Matrix Factorization at the Frontier of Non-convex Optimizations

Sewoong Oh

University of Illinois at Urbana Champaign
Industrial and Enterprise Systems Engineering Department

COLUMBIA UNIVERSITY, NEW YORK
JUNE 14-18, 2010

- Home
- Call for Papers
- Paper Submission
- Organizers
- Committee
- Program
- Workshops
- Registration
- Tutorials
- Hotel
- Venue

Welcome!

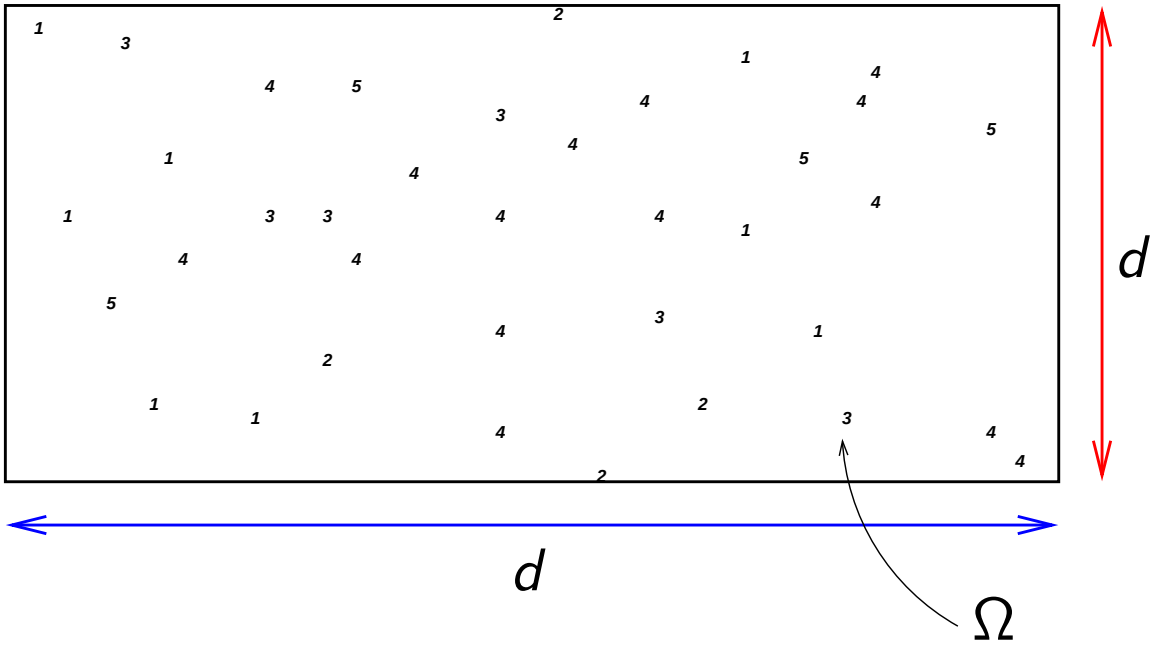
The organizing committee is excited to invite you to take part in ACM SIGMETRICS 2010. SIGMETRICS is the flagship conference of the ACM special interest group for the computer systems performance evaluation community.

The ACM SIGMETRICS conference solicits papers on the development and application of state-of-the-art, broadly applicable analytic, simulation, and measurement-based performance evaluation techniques. Of particular interest is work that furthers the state-of-the-art in performance evaluation methods, or work that creatively apply previously developed methods to understand or gain important insights into key design trade-offs in computer or network systems.

Announcements

- Congratulations to the winners of the **best paper award** given to **Load Balancing via Randomized Local Search in Closed and Open Systems** by A. Ganesh (Bristol University), S. Lilienthal (Cambridge University), D. Manjunath (IIT Mumbai), A. Proutiere (Microsoft Research), F. Simatos (INRIA)
- Congratulations to Amin Karbasi (EPFL) and Sewoong Oh (Stanford) for **best student paper award** with their paper **Distributed Sensor Network Localization from Local Connectivity: Performance Analysis for the HOP-TERRAIN Algorithm**
- Congratulations to the authors of the poster that won the **student research competition** with the poster **ASTUTE: Detecting a Different Class of Traffic Anomalies** by Fernando Silveira (Thomson and UPMC), Christophe Diot (Thomson), Nina Taft (Intel Labs Berkeley), Ramesh Govindan (University of Southern California)
- Congratulations to the authors of the the poster that won **honorable mention for best poster** with **Deep Diving into BitTorrent Locality** by Ruben Cuevas (Univ. Carlos III de Madrid),

Matrix completion



Suppose each entry is sampled i.i.d. with probability p,
how large should $|\Omega|$ be in order to recover the rank-r matrix M?

$$|\Omega| \quad \geq \quad 2\, r\, d$$

Convex relaxation

$$\underset{X \in \mathbb{R}^{d \times d}}{\text{minimize}} \qquad \sum_{(i,j) \in \Omega} (M_{ij} - X_{ij})^2 + \lambda \, \mathrm{rank}(X)$$

Convex relaxation

$$\underset{X \in \mathbb{R}^{d \times d}}{\text{minimize}} \qquad \sum_{(i,j) \in \Omega} (M_{ij} - X_{ij})^2 + \lambda \, \|X\|_*$$

(informal) Theorem

[CandesRecht08,CandesTao09,Gross10,Recht10,NegahbanWainwright11]

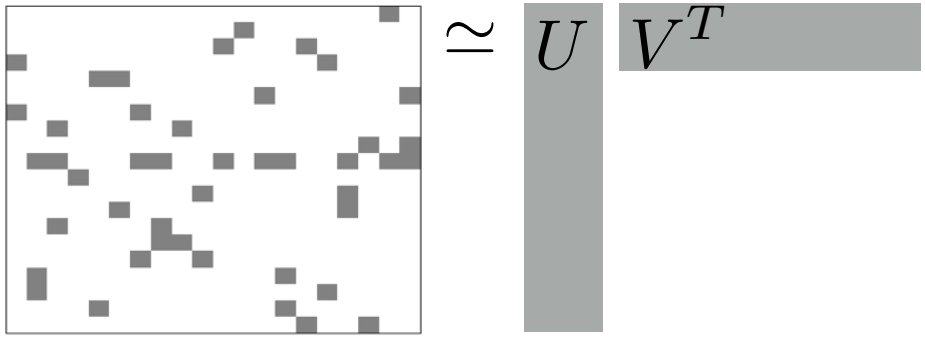
Exact recovery with high probability if and only if

$$|\Omega| \quad \geq \quad C \, r \, d \, \log d \, .$$

- SemiDefinite Program (SDP)
  - requires memory d^2
  - requires d^3 computation per iteration
  - low-rank solution is not guaranteed

Non-convex optimization

$$\underset{U,V\in\mathbb{R}^{d\times r}}{\text{minimize}} \quad \sum_{(i,j)\in\Omega} (M_{ij} - (UV^T)_{ij})^2$$

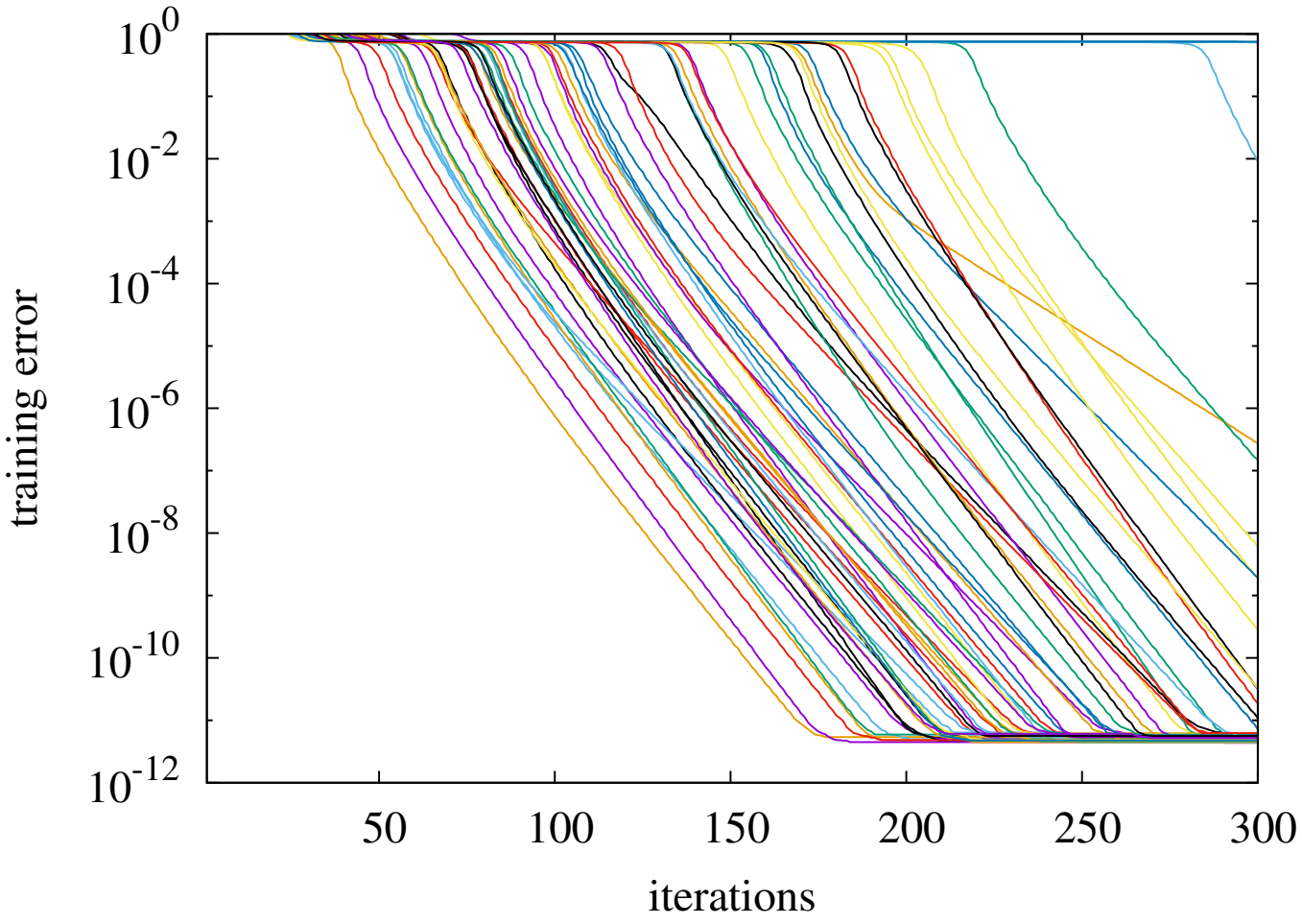


- Convex approach
  - d^2 memory
  - d^3 computation per iteration
  - low-rank not guaranteed
- Non-convex approach
  - $2rd$ memory
  - $|\Omega|$ computation per iteration
  - low-rank solution guaranteed

[BurerMoteiro03,Simon Funk 07]

Non-convex optimization

Random Initialization + Gradient Descent



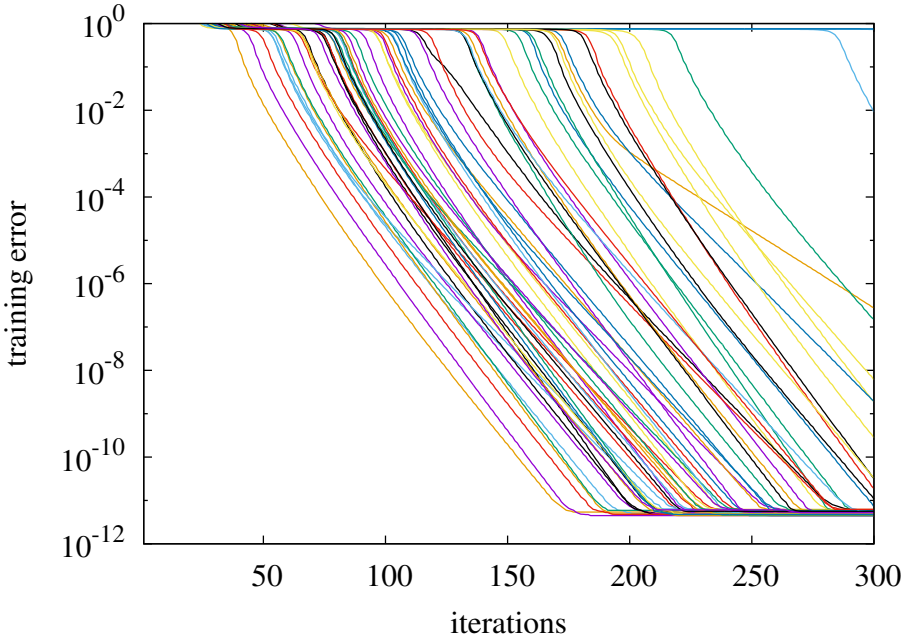
1000×1000 rank 10 matrix with 8% of entries sampled

Jellyfish [RechtRé13]

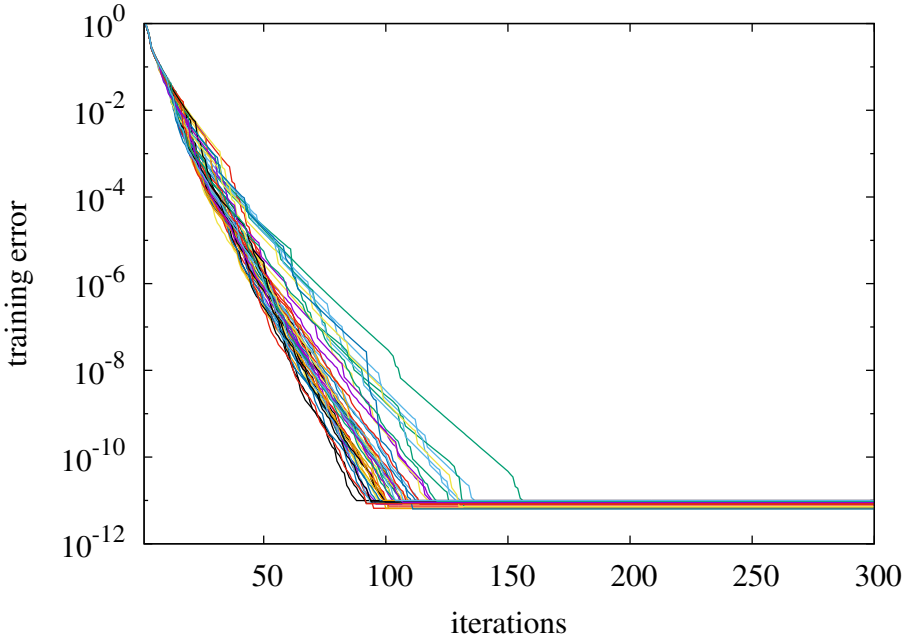
OptSpace [KeshavanMontanariOh09]

