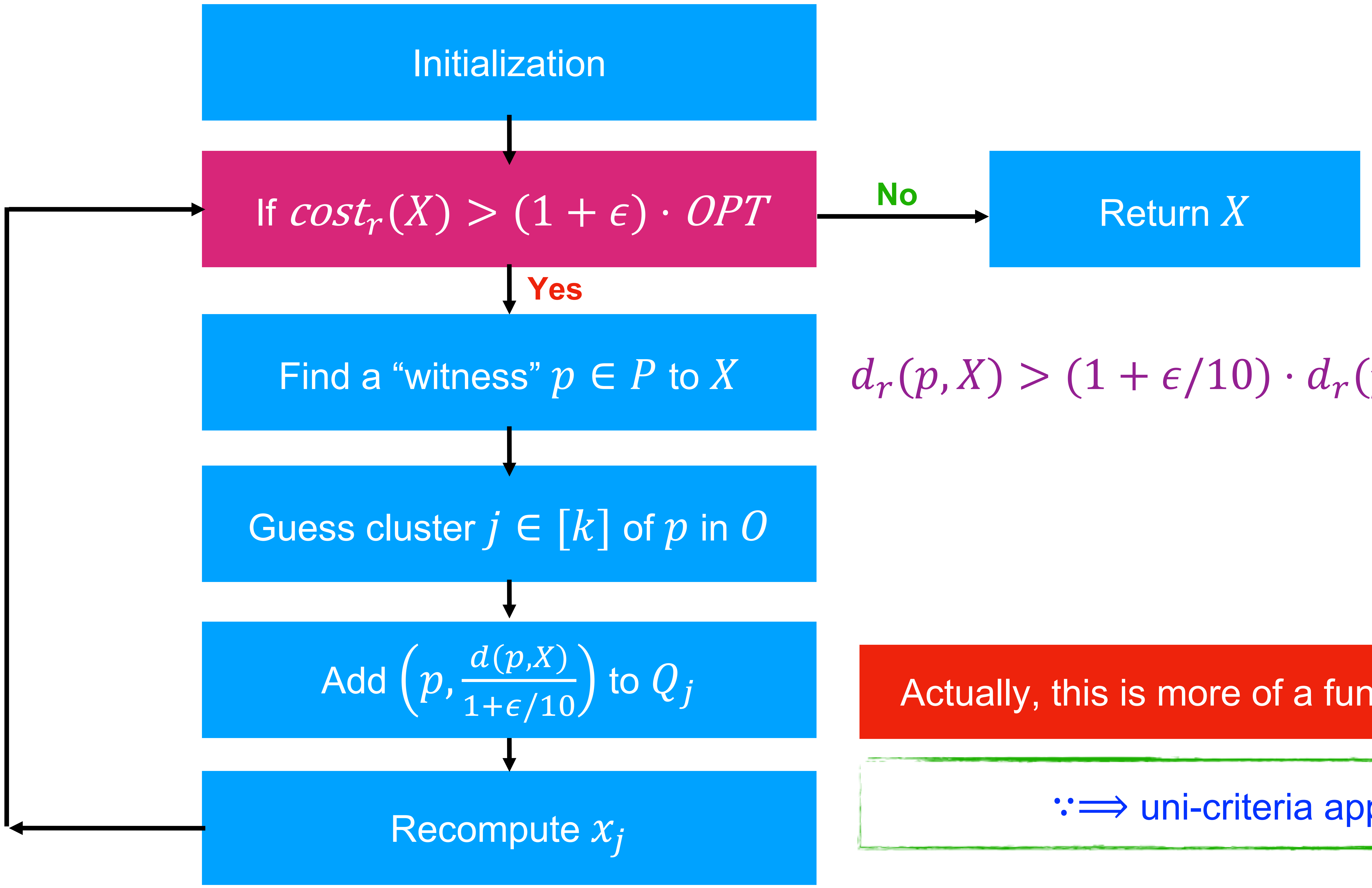
Request may not be feasible!