2-phase approach: Spectral Initialization + Gradient Descent

Random Initialization + GD

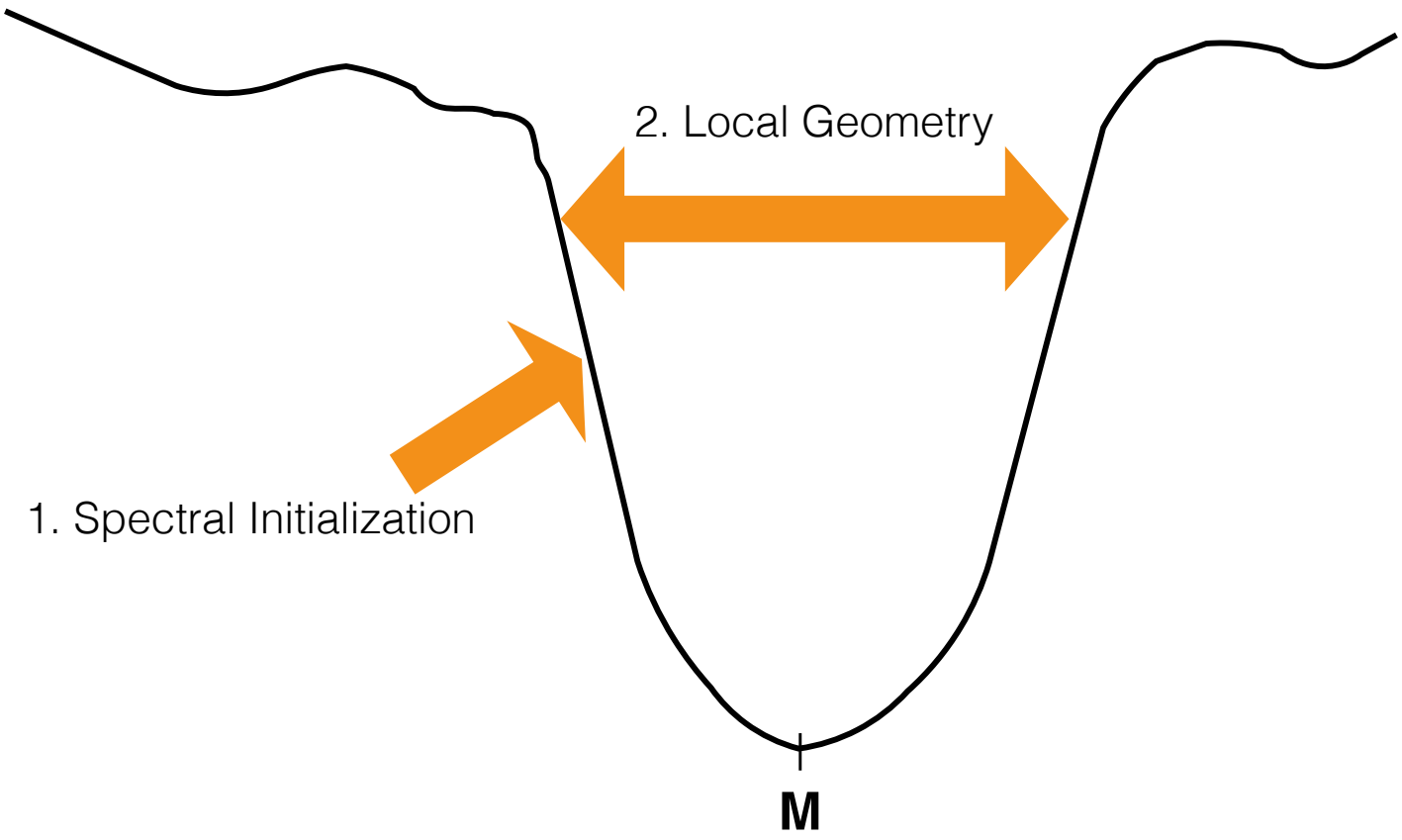


Spectral Initialization + GD

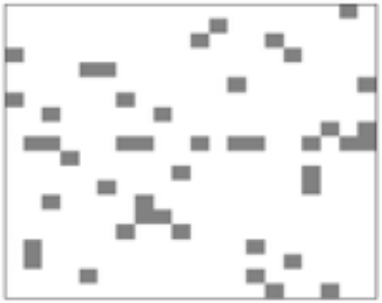


[FriedmanKahnSzemeredi89]

Conjectures for OptSpace

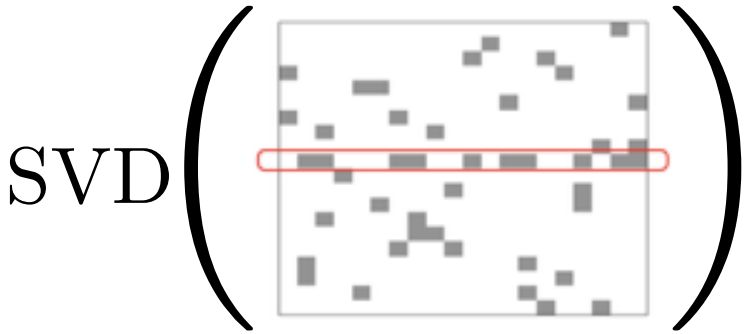


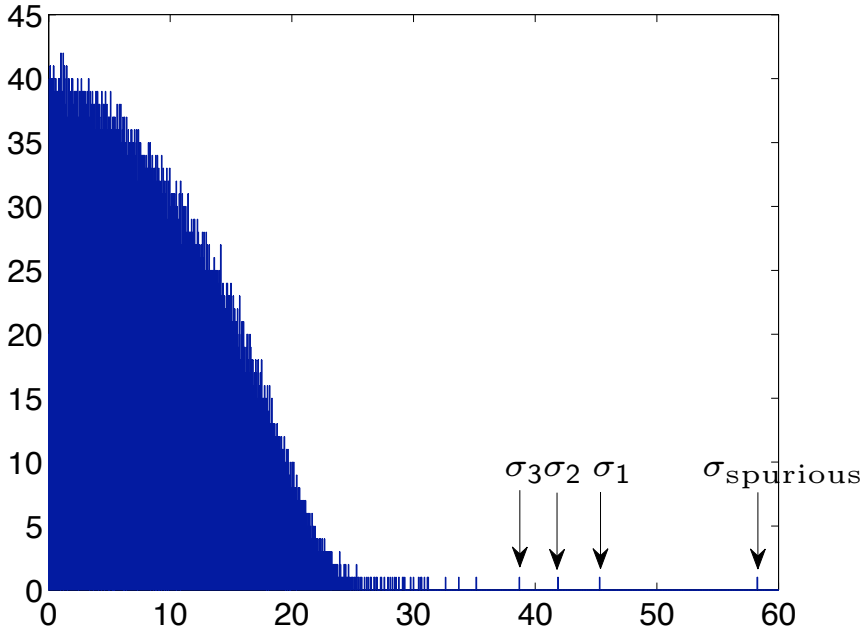
Spectral Initialization

$\mathrm{SVD}\Bigg($  $\Bigg)$

[FriedmanKahnSzemeredi89]

Spectral Initialization

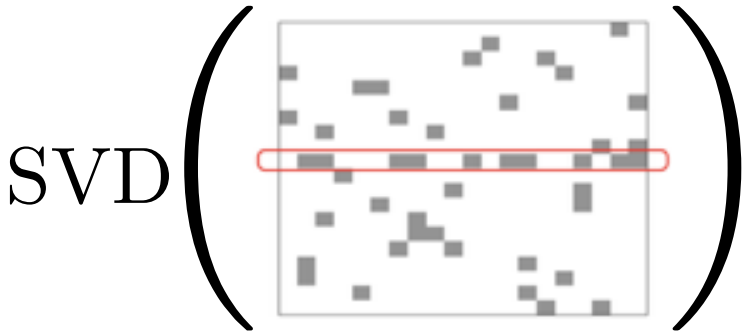


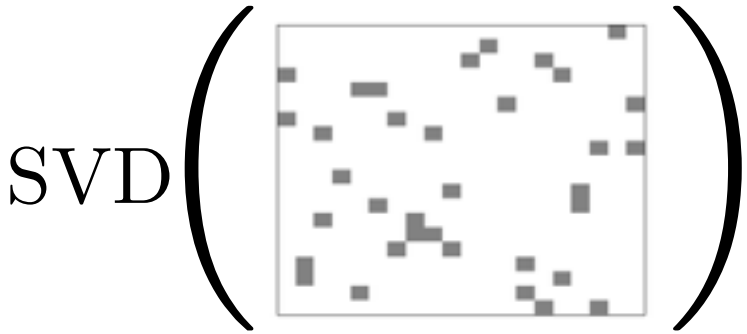


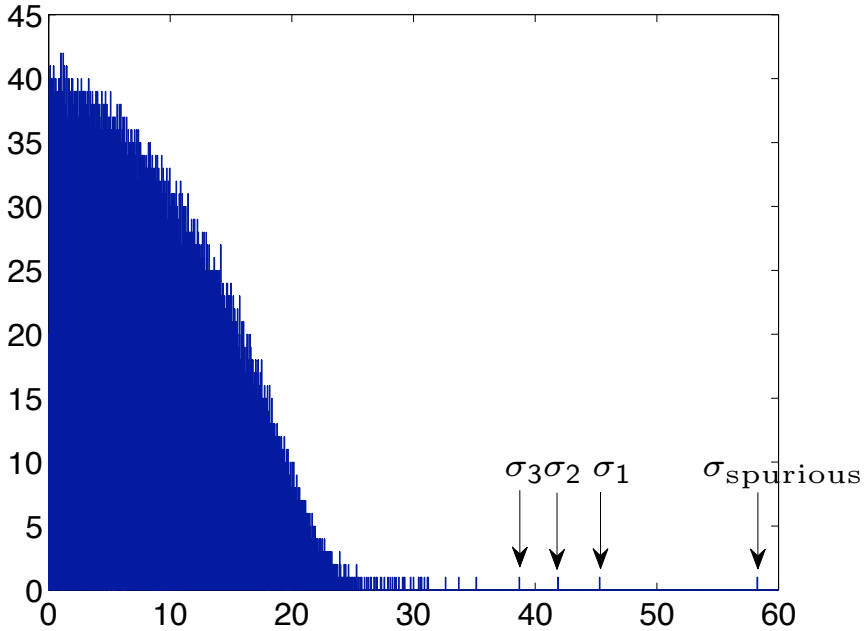
$$\sigma_{\text{spurious}} \asymp \frac{\log d}{\log \log d}$$

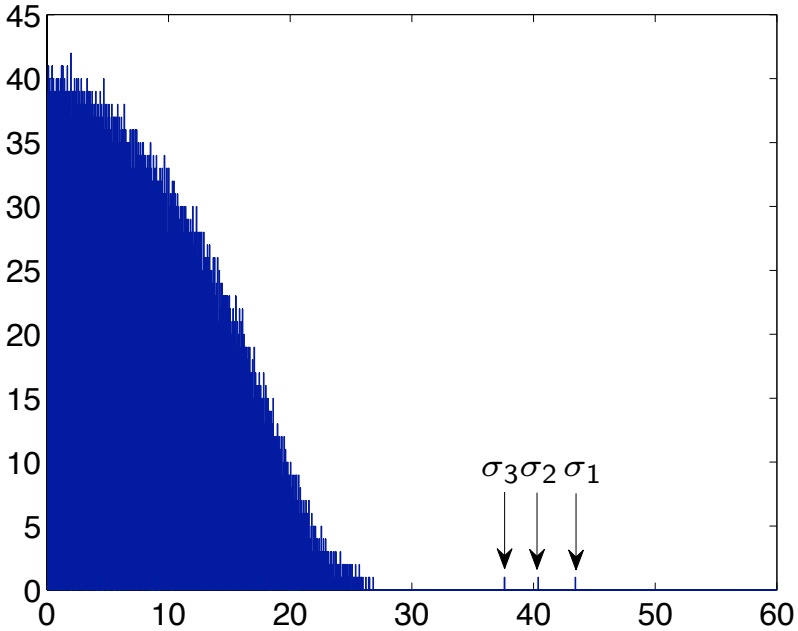
[FriedmanKahnSzemeredi89]

Spectral Initialization





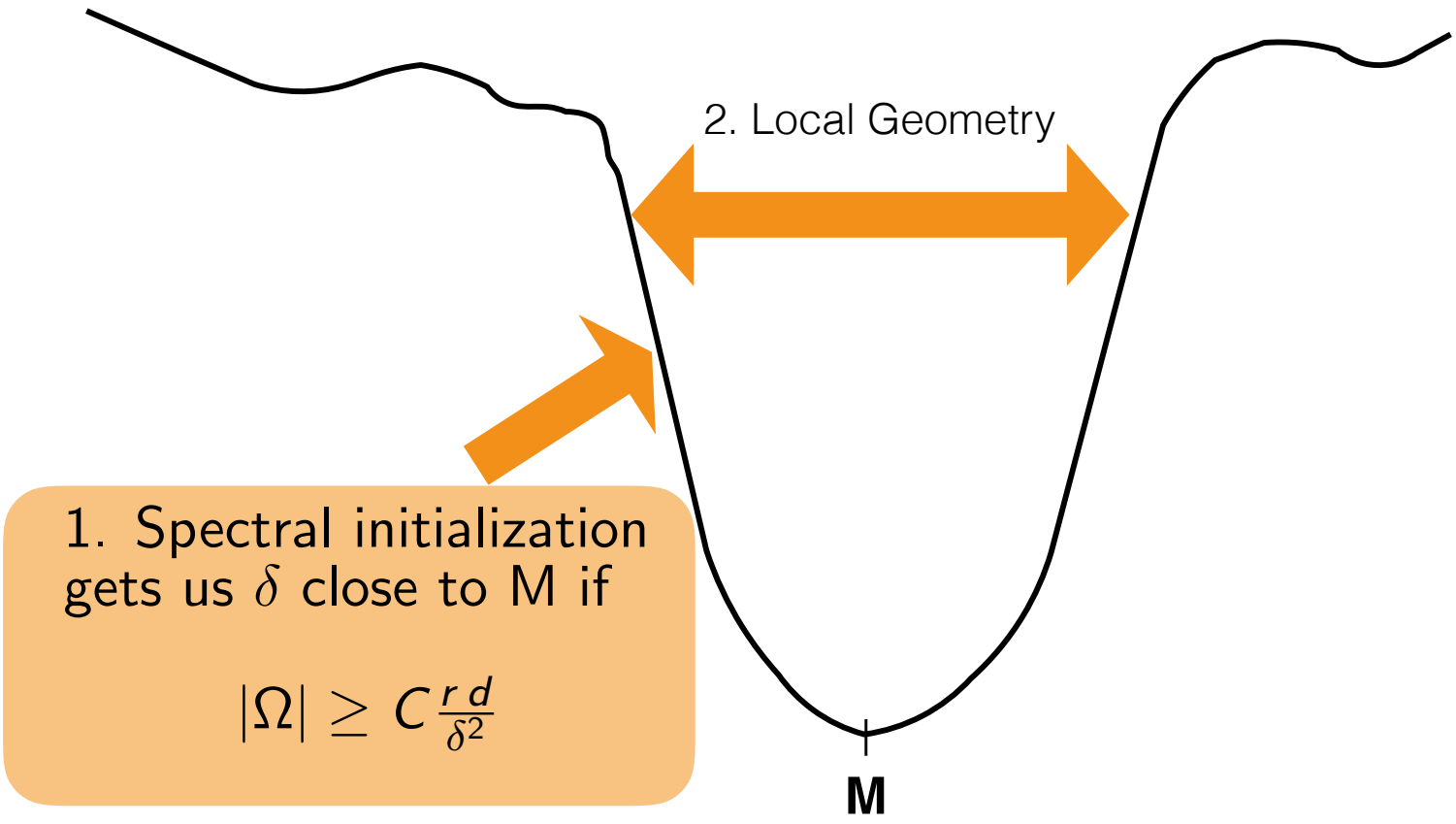




$$\sigma_{\text{spurious}} \asymp \frac{\log d}{\log \log d}$$

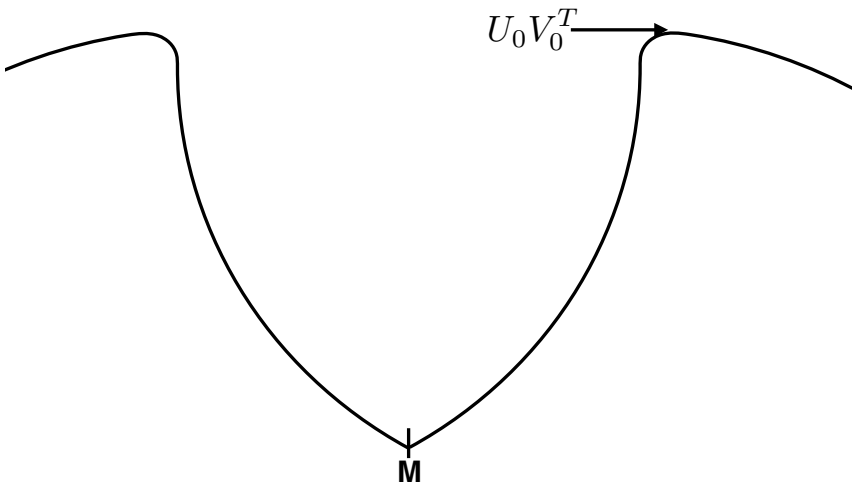
[FriedmanKahnSzemeredi89]