think when $d(p, X) \approx r$

Recompute x_j



Attempt 1



Initialization

If $cost_r(X) > (1+\epsilon) \cdot OPT$

No

Return X

Yes

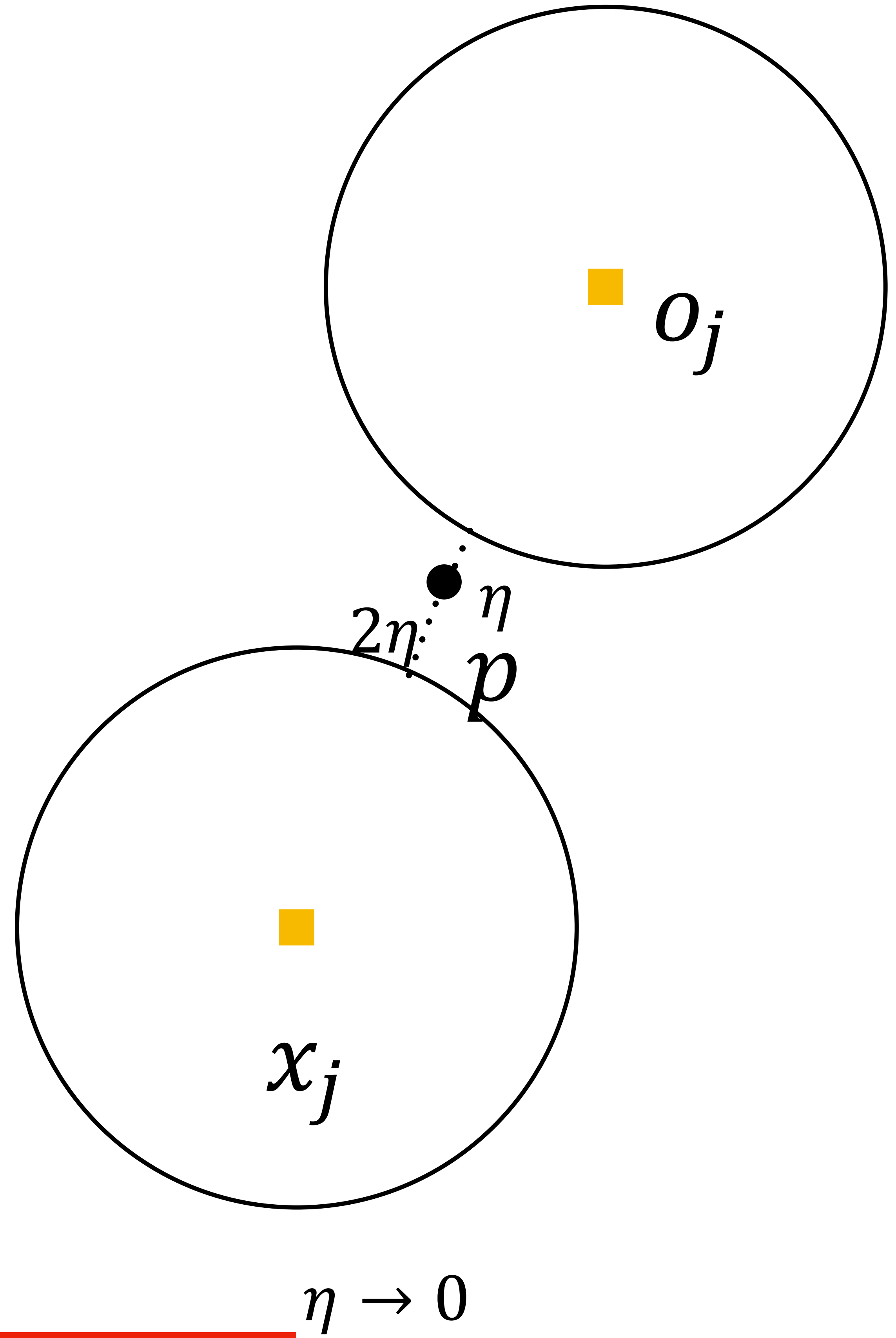
Find a “witness” $p \in P$ to X

Guess cluster $j \in [k]$ of p in O

Add $\left(p, \frac{d(p,X)}{1+\epsilon/10}\right)$ to Q_j

Recompute x_j

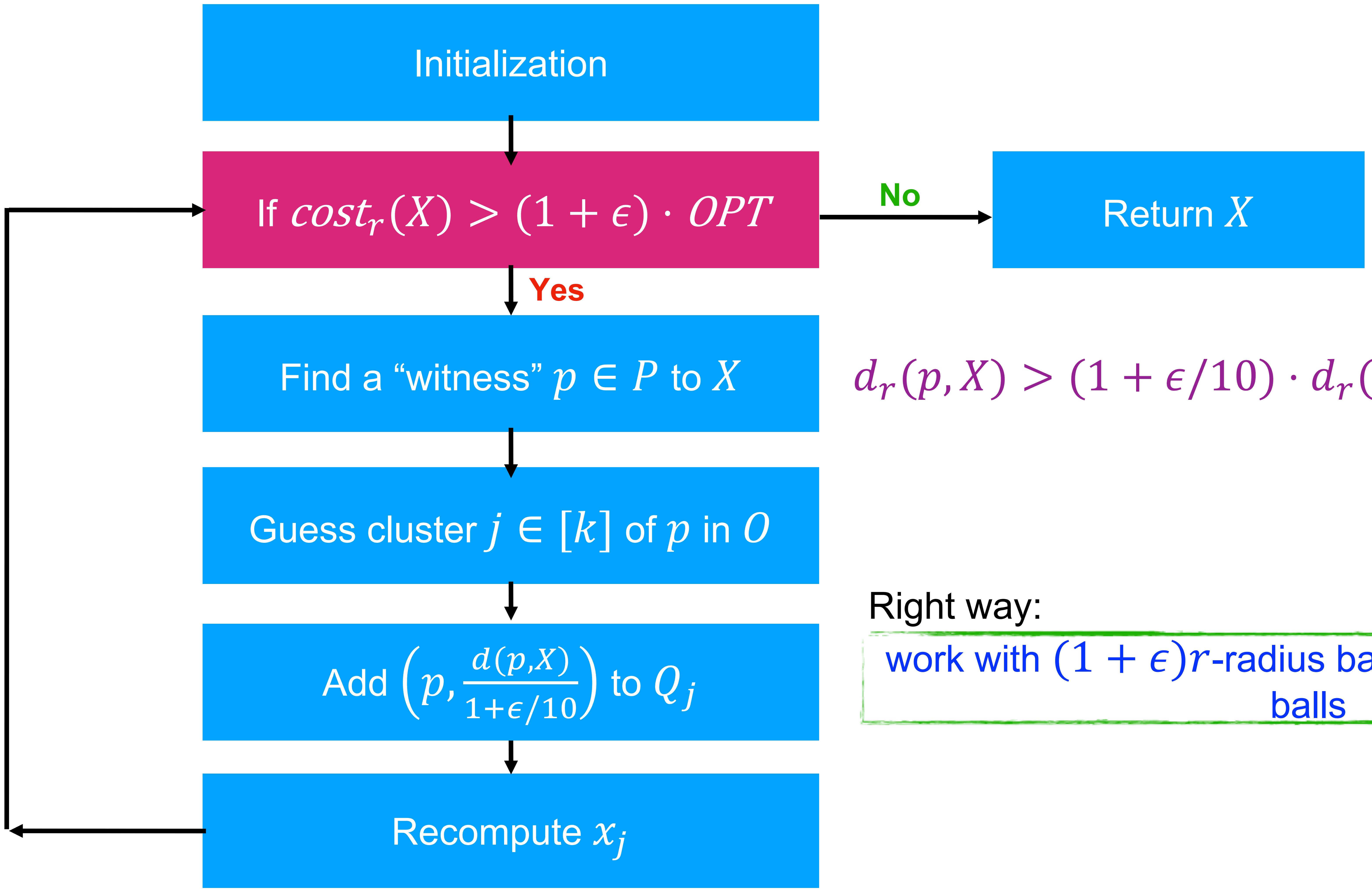
$d_r(p,X) > (1+\epsilon/10) \cdot d_r(p,O)$



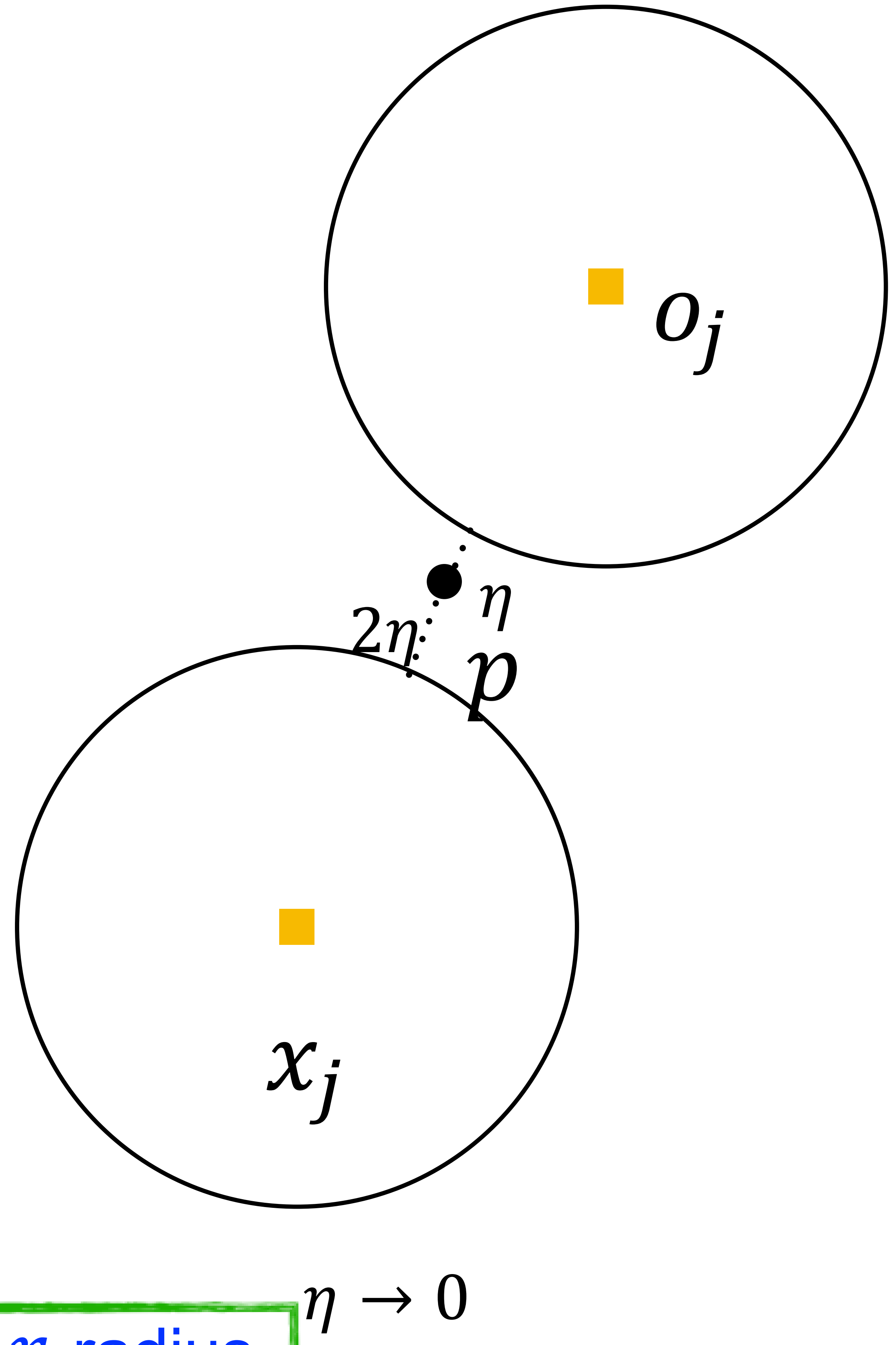
Actually, this is more of a fundamental bottleneck

$\because \Longrightarrow$ uni-criteria approximation

Attempt 1


Initialization
If $cost_r(X) > (1+\epsilon)\cdot OPT$
No
Return X
Yes
Find a “witness” $p \in P$ to X
Guess cluster $j \in [k]$ of p in O
Add $\left(p, \frac{d(p,X)}{1+\epsilon/10}\right)$ to Q_j
Recompute x_j

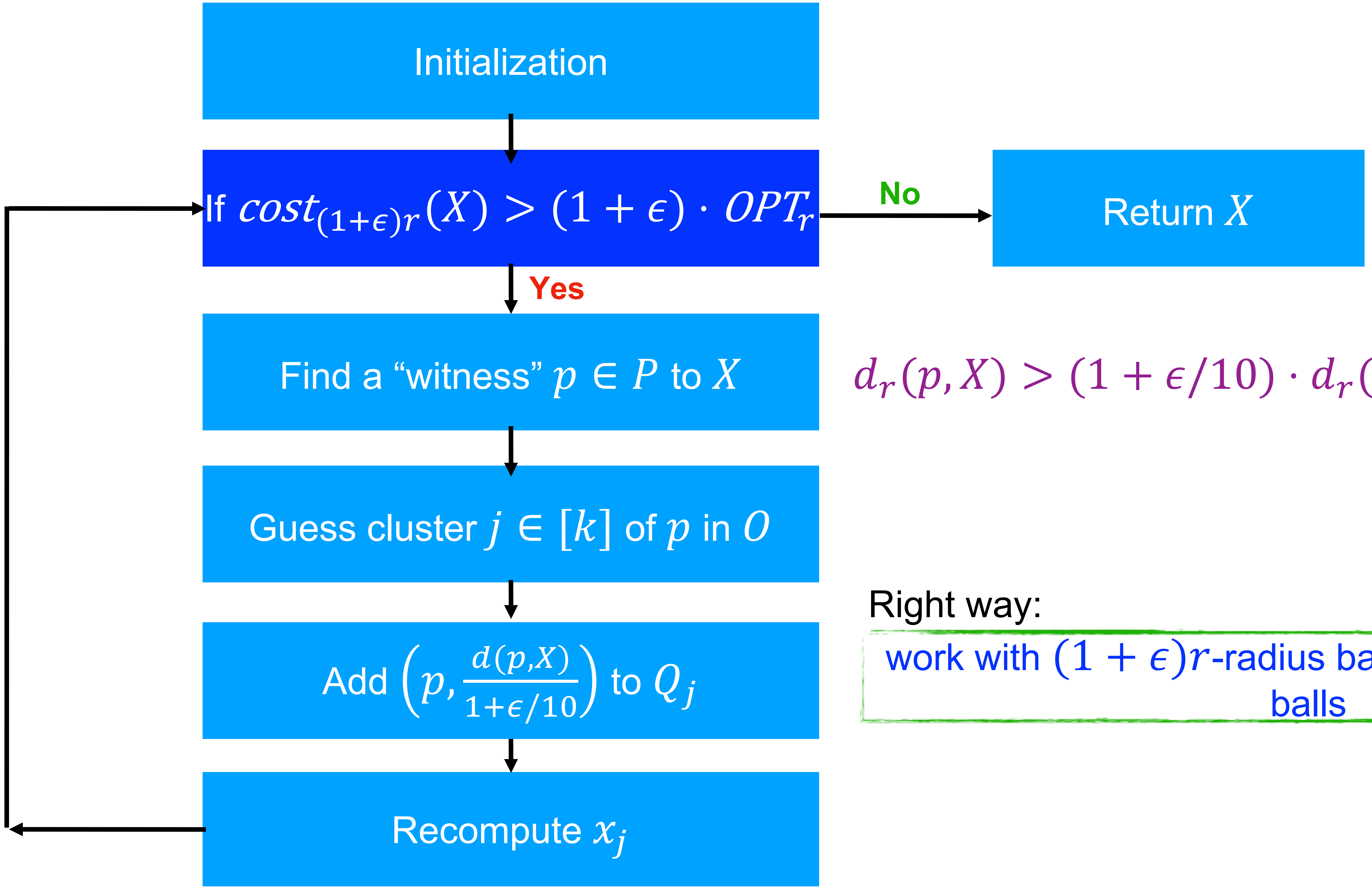
$$d_r(p,X) > (1+\epsilon/10)\cdot d_r(p,O)$$