Resolve conjecture 1



$$F(U,V) = \sum_{(i,j)\in\Omega} (M_{ij} - (UV^T)_{ij})^2$$

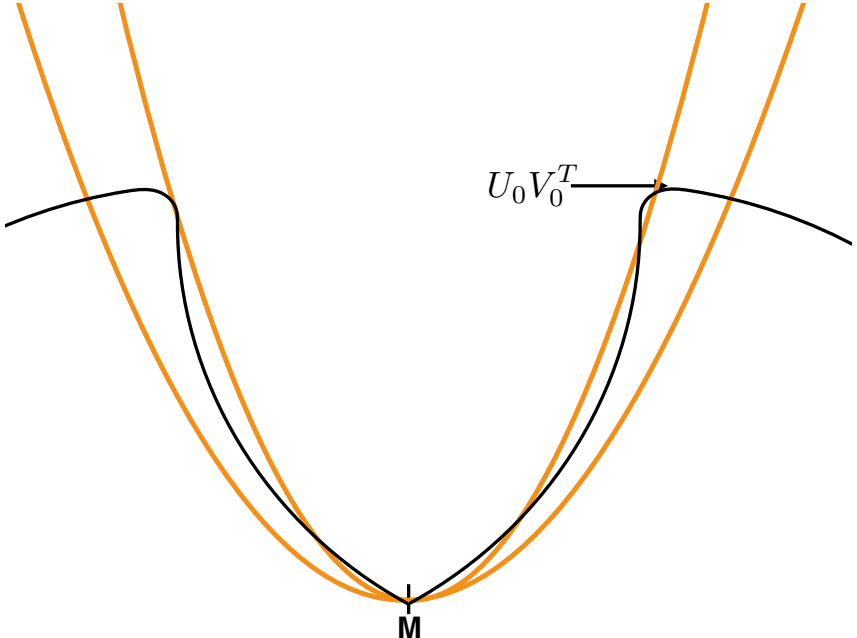
1. Gradient Descent stays inside $c\delta$-ball



$$F(U,V) = \sum_{(i,j)\in\Omega} (M_{ij} - (UV^T)_{ij})^2$$

1. Gradient Descent stays inside $c\delta$-ball

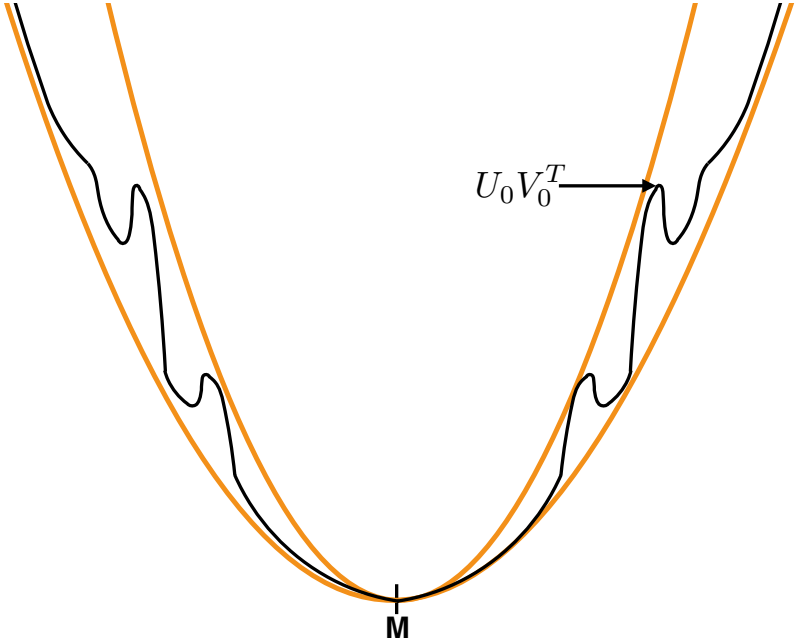
$$C_1 \|M - UV^T\|_F^2 \;\; \geq \;\; F(U,V) \;\; \geq \;\; C_2 \|M - UV^T\|_F^2$$



$$F(U,V) = \sum_{(i,j)\in\Omega} (M_{ij} - (UV^T)_{ij})^2$$

1. Gradient Descent stays inside $c\delta$-ball

$$C_1 \|M - UV^T\|_F^2 \;\; \geq \;\; F(U,V) \;\; \geq \;\; C_2 \|M - UV^T\|_F^2$$



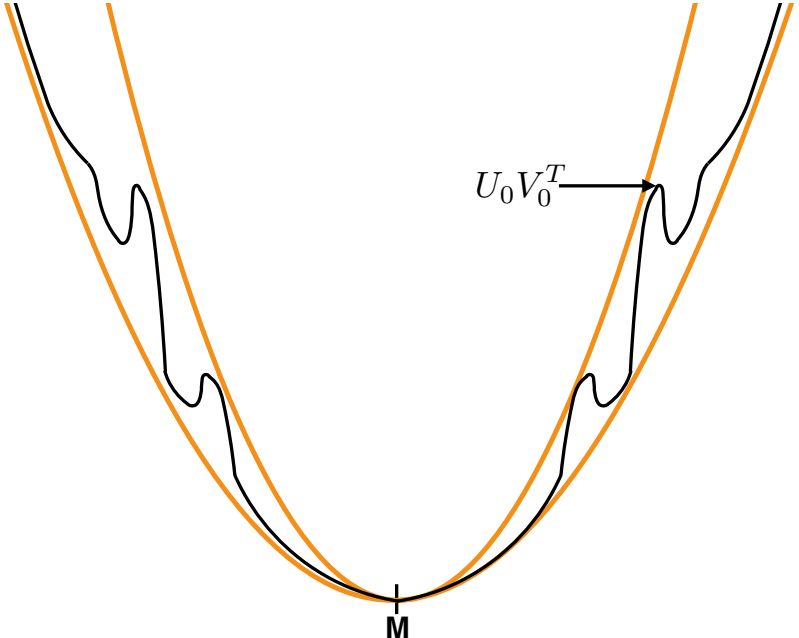
$$F(U,V) = \sum_{(i,j)\in\Omega} (M_{ij} - (UV^T)_{ij})^2$$

1. Gradient Descent stays inside $c\delta$-ball

$$C_1 \|M - UV^T\|_F^2 \;\; \geq \;\; F(U,V) \;\; \geq \;\; C_2 \|M - UV^T\|_F^2$$

2. Inside $c\delta$-ball, gradient vanishes only at global minima

$$\|\nabla F(U,V)\|_F^2 \;\; \geq \;\; C \|M - UV^T\|_F^2$$



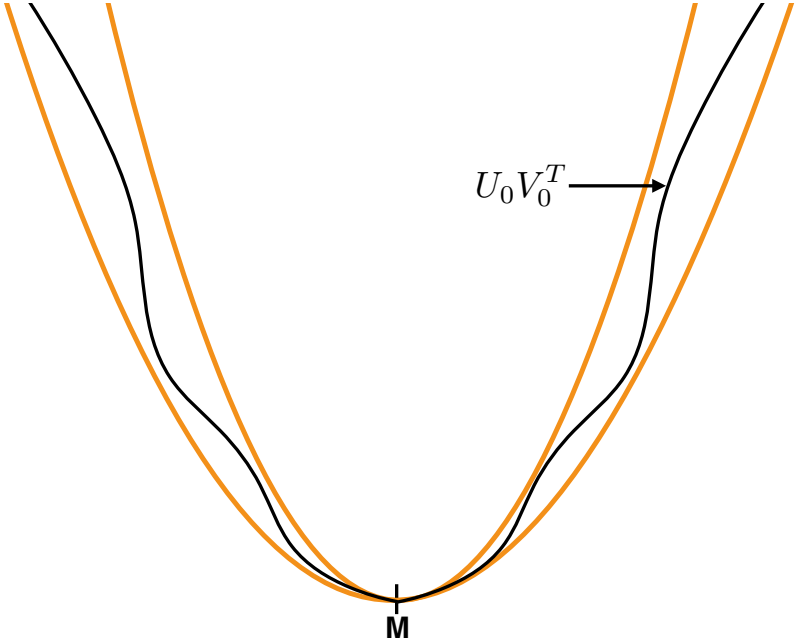
$$F(U,V) = \sum_{(i,j)\in\Omega} (M_{ij} - (UV^T)_{ij})^2$$

1. Gradient Descent stays inside $c\delta$-ball

$$C_1 \|M - UV^T\|_F^2 \;\; \geq \;\; F(U,V) \;\; \geq \;\; C_2 \|M - UV^T\|_F^2$$

2. Inside $c\delta$-ball, gradient vanishes only at global minima

$$\|\nabla F(U,V)\|_F^2 \;\; \geq \;\; C \|M - UV^T\|_F^2$$



Proof sketch

1. Gradient Descent stays inside $c\delta$-ball

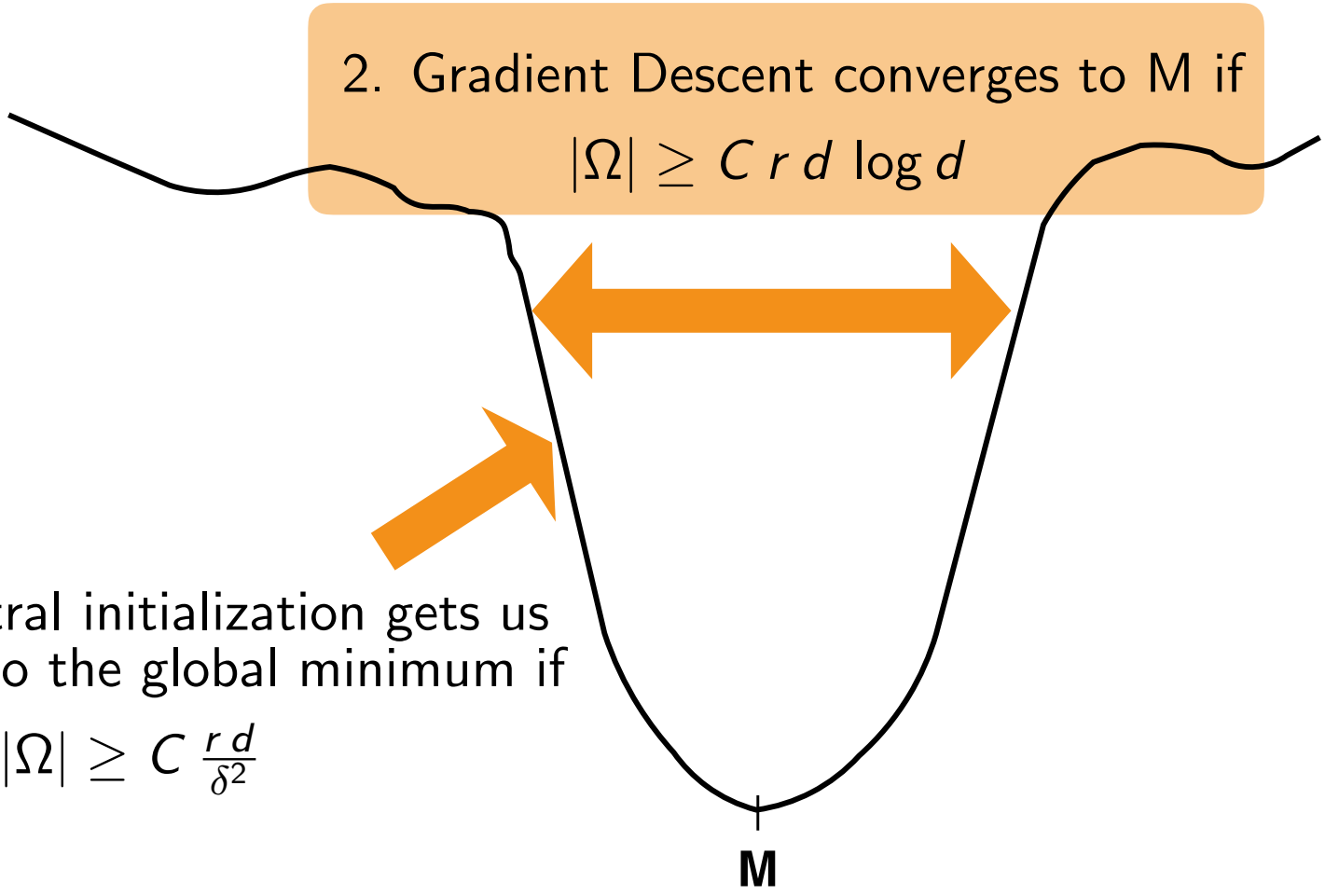
$$C_1 \|M - UV^T\|_F^2 \;\geq\; F(U,V) \;\geq\; C_2 \|M - UV^T\|_F^2$$

Lemma. [CandesRecht08]

If $|\Omega| \geq crd \log d$ then for all M and UV^T that are rank-r and incoherent,

$$\frac{1}{|\Omega|} \sum_{(i,j)\in\Omega} \left(M_{ij} - (UV^T)_{ij}\right)^2 \;\simeq\; \frac{1}{d^2}\|M - UV^T\|_F^2$$

Resolve conjecture 2

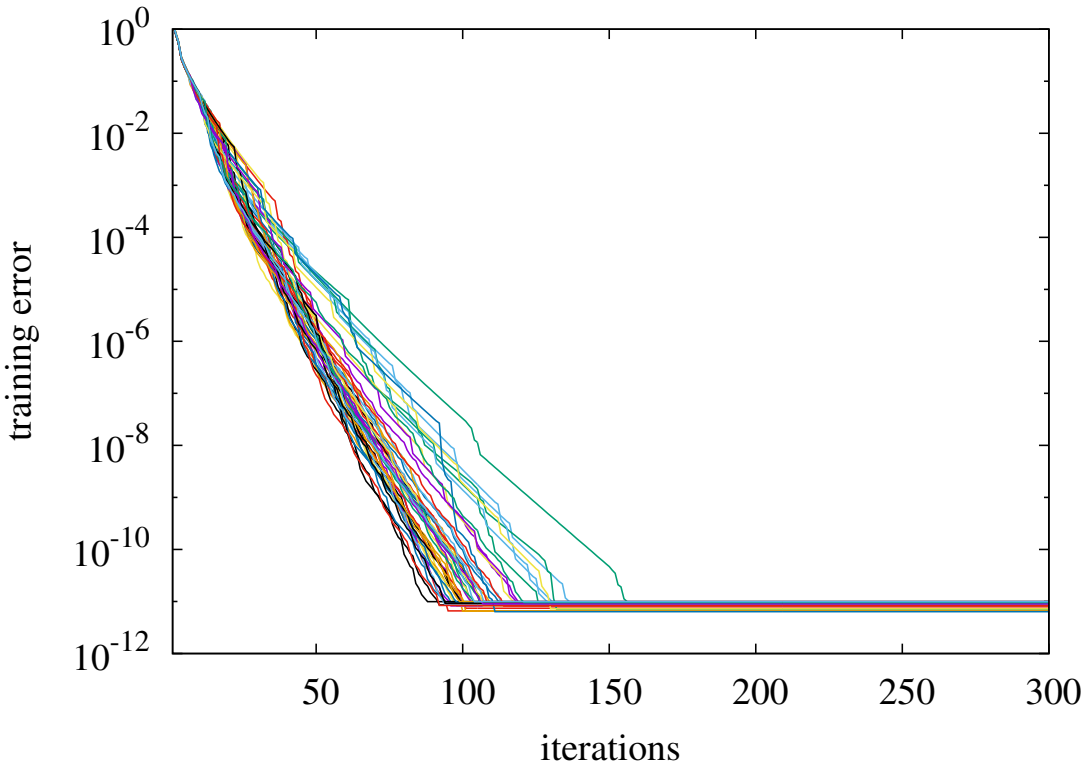


1. Spectral initialization gets us δ close to the global minimum if

$$|\Omega| \geq C\, \frac{r\, d}{\delta^2}$$

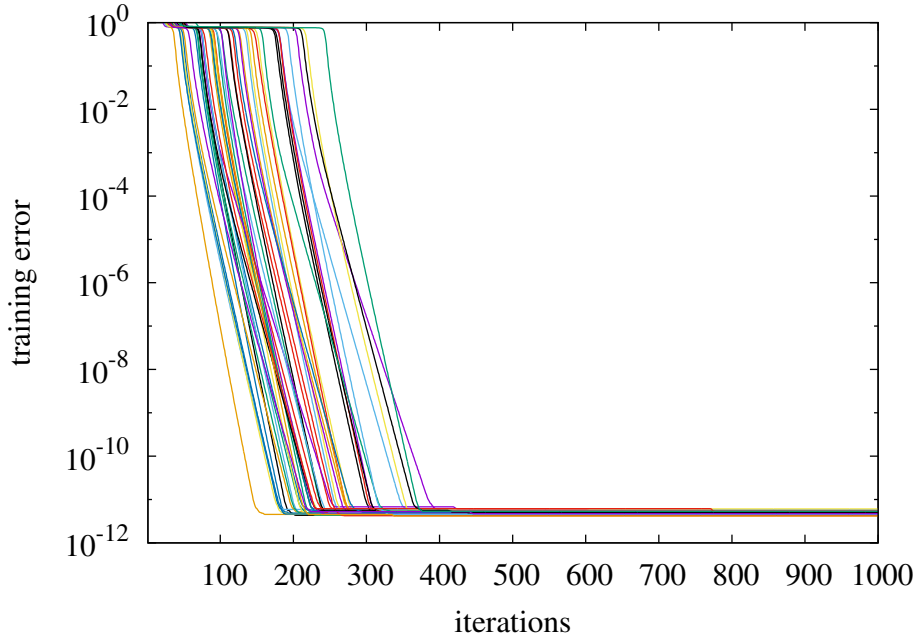
Recap: OptSpace

2-phase approach: Spectral Initialization + Gradient Descent



Followed by a productive line of research in
[JainNetrapalliSanghavi13] [Hardt14] [HardtWooters14] [JainNetrapalli14] [HastieMazumderLeeZadeh14] [SunLuo14] [ChenWainwright15] [ZhengLafferty16]

Are some non-convex optimizations as easy as convex optimizations?



- Does first-order methods converge from random initialization?
- there are no spurious local minima [GeLeeMa16,GeJinZheng17]
- saddle points can be escaped by [NesterovPolyak06,SunQuWright15,GeHuangJinYuan15, Levy16,LeeSimchowitzJordanRecht16,JinGeNetrapalliKakadeJordan17]

Standard Recipe

Matrix Completion

[CandesRecht08][CandesTao09]
[CandesPlan09][Gross11][Recht11]
[NegahbanWainwright10]
[BhojanapalliJainSanghavi15]
[SingerCucuringu08][RhodeTsybakov09]
[ChenXuCaramanisSanghavi11]
[FoygelSrebro11]

[SrebroJaaakkola03][CaiCandesShen08]
[MaGoldfardChen09]
[MazumdarHastieTibshirani09]
[TohYun09][WenYinZhang10]
[BalzanoNowakRecht10]

} **Convex relaxation**

[KeshavanMontanariOh09a,b]
[LeeBresler09][DaiMilenkovic09]
[MekaJainDhillon09][KimYedlaPfister10]
[RechtRe10][KeshavanMontanari10]
[JainNetrapalliSanghavi12][Hardt14]
[SunLuo15][ZhengLafferty16]

} **2-Phase non-convex method**

[GeLeeMa16][GeJinZheng17]
[BhojanapalliNeyshaburSrebro16]

} **Global geometry**

Matrix Completion

[CandesRecht08][SingerCucuringu08]
[RhodeTsybakov09][CandesTao09]
[CandesPlan09][NegahbanWainwright10]
[Gross11][Recht11]
[ChenXuCaramanisSanghavi11]
[FoygelSrebro11]
[BhojanapalliJainSanghavi15]

[SrebroJaaakkola03][CaiCandesShen08]
[MaGoldfardChen09]
[MazumdarHastieTibshirani09]
[TohYun09][WenYinZhang10]
[BalzanoNowakRecht10]

[KeshavanMontanariOh09a,b]
[LeeBresler09][DaiMilenkovic09]
[MekaJainDhillon09][KimYedlaPfister10]
[RechtRe10][KeshavanMontanari10]
[JainNetrapalliSanghavi12][Hardt14]
[SunLuo15][ZhengLafferty16]

[BhojanapalliNeyshaburSrebro16]
[GeLeeMa16][GeJinZheng17]

Choice Modeling

[Thurstone1927]

[LuNegahban14]
[ParkNeemanZhangSanghavi15]
[OhThekumparampilXu15]

Phase Retrieval

[Walther1963][Fienup1978][Fienup1982]

[CandesStrohmerVoroninski13]
[CandesEldarStrohmerVoroninski13a,13b]

[SchniterRangan12]
[NetrapalliJainSanghavi13]
[MrouehRosasco13][MarchesiniTuWu14]
[MukherjeeSeelamantula14]
[ShechtmanBeckEldar14][Mroueh14]
[ChenCandes15][ZhangChiLiang16]
[CandesLiSoltanolkotabi15]
[KolteOzgur16][CaiLiMa16][GauXu16]
[BendoryEldar16][SunQuWright16]

Tensor Completion

[GandyRechtYamada11]
[MuHuangWrightGoldfarb13][Romera-PeredesPontil13][LiuMusialskiWonkaYe13]
[HuangMuGoldfarbWright14][YuanZhang14]
[BhojanapalliSanghavi15][BarakMoitra15]
[ShahRaoTang15][Aswani16]
[PotechinSteurer17]

[JainOh14]
[KressnerSteinlechnerVandereycken14]
[ZhaoZhangCichocki15][SchrammWeitz15]
[Zhang16][LiuAeronAggarwalWang16]
[ZhangAeron17]

Robust PCA

[DevlinGnanadesikanKettenring1981]
[XuYuille1995]

[CandesLiMaWright09]
[WrightPengMaGaneshRao09]
[ChandrasekaranSanghaviParriloWillsky09]
[LinChenMa10][XuCaramanisSanghavi10]
[HsuKakadeZhang11]
[AgarwalNegahbanWainwright12]
[WangHongMaLuo13]
[SedghiAnandkumarJonckheere14]

[WatersSankaranarayananBaraniuk11]
[KyrillidisCevher12][FengXuYan13]
[NetrapalliNiranjanSanghaviAnandkumarJain14][YanYeShen15][ChenWainwright15]
[YiParkChenCaramanis16][GuWangLiu16]
[CherapanamjeriGuptaJain16]

Crowdsourcing

[DawidSkene79]

[SmythFayyadBurlPeronaBaldi1994]
[ShengProvostIpeirotis08]
[WauthierJordan11]**[KargerOhShah11,13]**
[HoVaughan12][ZhouBasuMaoPlatt12]
[LiuPenglhler12]
[MozafariSarkarFranklinJordan12,14]
[DalviDasguptaKumarRastogi13]
[ZhangChenZhouJordan14]
[OkOhShinYi16][KhetanOh16]
[OkOhJangShinYi17]

Dictionary Learning, Community Detection, Synchronization…

Crowdsourcing systems: When does adaptivity help?

Market to get labels for training data

Quality of the labels can be very low

Add redundancy to cope with noise

Help find missing...

Help find missing...

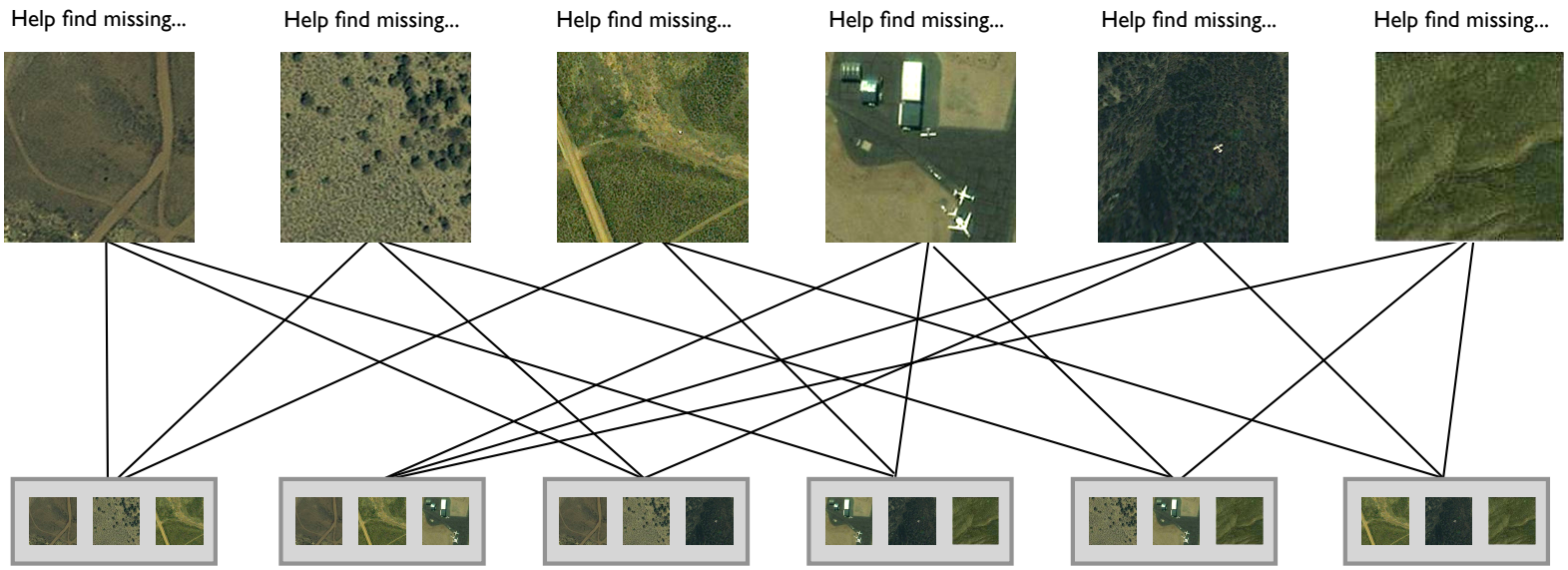
Help find missing...

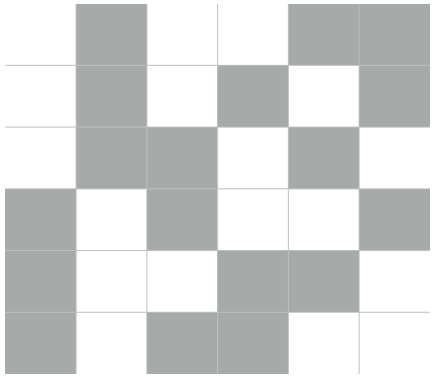
Help find missing...

Help find missing...

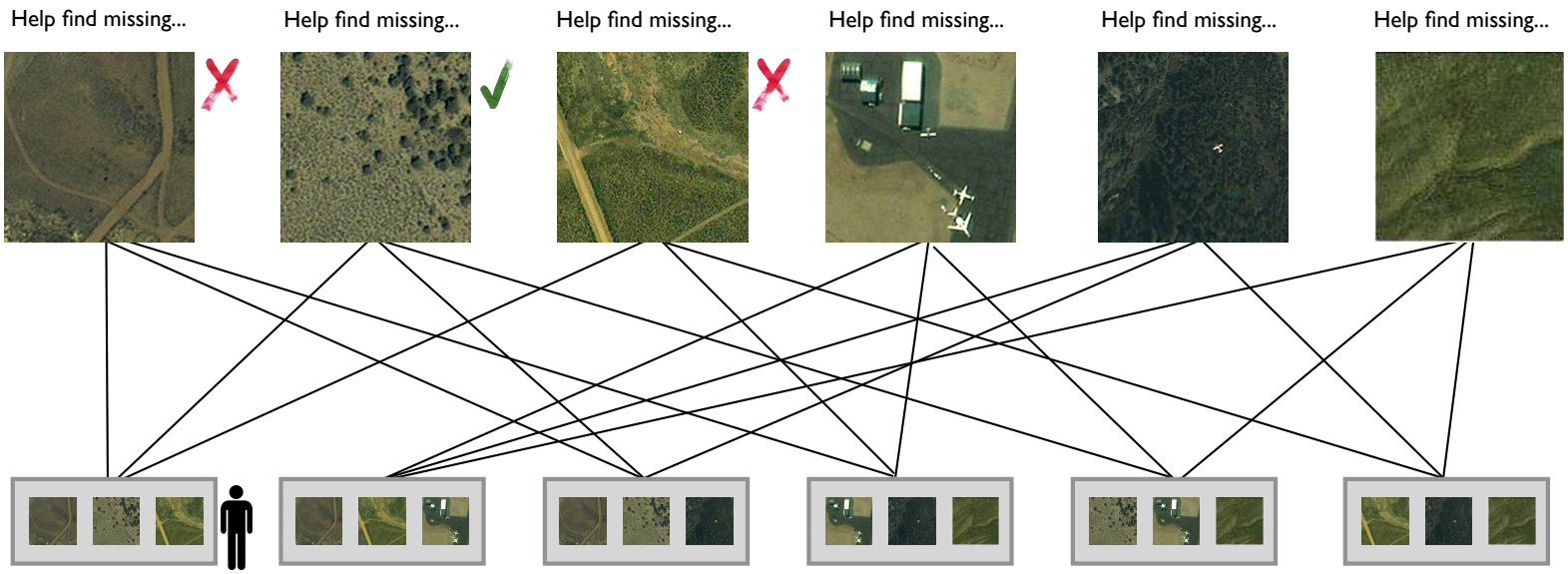
Help find missing...