Right way:

work with $(1+\epsilon)r$-radius balls instead of r-radius balls

Attempt 1



Initialization

If $cost_{(1+\epsilon)r}(X) > (1+\epsilon) \cdot OPT_r$

No

Return X

Yes

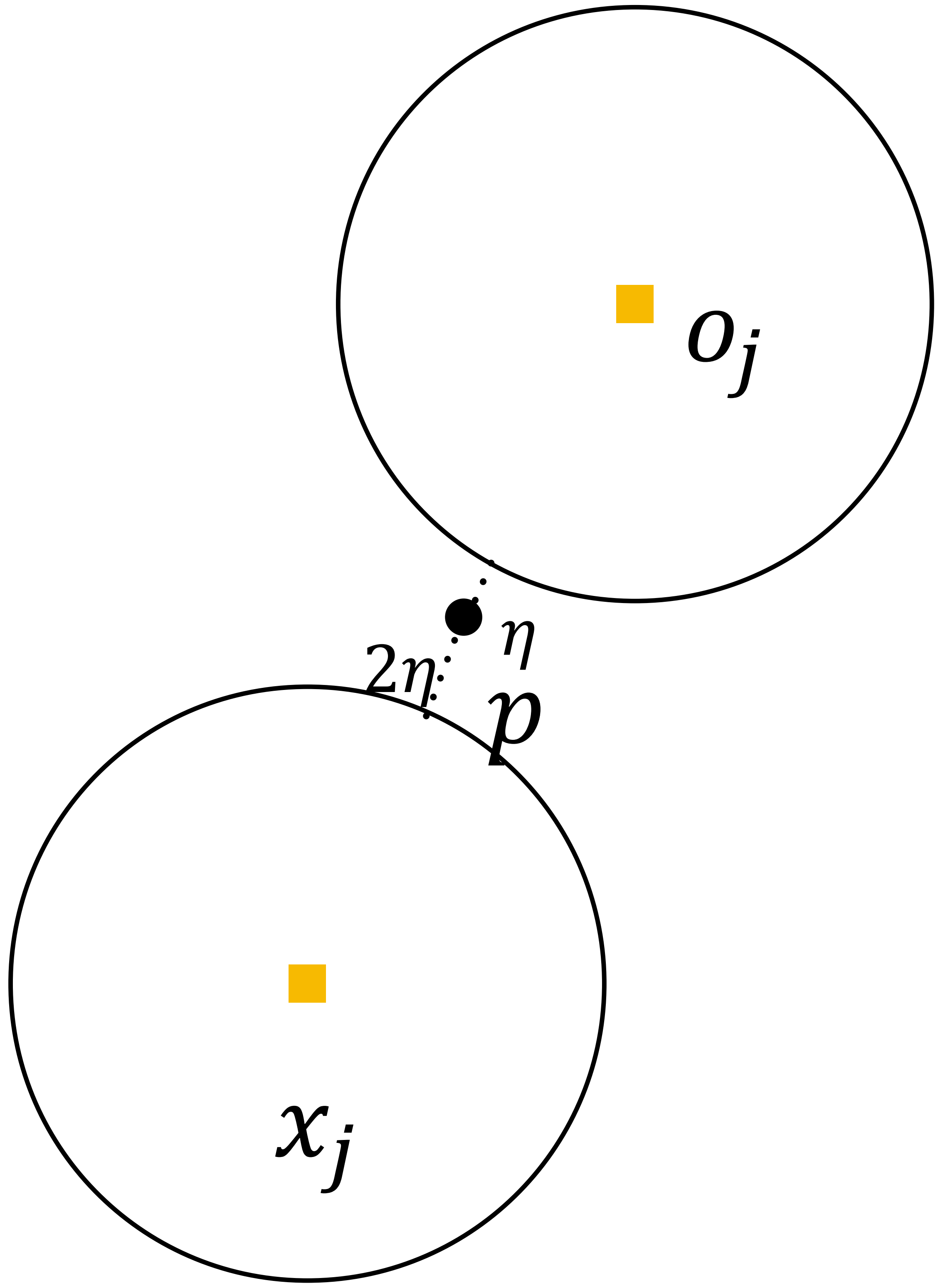
Find a "witness" $p \in P$ to X

Guess cluster $j \in [k]$ of p in O

Add $\left(p, \frac{d(p,X)}{1+\epsilon/10}\right)$ to Q_j

Recompute x_j

$$d_r(p,X) > (1+\epsilon/10) \cdot d_r(p,O)$$

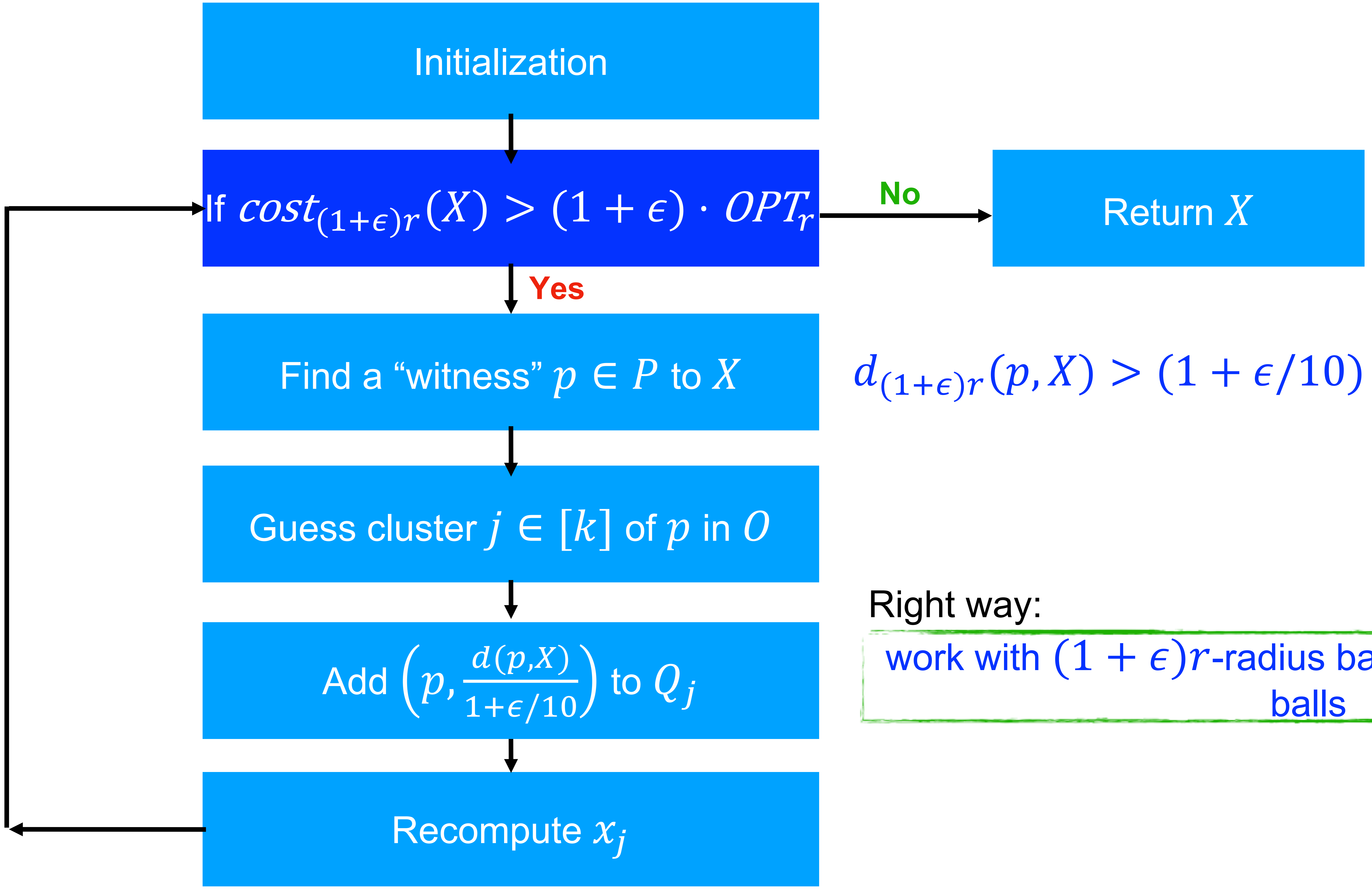


$\eta \to 0$

Right way:

work with $(1+\epsilon)r$-radius balls instead of r-radius balls

Attempt 1



Initialization

If $cost_{(1+\epsilon)r}(X) > (1+\epsilon) \cdot OPT_r$

No

Return X

Yes

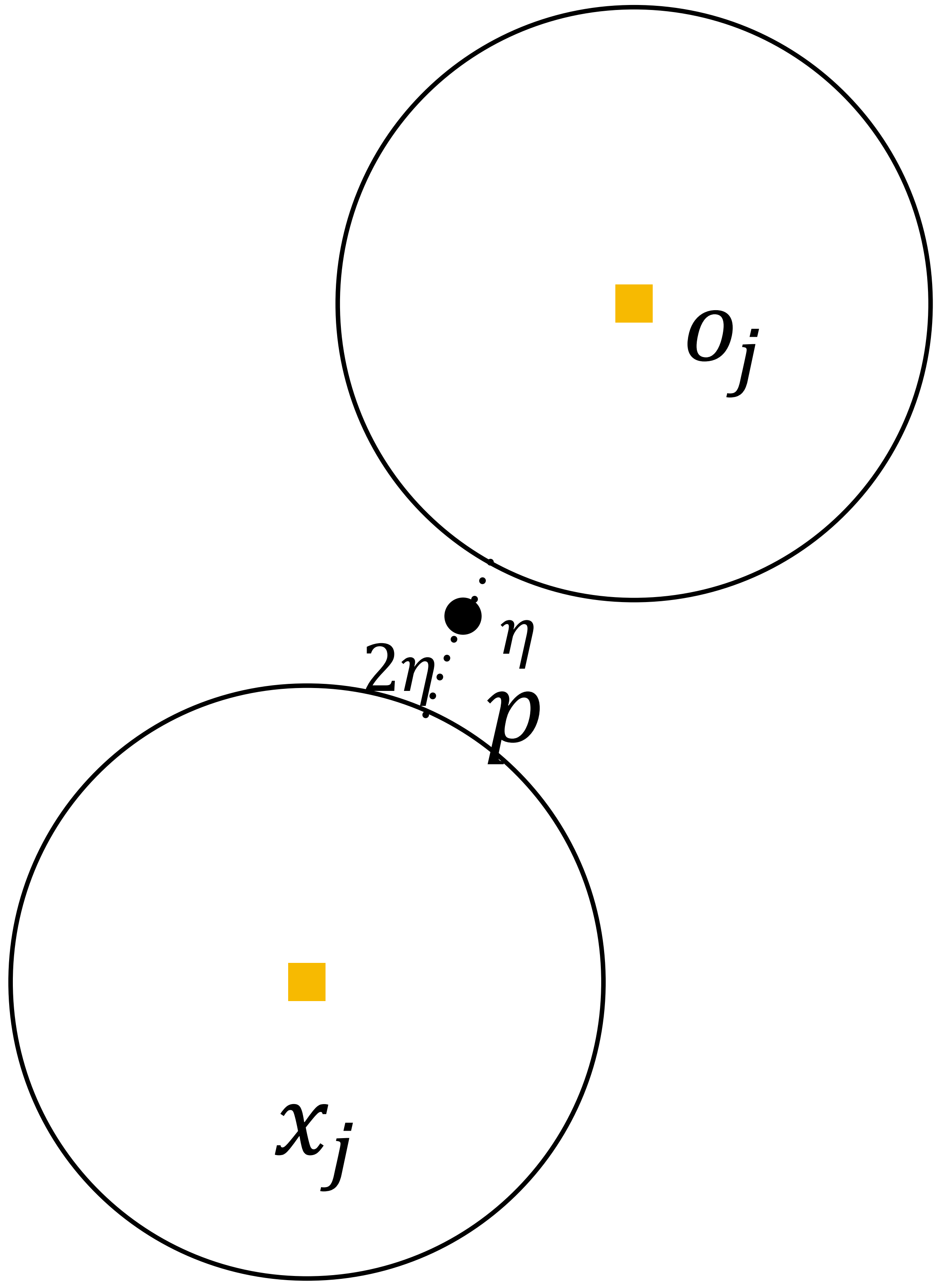
Find a “witness” $p \in P$ to X

Guess cluster $j \in [k]$ of p in O

Add $\left(p, \frac{d(p,X)}{1+\epsilon/10}\right)$ to Q_j

Recompute x_j

$d_{(1+\epsilon)r}(p,X) > (1+\epsilon/10) \cdot d_r(p,O)$

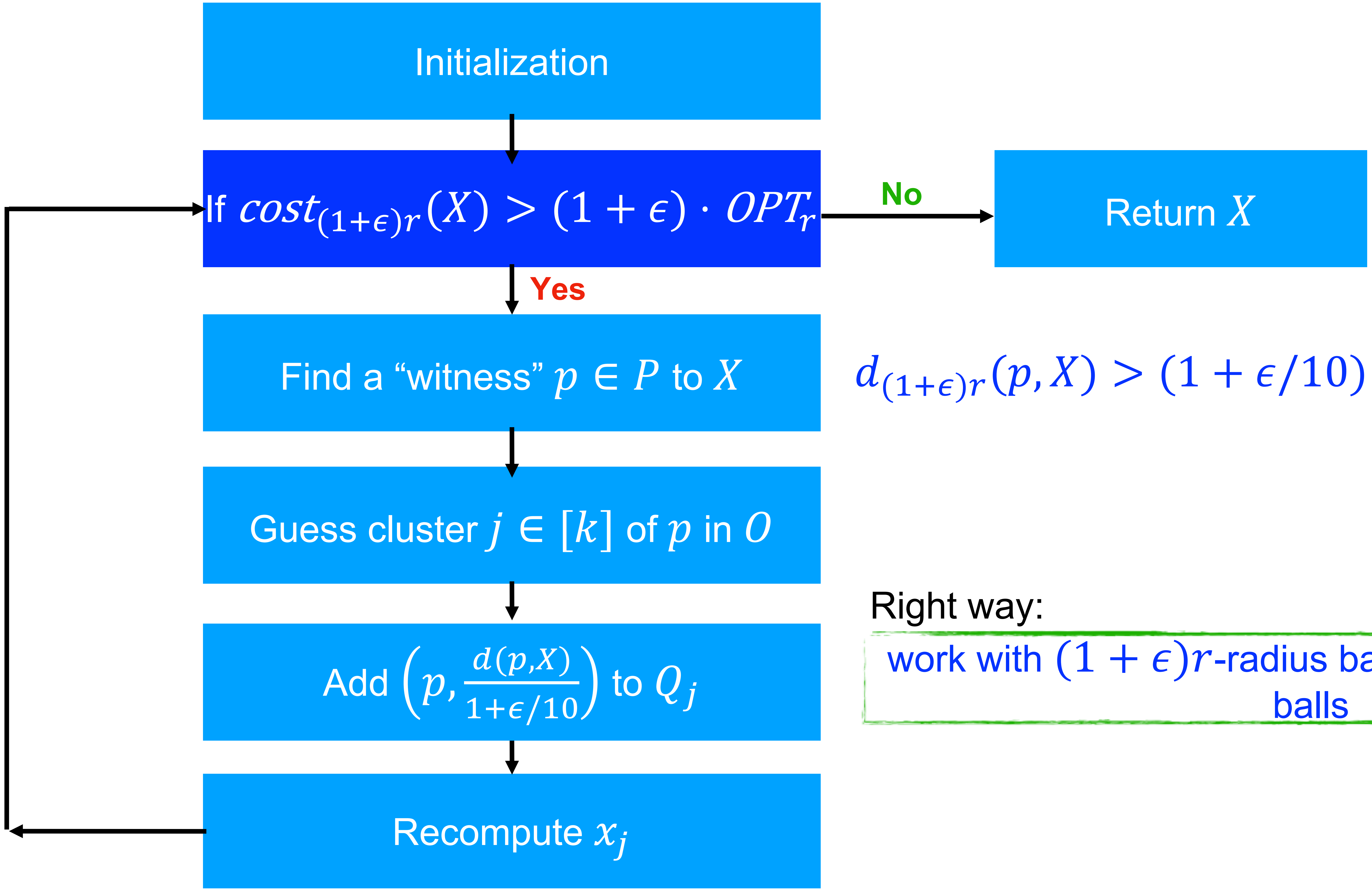


Right way:

work with $(1+\epsilon)r$-radius balls instead of r-radius balls

$\eta \to 0$

Attempt 1



Initialization

If $cost_{(1+\epsilon)r}(X) > (1+\epsilon) \cdot OPT_r$

No

Return X

Yes

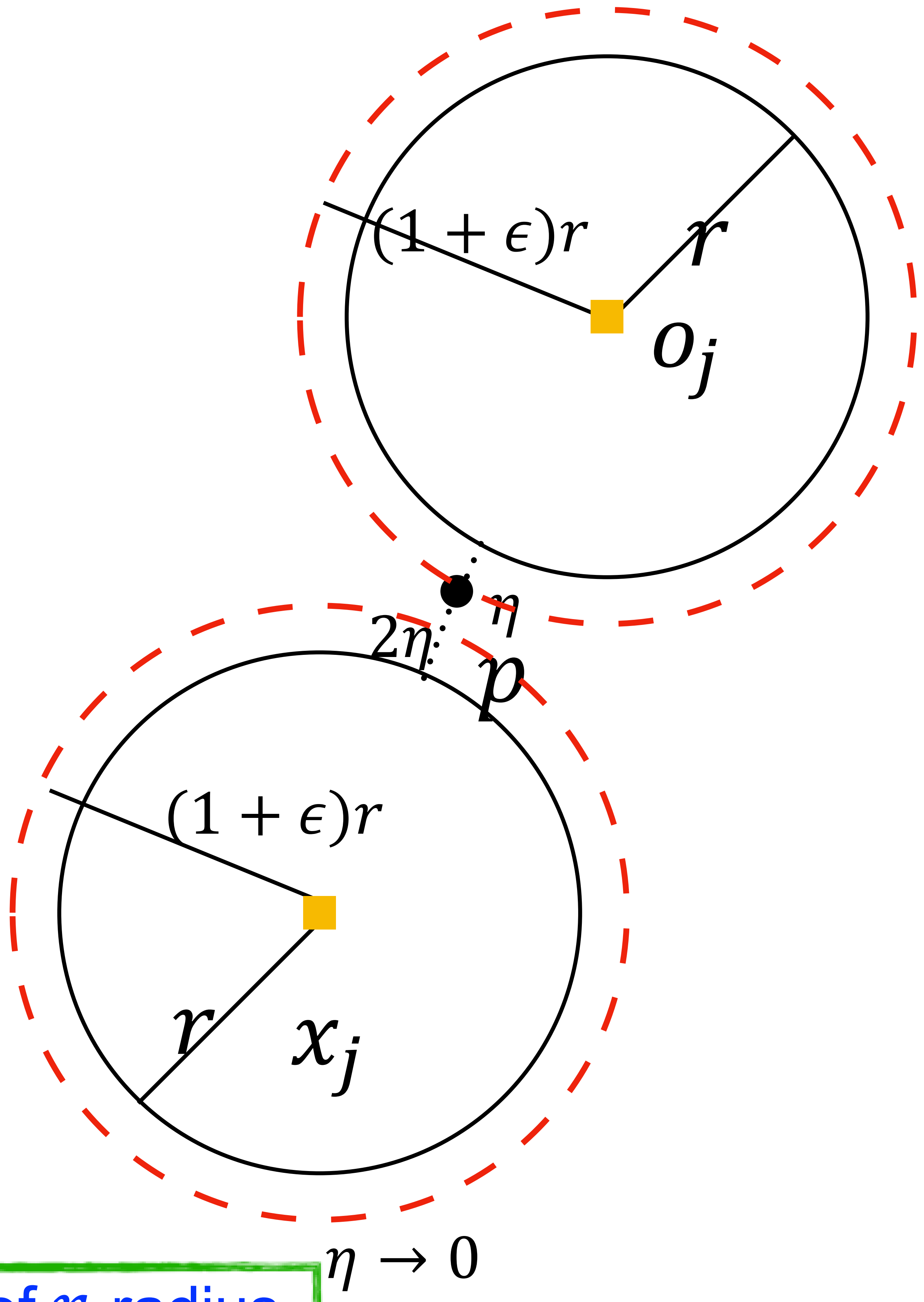
Find a “witness” $p \in P$ to X

Guess cluster $j \in [k]$ of p in O

Add $\left(p, \frac{d(p,X)}{1+\epsilon/10}\right)$ to Q_j

Recompute x_j

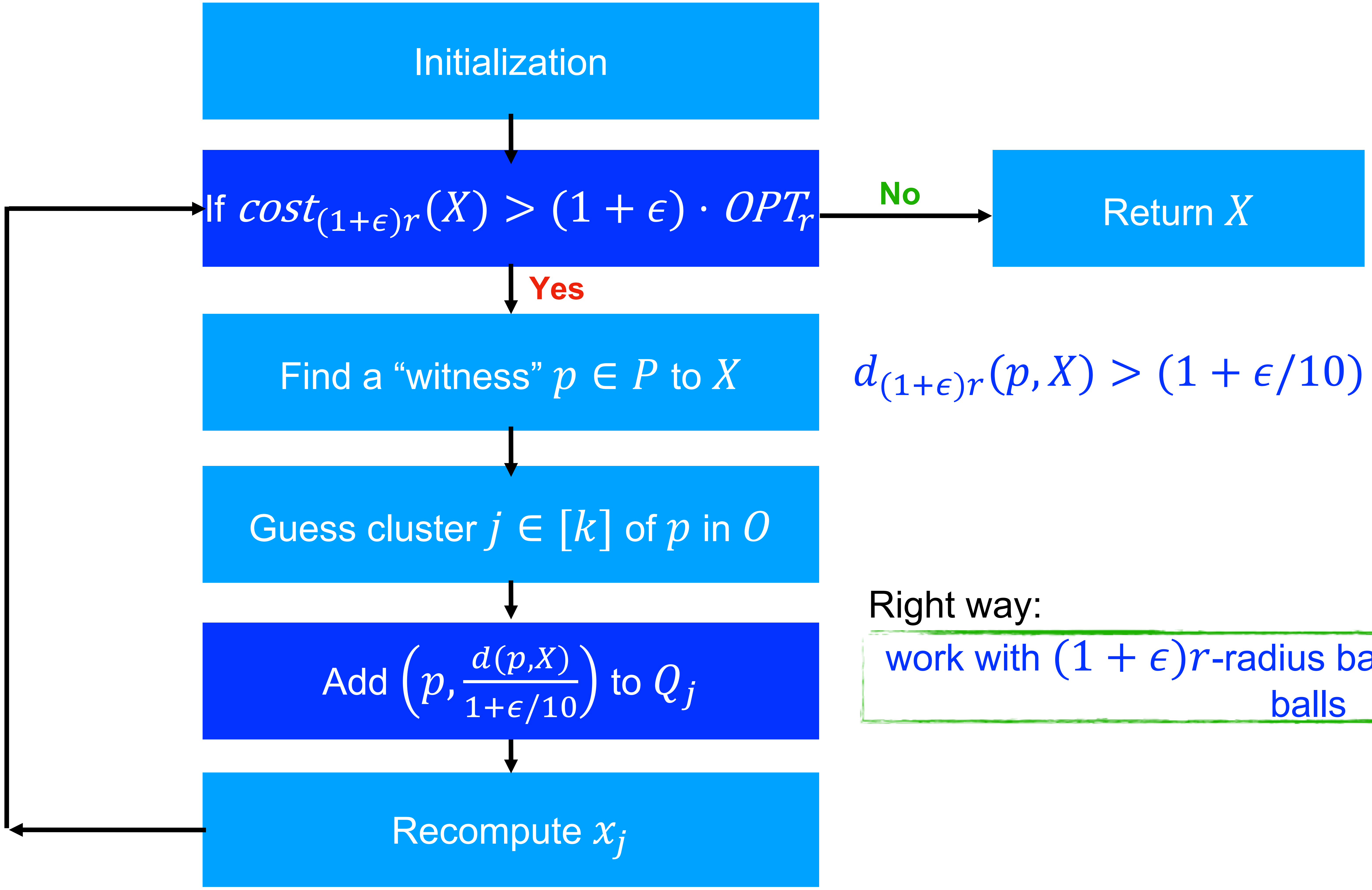
$$d_{(1+\epsilon)r}(p, X) > (1 + \epsilon/10) \cdot d_r(p, O)$$



Right way:

work with $(1+\epsilon)r$-radius balls instead of r-radius balls

Attempt 1



Initialization

If $cost_{(1+\epsilon)r}(X) > (1+\epsilon) \cdot OPT_r$

No

Return X

Yes

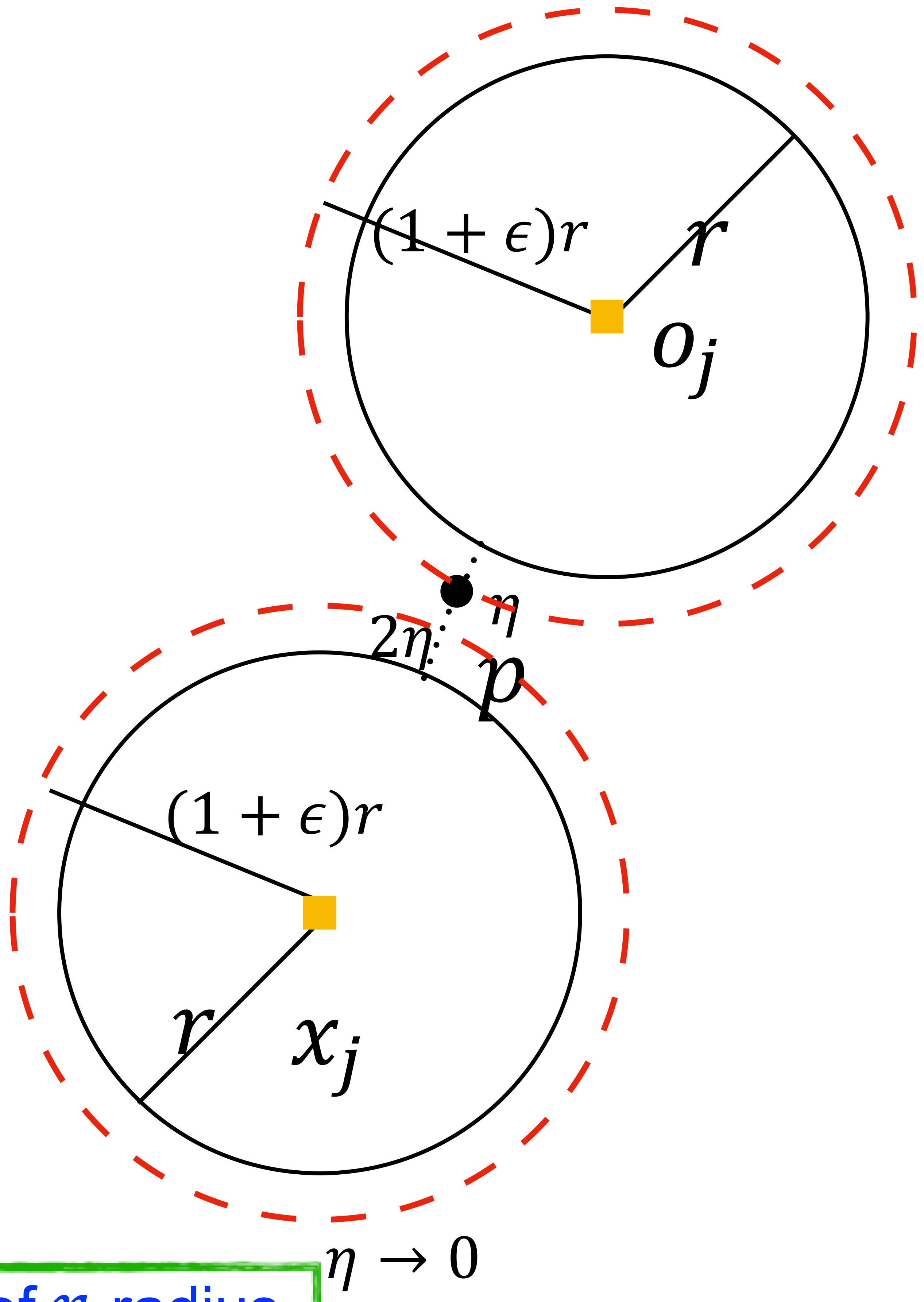
Find a “witness” $p \in P$ to X

Guess cluster $j \in [k]$ of p in O

Add $\left(p, \frac{d(p,X)}{1+\epsilon/10}\right)$ to Q_j

Recompute x_j

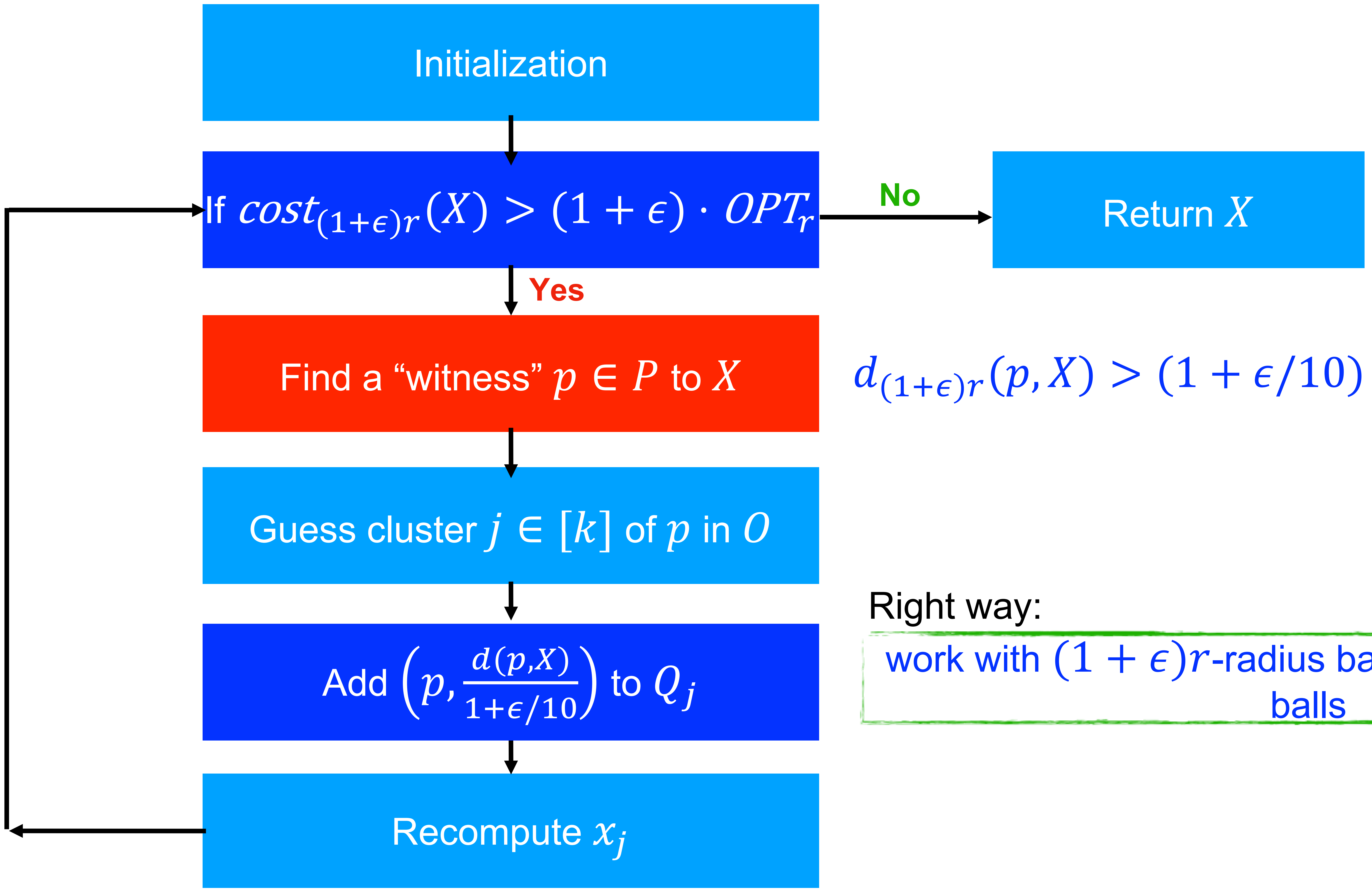
$d_{(1+\epsilon)r}(p, X) > (1 + \epsilon/10) \cdot d_r(p, O)$



Right way:

work with $(1+\epsilon)r$-radius balls instead of r-radius balls

Attempt 1



Initialization

If $cost_{(1+\epsilon)r}(X) > (1+\epsilon)\cdot OPT_r$

No

Return X

Yes

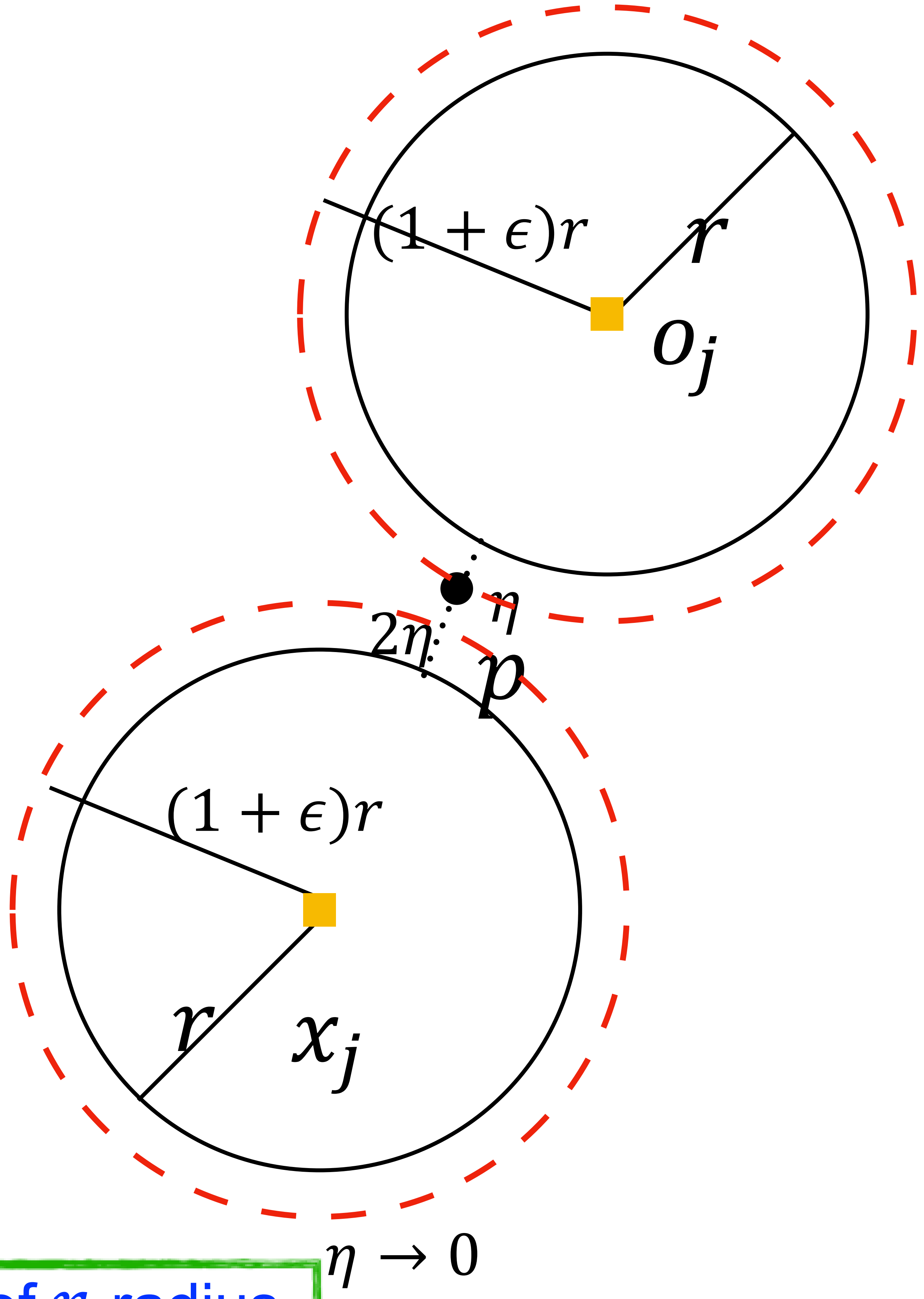
Find a “witness” $p \in P$ to X

Guess cluster $j \in [k]$ of p in O

Add $\left(p, \frac{d(p,X)}{1+\epsilon/10}\right)$ to Q_j

Recompute x_j

$d_{(1+\epsilon)r}(p,X) > (1+\epsilon/10)\cdot d_r(p,O)$



Right way:

work with $(1+\epsilon)r$-radius balls instead of r-radius balls

Sampling Witness

Witness: $d_{(1+\epsilon)r}(p, X) > (1+\epsilon/10) \cdot d_r(p, O)$

d_r does not satisfy triangle inequality $\Longrightarrow$ FOCS’23 sampling fails

Sampling Witness

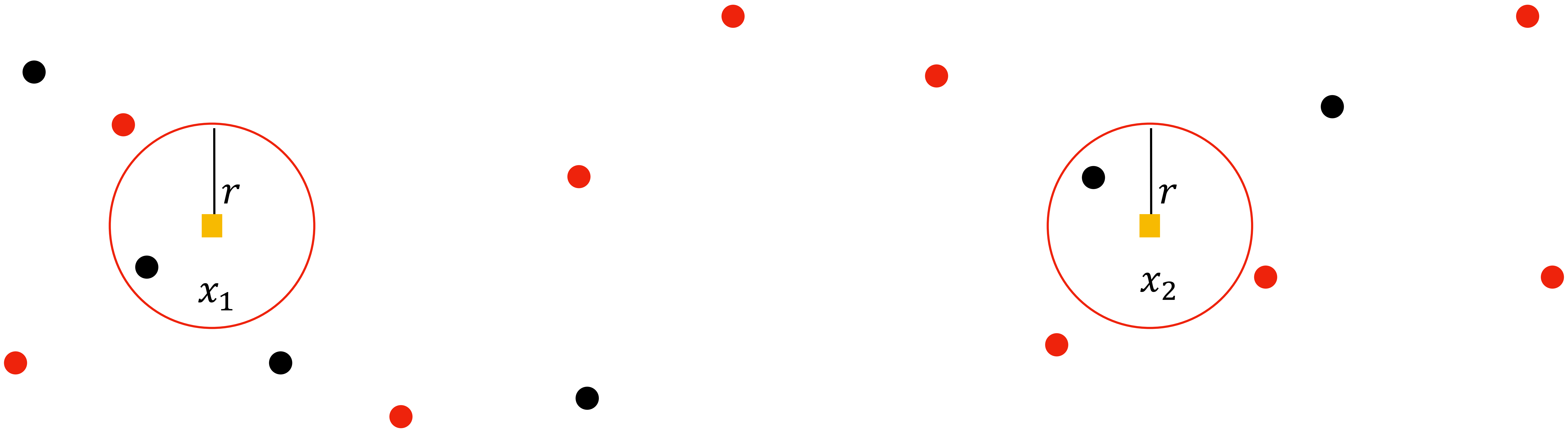
Witness: $d_{(1+\epsilon)r}(p, X) > (1 + \epsilon/10) \cdot d_r(p, O)$

d_r does not satisfy triangle inequality. But, $d_r \approx d$ when $d_r = \Omega(r/\epsilon)$

Sampling Witness

Witness: $d_{(1+\epsilon)r}(p, X) > (1 + \epsilon/10) \cdot d_r(p, O)$

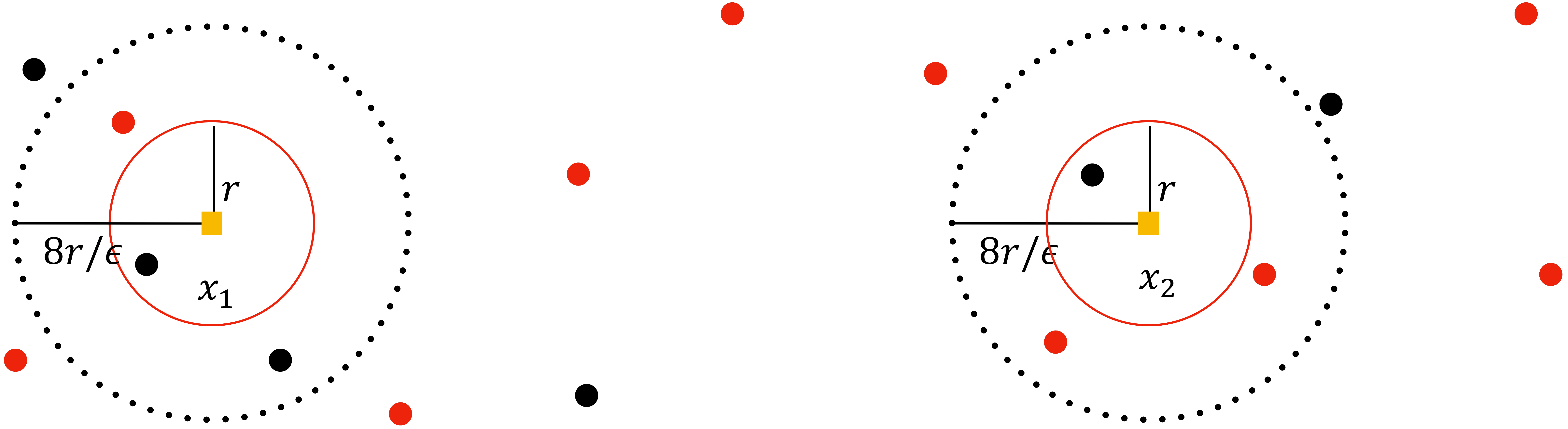
d_r does not satisfy triangle inequality. But, $d_r \approx d$ when $d_r = \Omega(r/\epsilon)$



Sampling Witness

Witness: $d_{(1+\epsilon)r}(p, X) > (1 + \epsilon/10) \cdot d_r(p, O)$

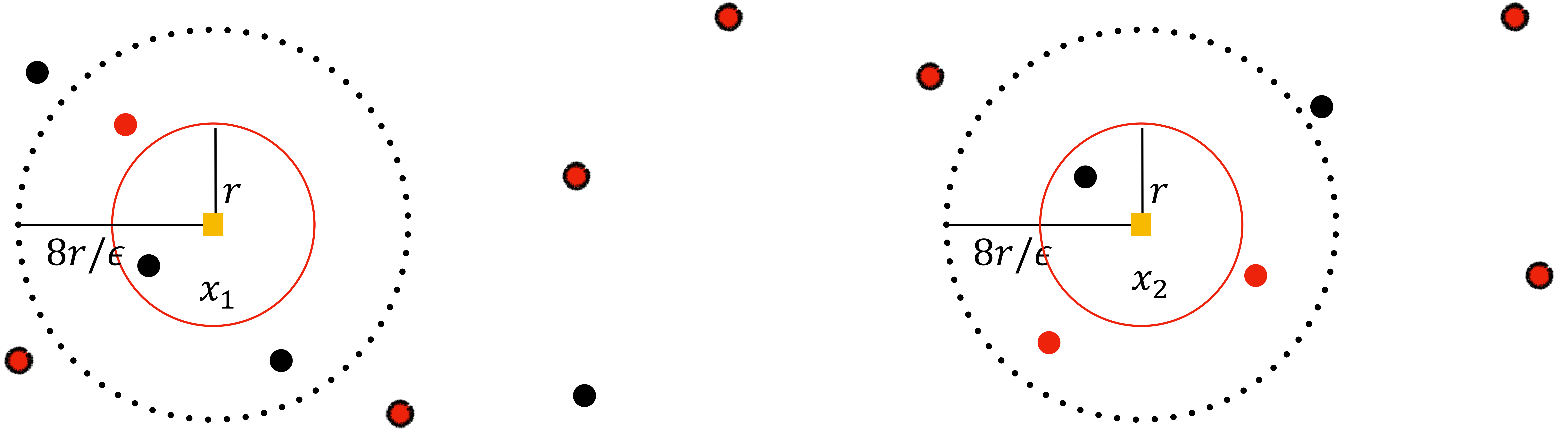
d_r does not satisfy triangle inequality. But, $d_r \approx d$ when $d_r = \Omega(r/\epsilon)$



Sampling Witness

Witness: $d_{(1+\epsilon)r}(p, X) > (1 + \epsilon/10) \cdot d_r(p, O)$

d_r does not satisfy triangle inequality. But, $d_r \approx d$ when $d_r = \Omega(r/\epsilon)$

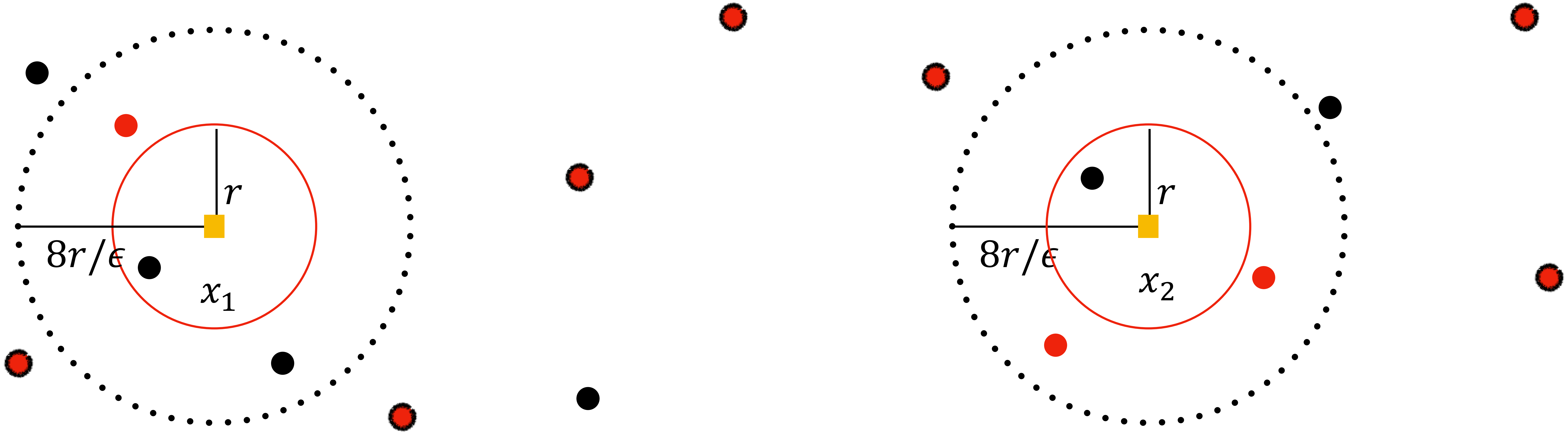


Sampling Witness

Witness: $d_{(1+\epsilon)r}(p, X) > (1 + \epsilon/10) \cdot d_r(p, O)$

Far away witnesses

d_r does not satisfy triangle inequality. But, $d_r \approx d$ when $d_r = \Omega(r/\epsilon)$