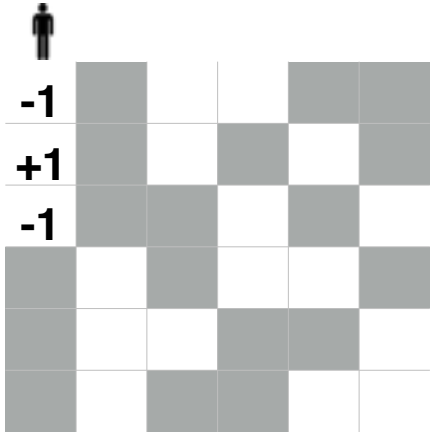
Add redundancy to cope with noise



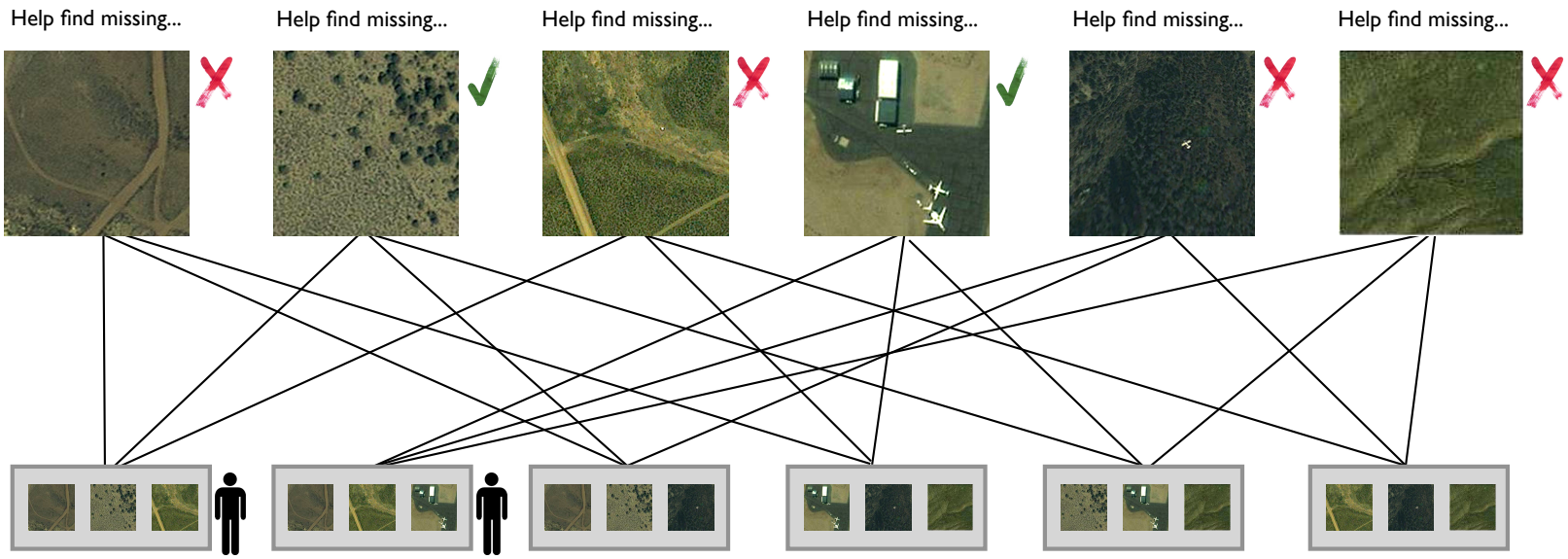


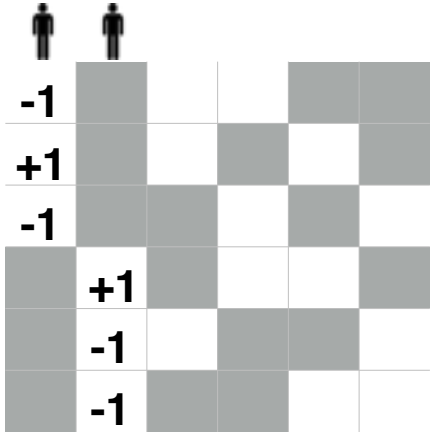
Add redundancy to cope with noise



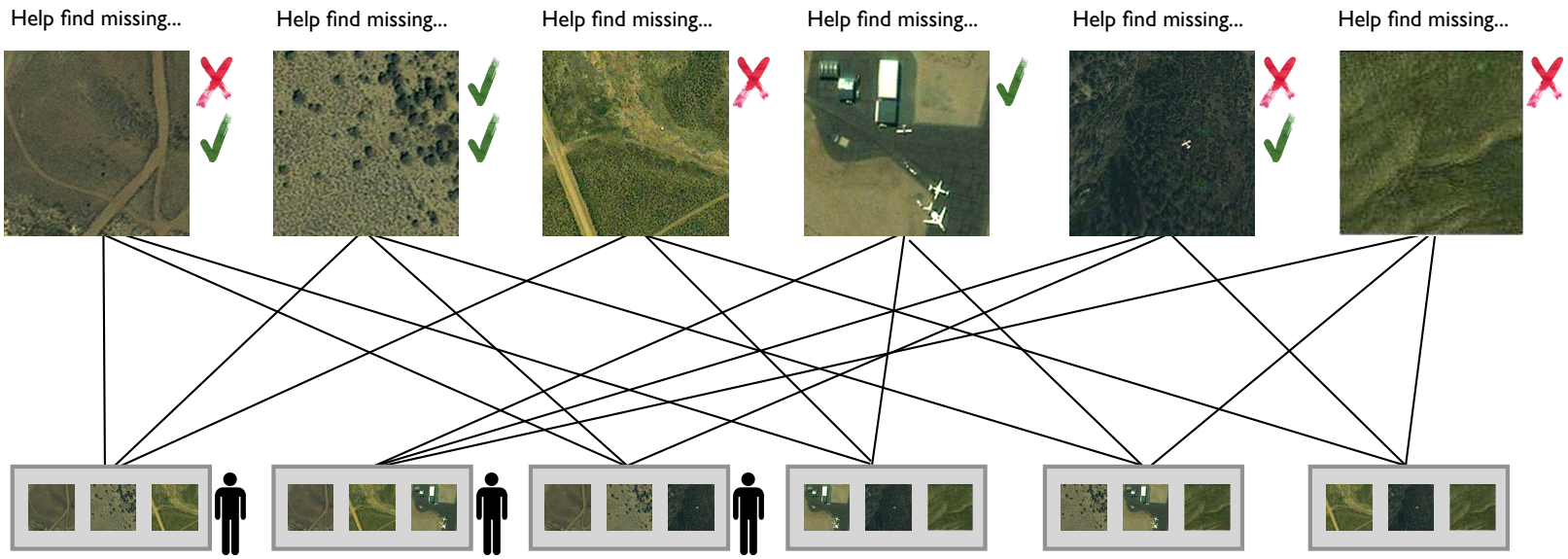


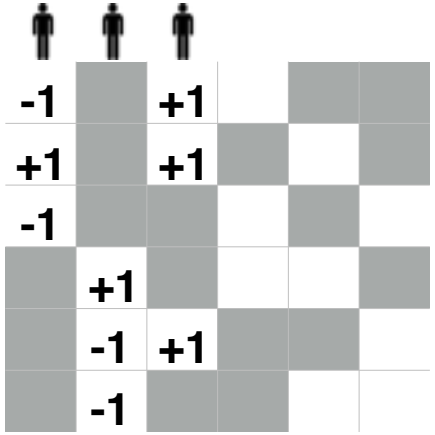
Add redundancy to cope with noise



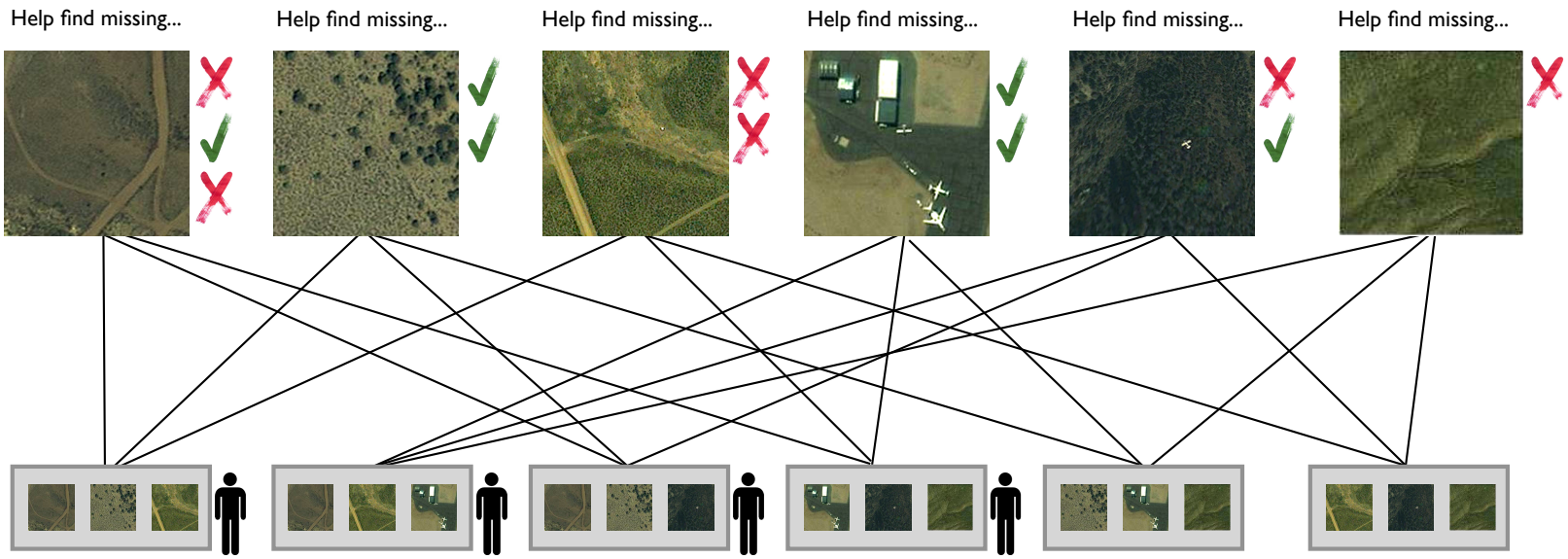


Add redundancy to cope with noise



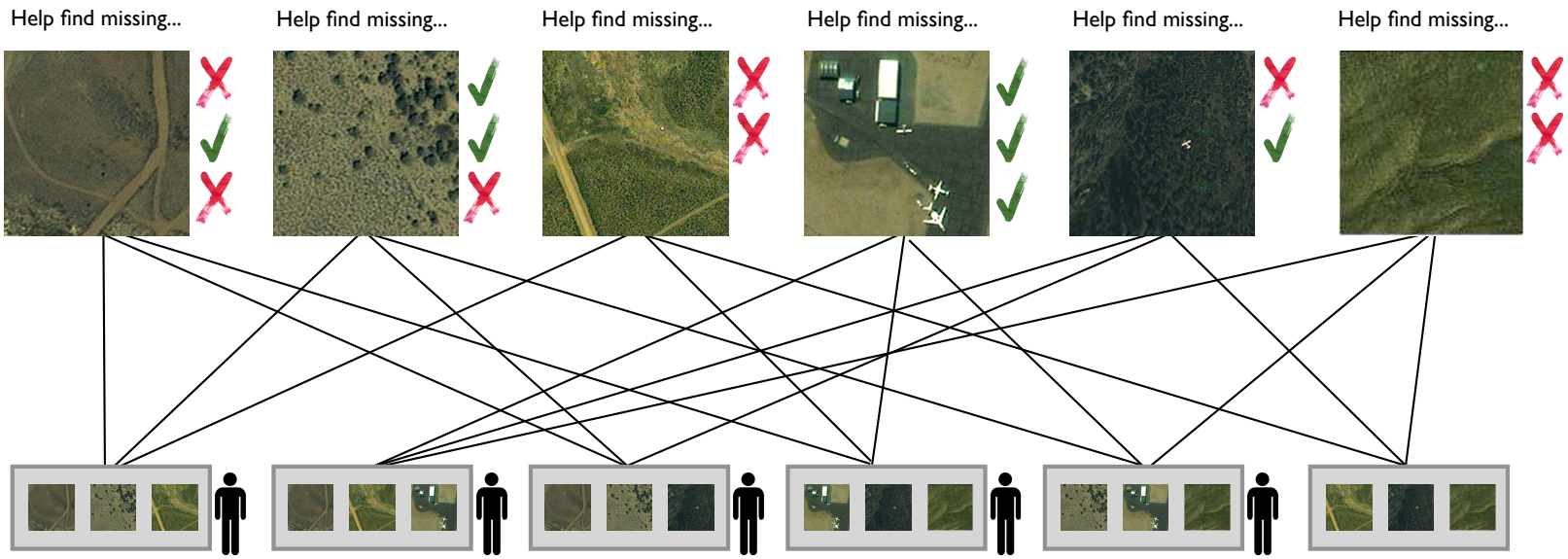


Add redundancy to cope with noise



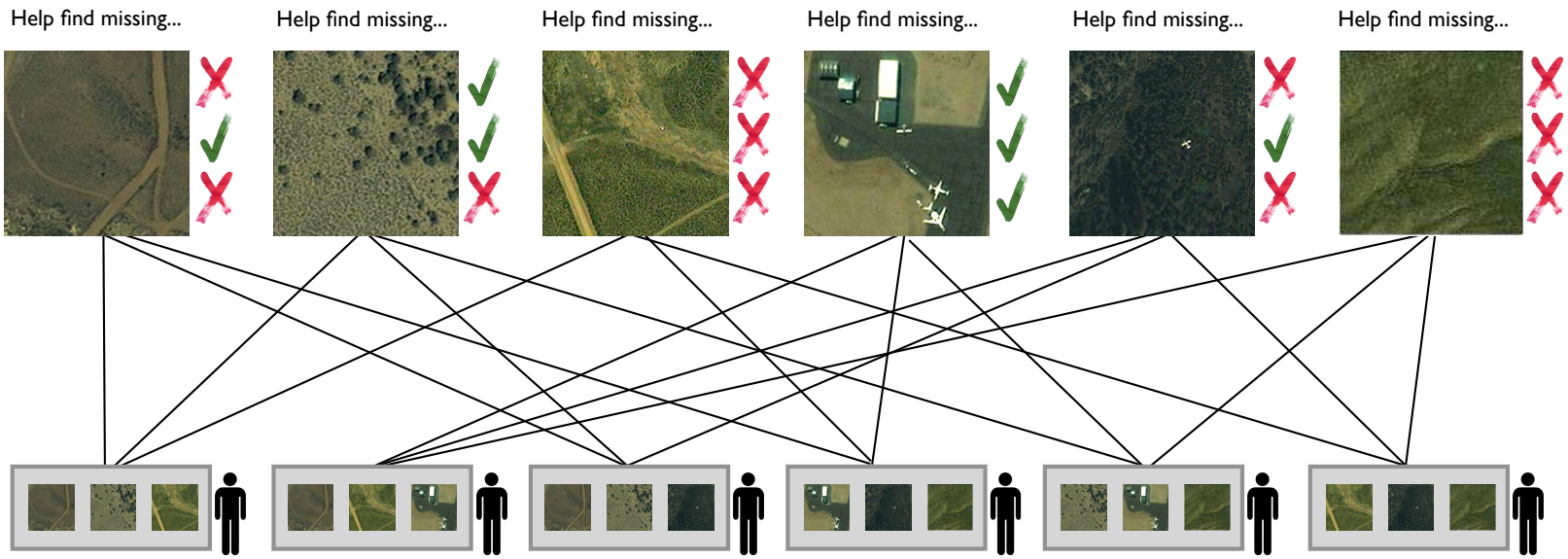
-1		+1	-1		
+1		+1			
-1			-1		
	+1		+1		
	-1	+1			
	-1				