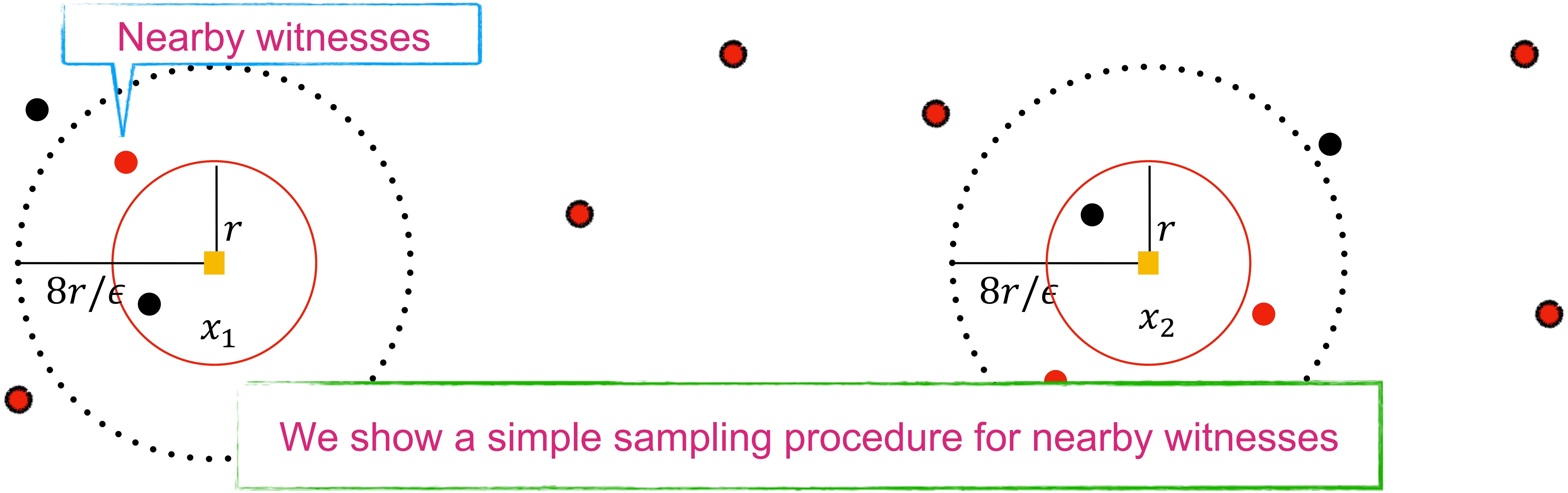


Sampling Witness

Witness: $d_{(1+\epsilon)r}(p, X) > (1+\epsilon/10) \cdot d_r(p, O)$

Far away witnesses

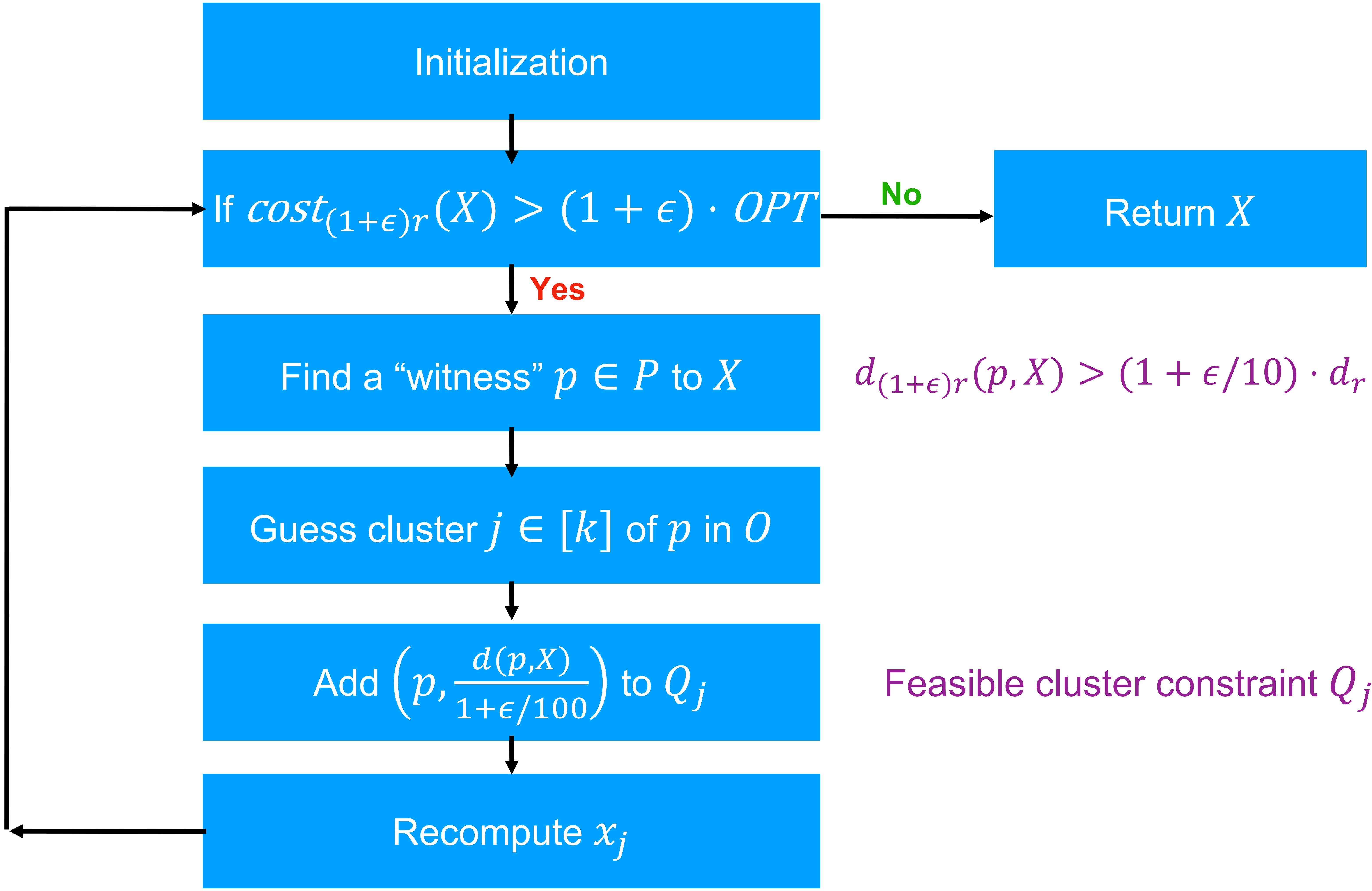
d_r does not satisfy triangle inequality. But, $d_r \approx d$ when $d_r = \Omega(r/\epsilon)$



Nearby witnesses

We show a simple sampling procedure for nearby witnesses

Our Algorithm



Initialization

If $cost_{(1+\epsilon)r}(X) > (1+\epsilon)\cdot OPT$

No

Return X

Yes

Find a “witness” $p \in P$ to X

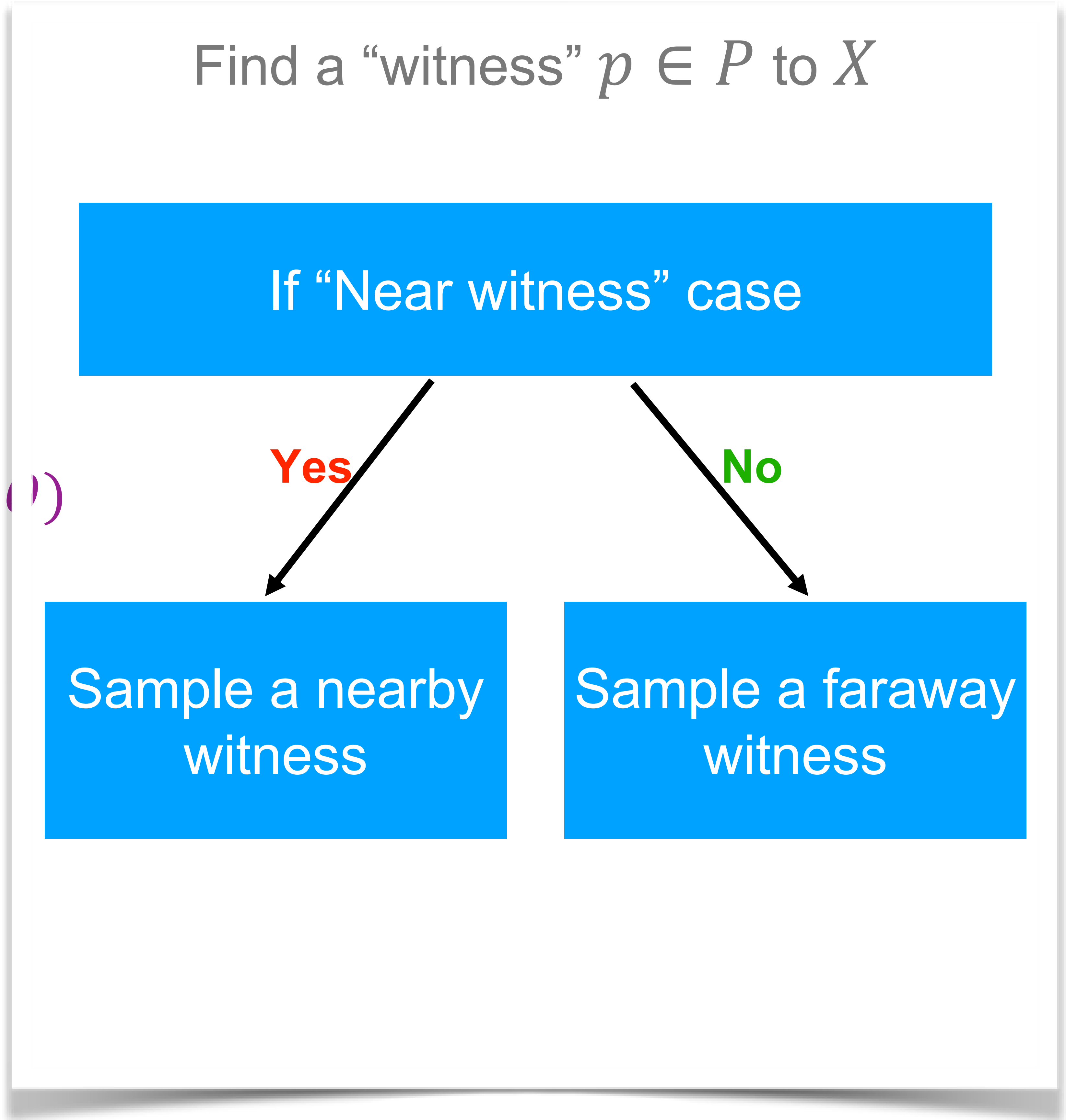
Guess cluster $j \in [k]$ of p in O

Add $\left(p, \frac{d(p,X)}{1+\epsilon/100}\right)$ to Q_j

Recompute x_j

$d_{(1+\epsilon)r}(p,X) > (1+\epsilon/10)\cdot d_r(p,\,)$

Feasible cluster constraint Q_j



Find a “witness” $p \in P$ to X

If “Near witness” case

Yes

No

Sample a nearby witness

Sample a faraway witness

Summary

Showed a bi-criteria EPAS for Hybrid Clustering

Metric spaces with bounded scatter dimension

Norm objective of r-distances

Generalize FOCS'23 EPAS framework for r-distances

Designed coresets for Hybrid Clustering in doubling dimensions

Derandomization?

Constrained variants of Hybrid Clustering?

capacities, outlier, fairness

Polynomial-time approximability?

$(18,6)$ is known

Thank You!