Add redundancy to cope with noise



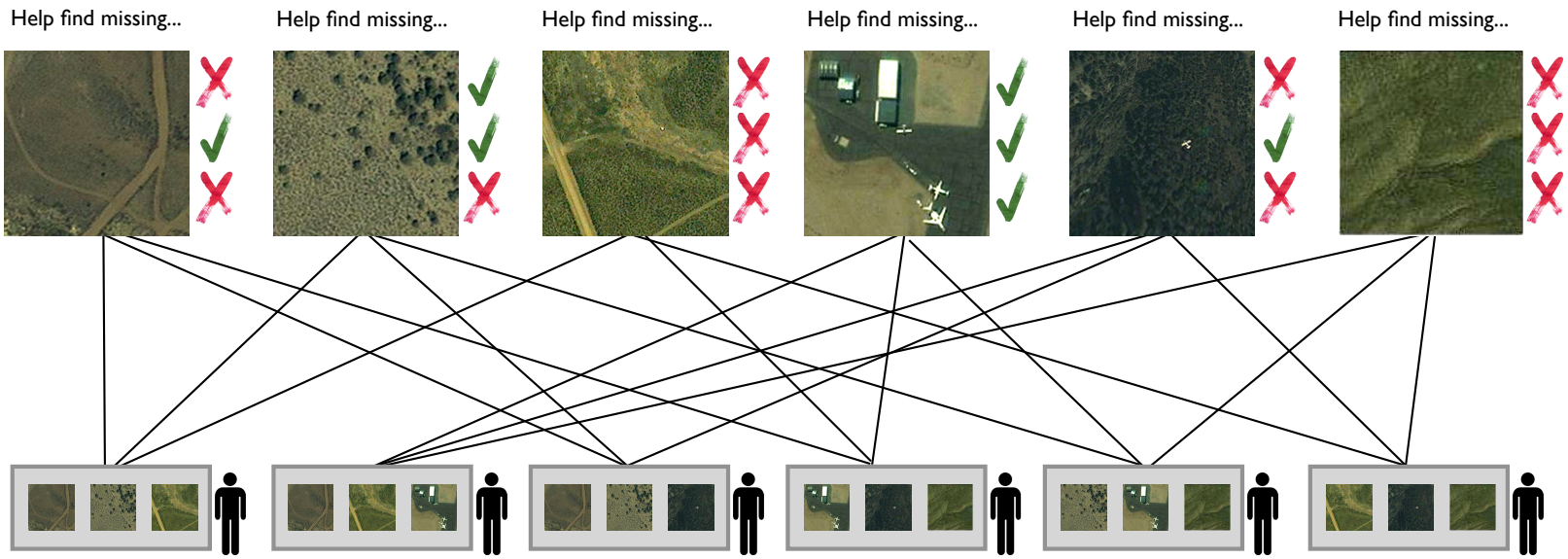
-1		+1	-1		
+1		+1		-1	
-1			-1		
	+1		+1	+1	
	-1	+1			
	-1			-1	

Add redundancy to cope with noise



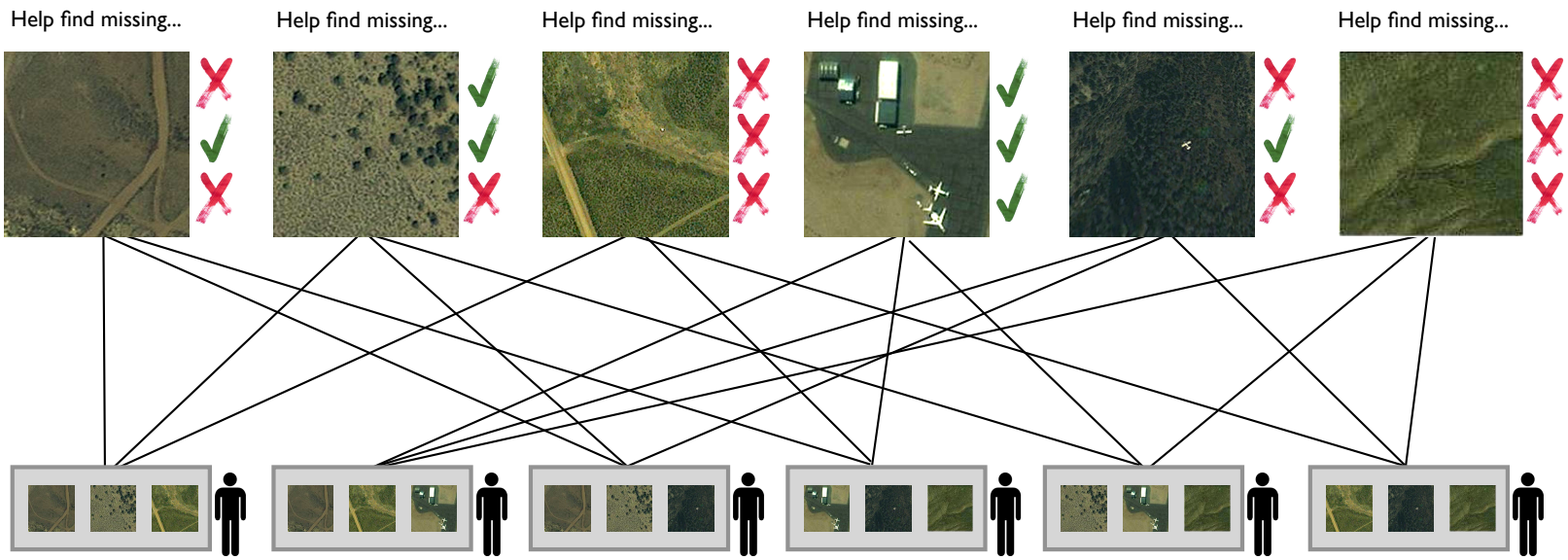
-1		+1	-1		
+1		+1		-1	
-1			-1		-1
	+1		+1	+1	
	-1	+1			-1
	-1			-1	-1

Add redundancy to cope with noise



						Majority Voting
-1		+1	-1			-1
+1		+1		-1		+1
-1			-1		-1	-1
	+1		+1	+1		+1
	-1	+1			-1	-1
	-1			-1	-1	-1

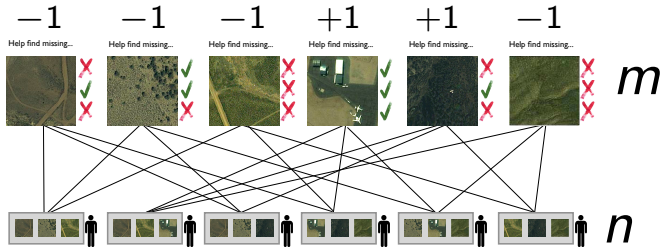
Add redundancy to cope with noise



Tradeoff: redundancy vs. accuracy

- special to crowdsourcing system:
  - we design the graph
  - workers arrive online fashion

Notations



- m questions
- unknown true answers $t \in \{-1, +1\}^m$
- task-assignment graph $G([m], [n], E)$

  $(i, j) \in E$ indicates that question i is asked to the j-th arriving worker
- n workers arrive online fashion and submit responses
- response matrix $A \in \{-1, 0, 1\}^{m \times n}$

$$A_{ij} = \begin{cases} 0 & \text{if not assigned} \\ +1 & \text{if worker answers "}+1\text{"} \\ -1 & \text{if worker answers "}-1\text{"} \end{cases}$$

$$\begin{array}{cc} & n \text{ workers} \longrightarrow \\ m \text{ tasks} \downarrow & \begin{bmatrix} \quad A \quad \end{bmatrix} \end{array}$$

- Dawid-Skene model from 1979:
  - Each worker is parametrized by a scalar value $p_j \in [0, 1]$
  - For each assigned question, answers correctly with probability p_j

$$A_{ij} = \begin{cases} t_i & \text{with probability } p_j \\ -t_i & \text{with probability } 1 - p_j \end{cases}$$

  - Criticism: all tasks are assumed to be equally difficult

Redundancy vs. Error Rate

$$\mu \triangleq \frac{1}{n}\sum_{j=1}^{n}(2p_j - 1) \qquad \sigma^2 \triangleq \frac{1}{n}\sum_{j=1}^{n}(2p_j - 1)^2$$

	p_1	p_2	p_3	p_4	p_5	p_6
t_1	-1		+1	-1		
t_2	+1		+1		-1	
t_3	-1			-1		-1
t_3		+1		+1	+1	
t_4		-1	+1			-1
t_5		-1			-1	-1

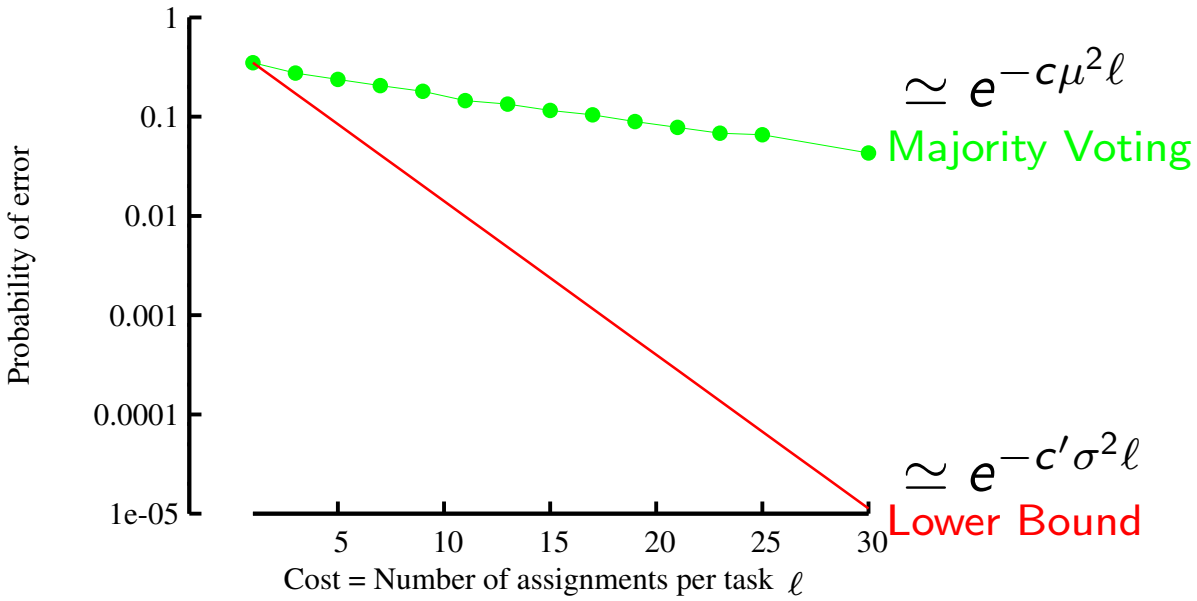
ℓ answers per task

Redundancy vs. Error Rate

$$\mu \triangleq \frac{1}{n}\sum_{j=1}^{n}(2p_j - 1) \qquad \sigma^2 \triangleq \frac{1}{n}\sum_{j=1}^{n}(2p_j - 1)^2$$

	p_1	p_2	p_3	p_4	p_5	p_6
t_1	-1		+1	-1		
t_2	+1		+1		-1	
t_3	-1			-1		-1
t_3		+1		+1	+1	
t_4		-1	+1			-1
t_5		-1			-1	-1

ℓ answers per task

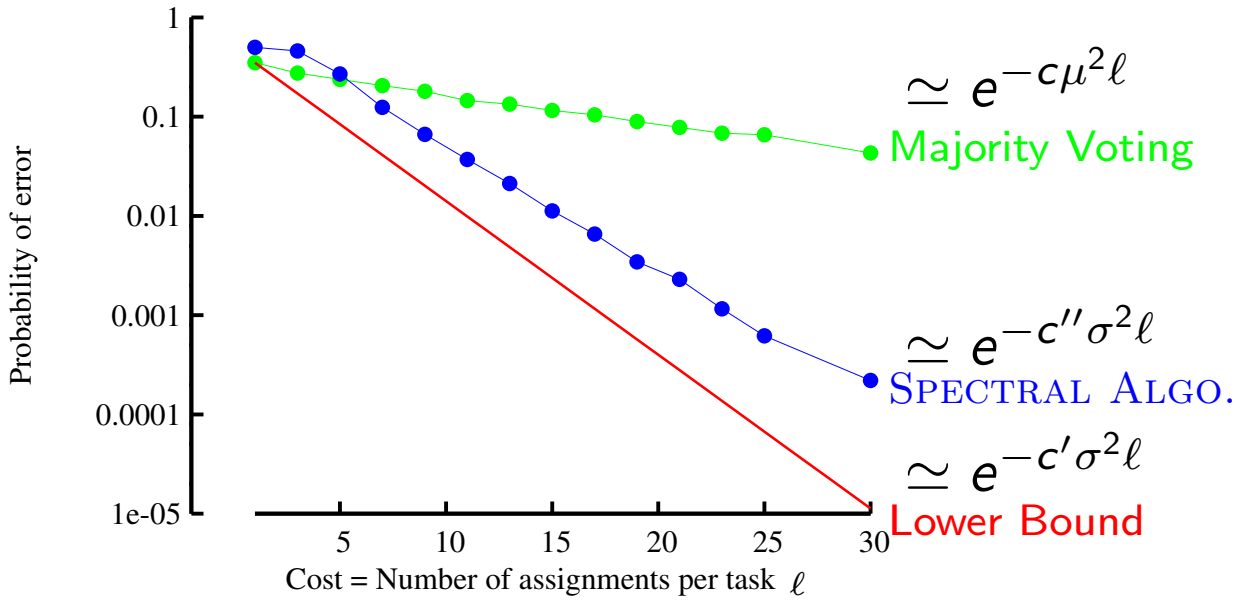


Redundancy vs. Error Rate

$$\mu \triangleq \frac{1}{n}\sum_{j=1}^{n}(2p_j - 1) \qquad \sigma^2 \triangleq \frac{1}{n}\sum_{j=1}^{n}(2p_j - 1)^2$$

	p_1	p_2	p_3	p_4	p_5	p_6
t_1	-1		+1	-1		
t_2	+1		+1		-1	
t_3	-1			-1		-1
t_3		+1		+1	+1	
t_4		-1	+1			-1
t_5		-1			-1	-1

ℓ answers per task



Dawid-Skene model [KargerOhShah11]

quality of the crowd $\sigma^2 \triangleq \frac{1}{n}\sum_{j=1}^{n}(2p_j - 1)^2$

Achievability: random graph E and SPECTRAL ALGORITHM achieves

$$P_{\mathrm{err}} \leq e^{-c\sigma^2\ell}$$

Fundamental limit: the best task-assignment and inference is limited by

$$\min_{E,\hat{t}} \max_{t\in\{\pm1\}^m, p\in\mathcal{F}_{\sigma^2}} P_{\mathrm{err}} \geq e^{-c'\sigma^2\ell}$$

Dawid-Skene model [KargerOhShah11]

quality of the crowd $\sigma^2 \triangleq \frac{1}{n}\sum_{j=1}^{n}(2p_j - 1)^2$

Achievability: random graph E and SPECTRAL ALGORITHM achieves

$$P_{\mathrm{err}} \leq e^{-c\sigma^2\ell}$$

Fundamental limit: the best task-assignment and inference is limited by

$$\min_{E,\hat{t}} \max_{t\in\{\pm1\}^m, p\in\mathcal{F}_{\sigma^2}} P_{\mathrm{err}} \geq e^{-c'\sigma^2\ell}$$

Disturbing fact: the best adaptive scheme is limited by

$$\min_{\text{all adaptive schemes},\hat{t}} \max_{t\in\{\pm1\}^m, p\in\mathcal{F}_{\sigma^2}} P_{\mathrm{err}} \geq e^{-c''\sigma^2\ell}$$

Adaptive schemes are common in practice

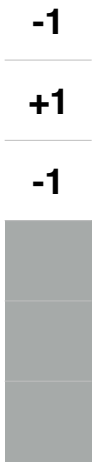
- in practice, adaptive schemes improve significantly
- in theory, the gain is minimal

Adaptive schemes are common in practice

- in practice, adaptive schemes improve significantly
- in theory, the gain is minimal

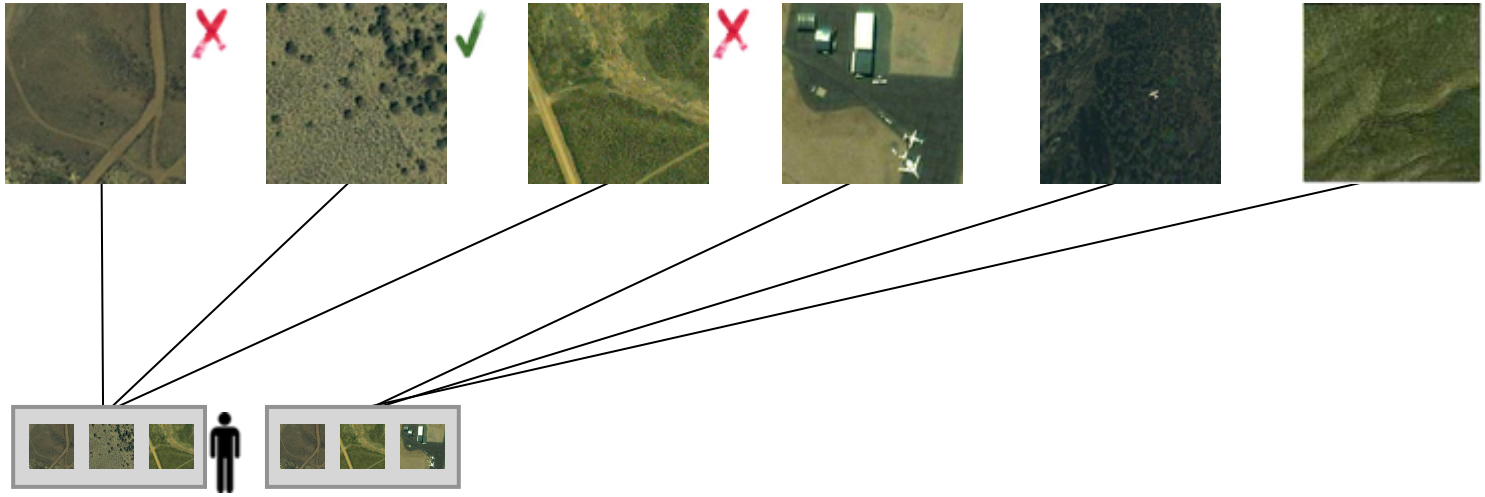
Adaptive schemes are common in practice

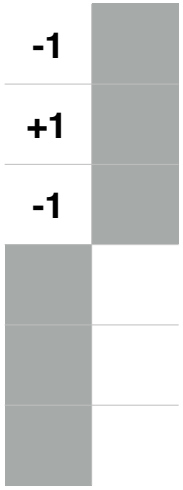
- in practice, adaptive schemes improve significantly
- in theory, the gain is minimal



Adaptive schemes are common in practice

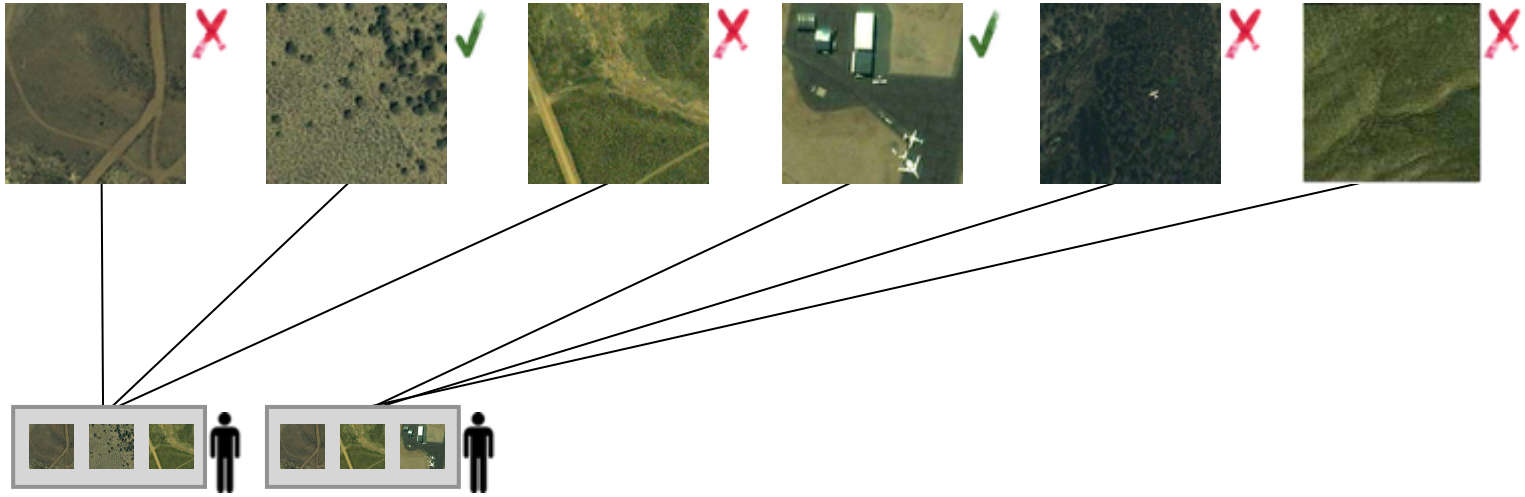
- in practice, adaptive schemes improve significantly
- in theory, the gain is minimal





Adaptive schemes are common in practice

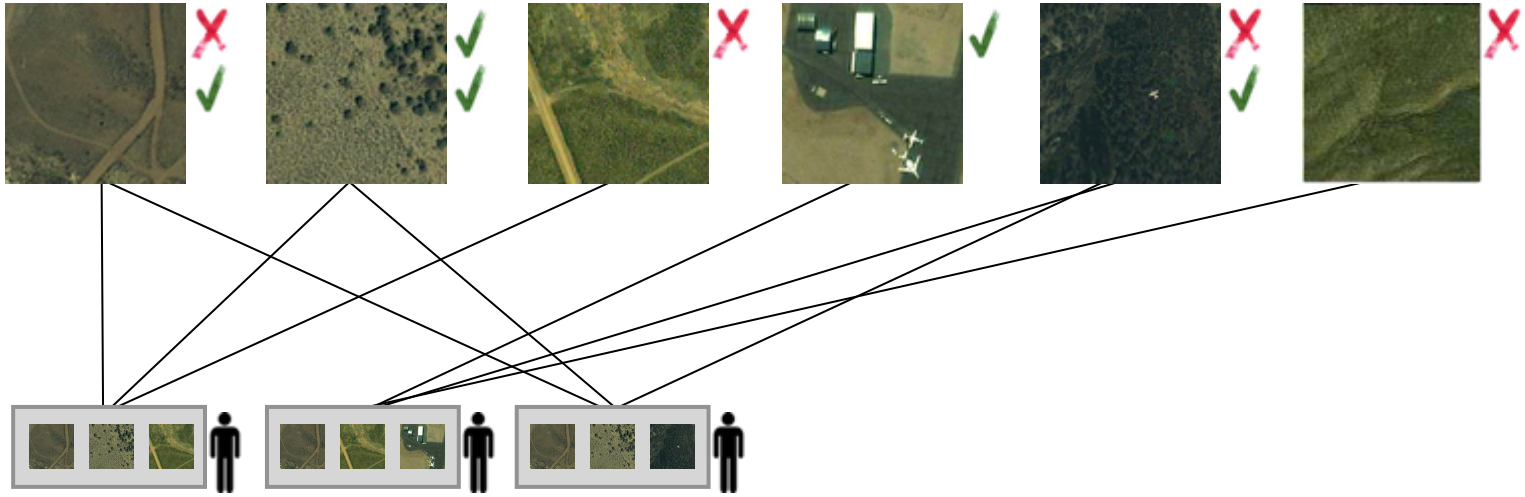
- in practice, adaptive schemes improve significantly
- in theory, the gain is minimal



-1	
+1	
-1	
	+1
	-1
	-1

Adaptive schemes are common in practice

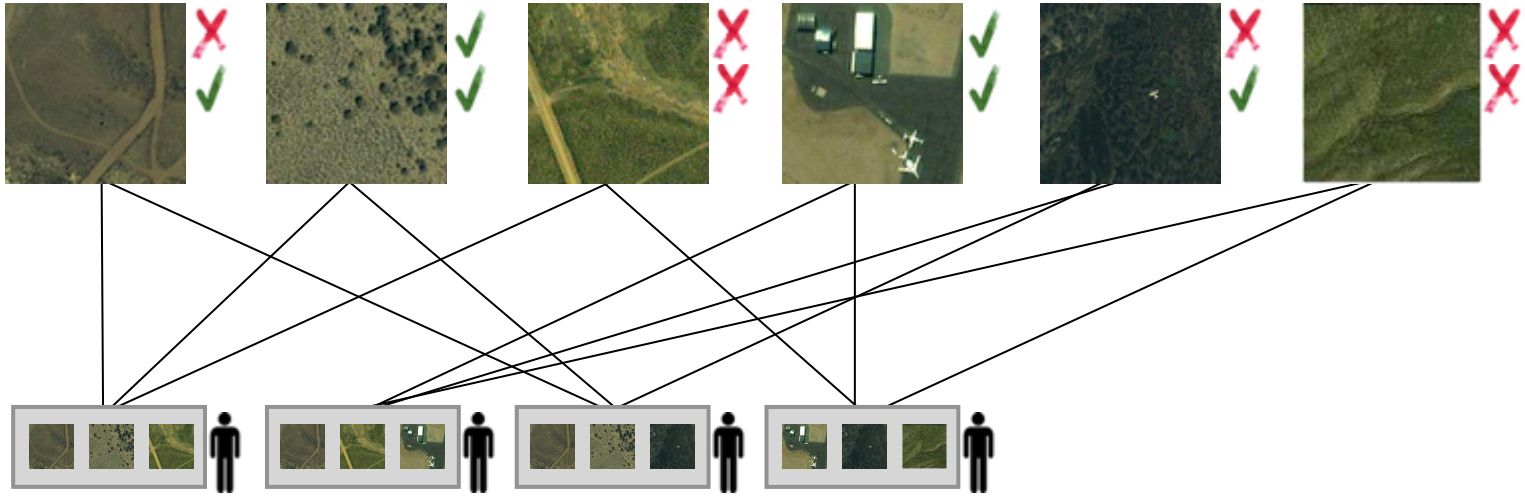
- in practice, adaptive schemes improve significantly
- in theory, the gain is minimal



-1		+1
+1		+1
-1		
	+1	
	-1	+1
	-1	

Adaptive schemes are common in practice

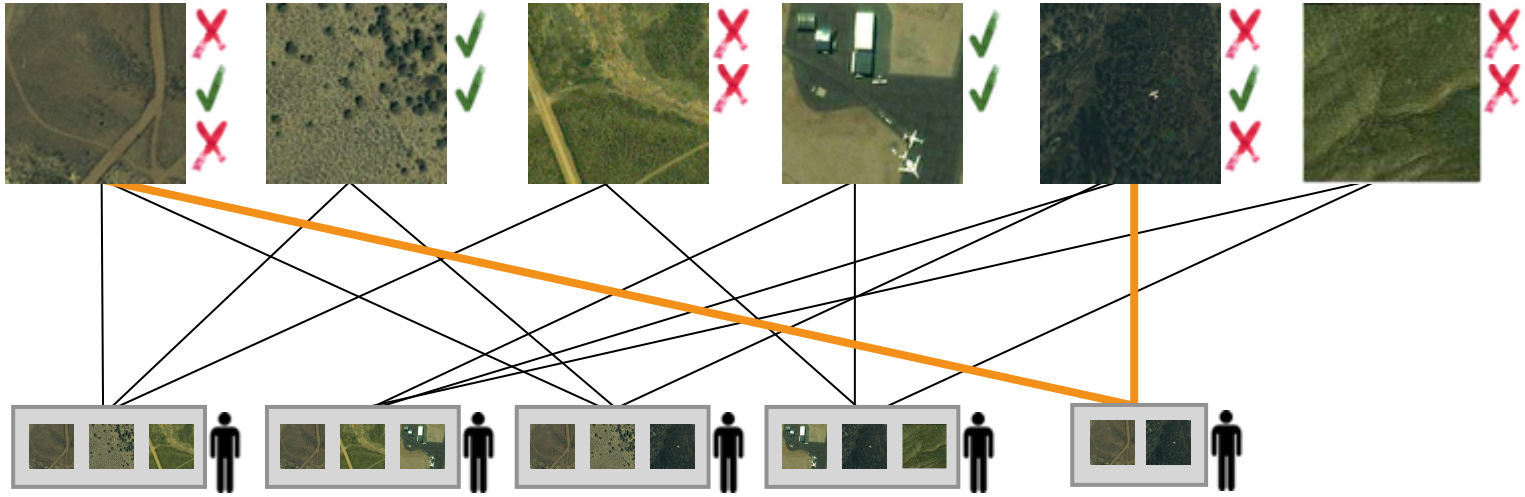
- in practice, adaptive schemes improve significantly
- in theory, the gain is minimal



-1		+1	
+1		+1	
-1			-1
	+1		+1
	-1	+1	
	-1		-1

Adaptive schemes are common in practice

- in practice, adaptive schemes improve significantly
- in theory, the gain is minimal



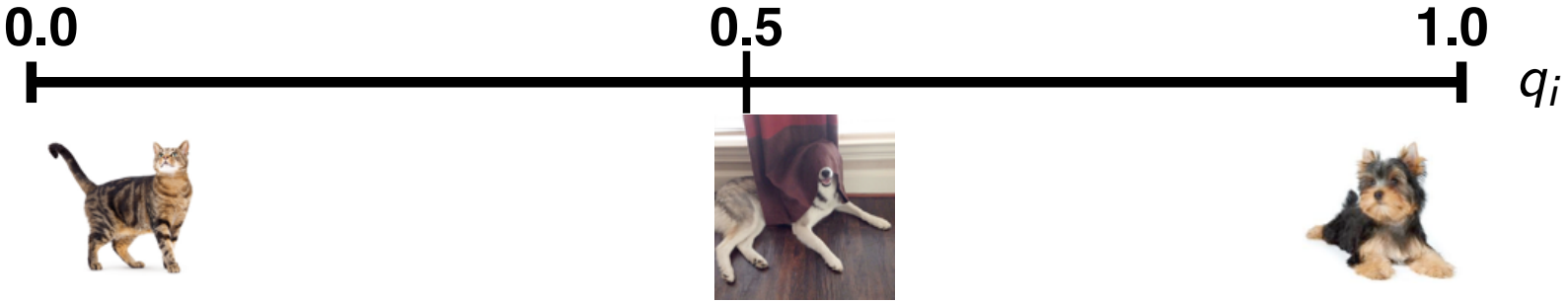
-1		+1		-1
+1		+1		
-1			-1	
	+1		+1	
	-1	+1		-1
	-1		-1	

Generalized DS model [ZhouLiuPlattMeekShah15]

- Definition
  - Each task is parametrized by a scalar value $q_i \in [0, 1]$
  - When a task is presented to a worker, it is perceived as a positive task with probability q_i

$$A_{ij} = \begin{cases} +1 & \text{with probability } q_i p_j + (1 - q_i)(1 - p_j) \\ -1 & \text{with probability } q_i(1 - p_j) + (1 - q_i)p_j \end{cases}$$

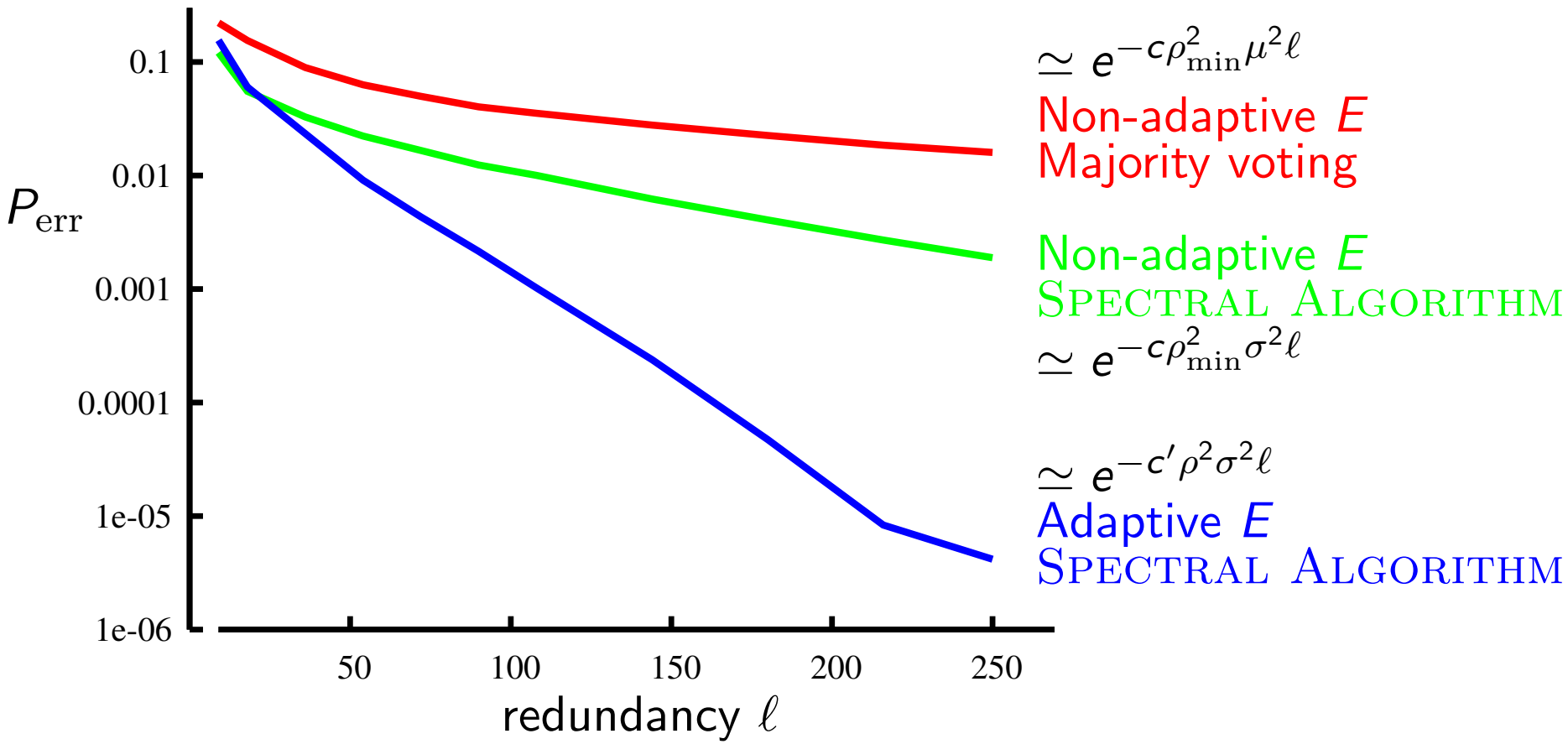
  - ground truth is $t_i = \mathbb{I}(q_i > 0.5) - \mathbb{I}(q_i < 0.5)$
  - difficulty level of a task i measured by $(2q_i - 1)^2$
  - recovers the DS model when all $q_i \in \{0, 1\}$



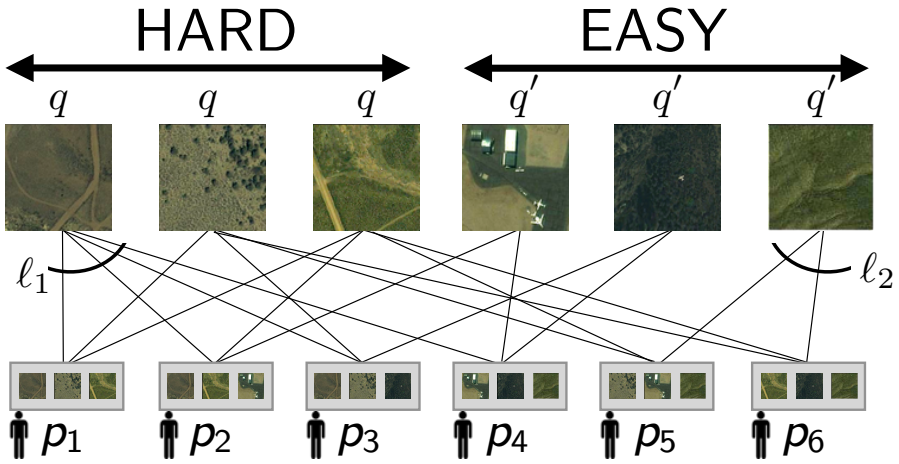
Does adaptivity help?

Collective task difficulty:

$$\rho^2_{\min} = \min_i (q_i - 1)^2 \,, \qquad \rho^2 = \frac{1}{\frac{1}{m}\sum_{i=1}^{m} \frac{1}{(2q_i-1)^2}}$$



Minimax rate for adaptive scenario [KhetanOh16]



- task difficulty $\rho^2 = \frac{1}{\frac{1}{m}\sum_{i=1}^{m}\frac{1}{(2q_i-1)^2}}$

Fundamental limit: the best adaptive scheme and inference is limited by

$$\min_{\text{adaptive}E,\hat{t}} \max_{q\in\mathcal{G}_{\rho^2},p\in\mathcal{F}_{\sigma^2}} P_{\text{err}} \geq e^{-C'\rho^2\sigma^2\ell}$$

Achievability: efficient adaptive scheme and inference achieves

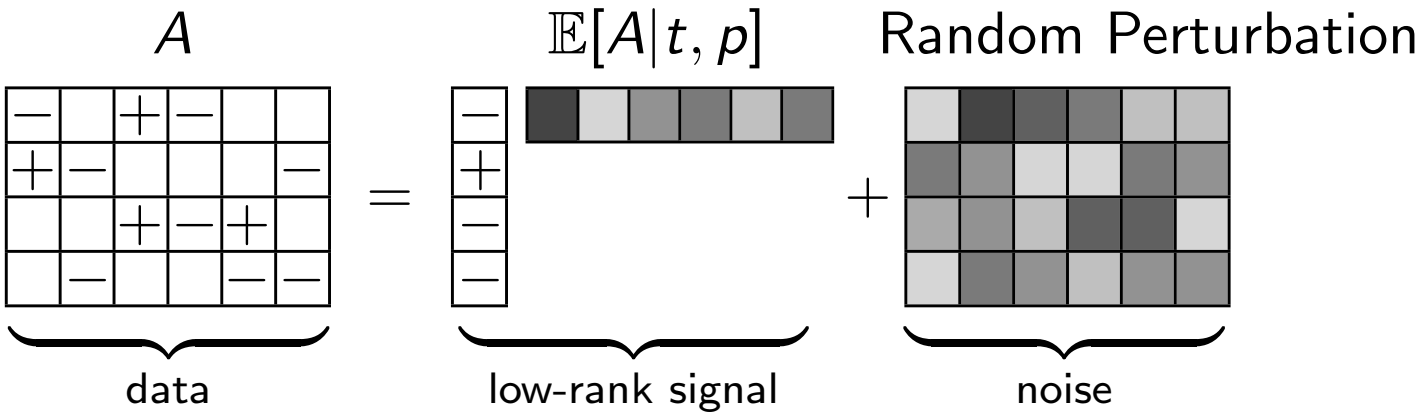
$$P_{\text{err}} \leq e^{-C\rho^2\sigma^2\ell}$$

Fundamental limit: the best non-adaptive scheme is limited by

$$\min_{E,\hat{t}} \max_{q\in\mathcal{G}_{\rho^2},p\in\mathcal{F}_{\sigma^2}} P_{\text{err}} \geq e^{-C''\rho^2_{\min}\sigma^2\ell}$$

Spectral Algorithm for the original DS model

- Noisy Matrix Completion



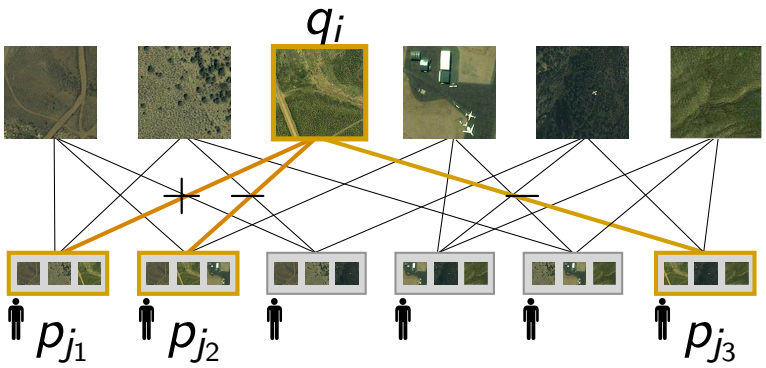
$$\mathbb{E}[A_{ij}|t_i,p_j] = \frac{\ell}{n}\cdot t_i\cdot(2p_j-1)$$

$$\mathbb{E}[A|t,p] = \frac{\ell}{n}\cdot t\cdot(2p-\mathbb{1})^T$$

- Matrix completion algorithm only gives

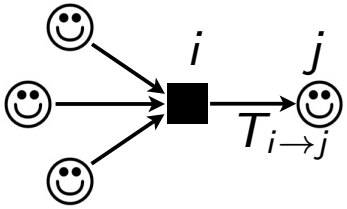
$$P_{\mathrm{err}} \leq \frac{C}{\sigma^2\ell}$$

Spectral Algorithm



- Message-passing algorithm
  - Two sets of messages:
    - ★ Task messages $\{T_{i\to j}\}$, and worker messages $\{W_{j\to i}\}$
  - Initialize worker messages as random Gaussian: $W_{j\to i} \sim \mathcal{N}(1,1)$
  - Iteratively update messages

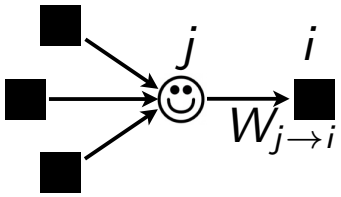
Task-likelihood update



$$\underbrace{T_{i\to j}}_{\textit{likelihood}} = \sum_{j'\neq j} \underbrace{W_{j'\to i}}_{\textit{reliability}} A_{ij'}$$

A task is likely to be '+' if reliable workers agree that it is '+'

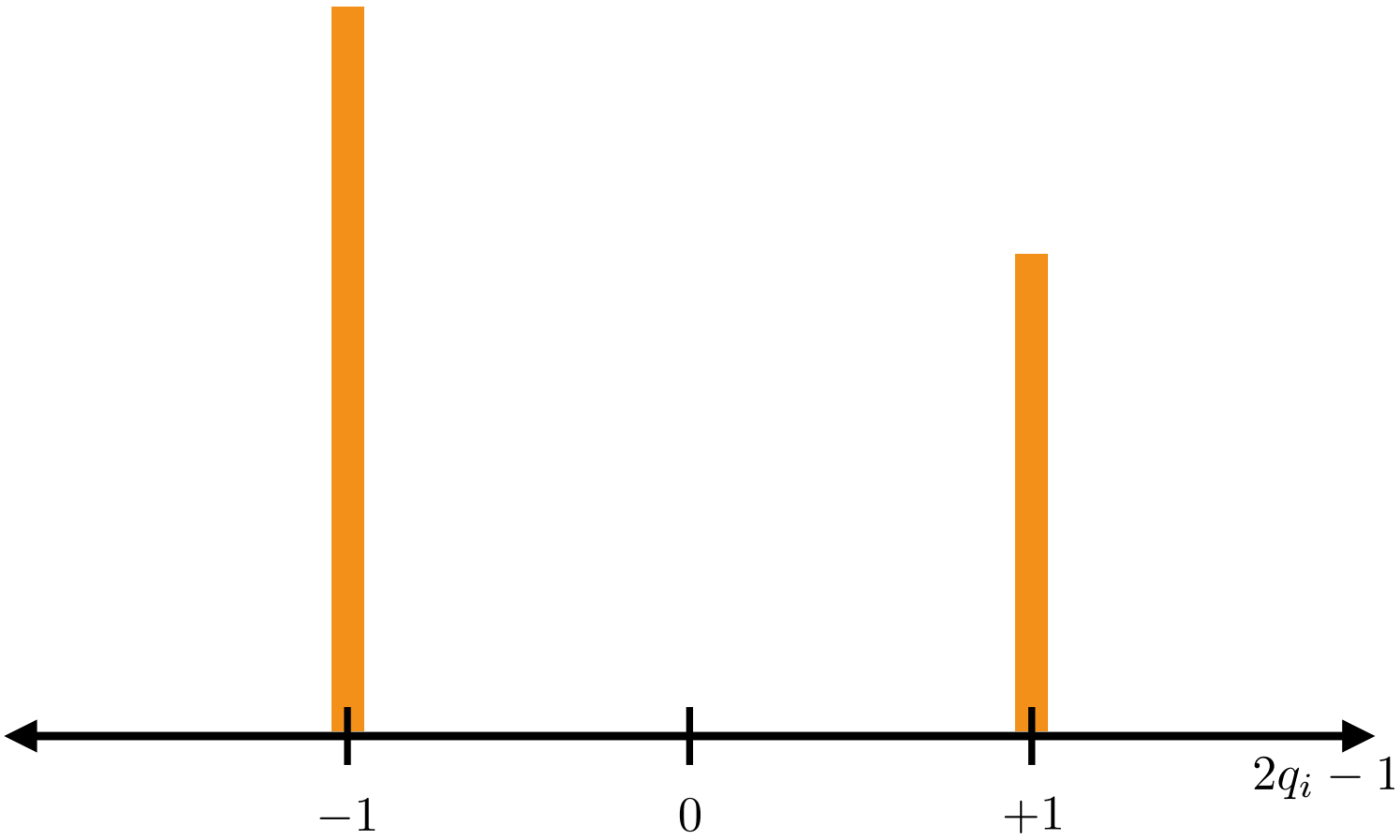
Worker-reliability update



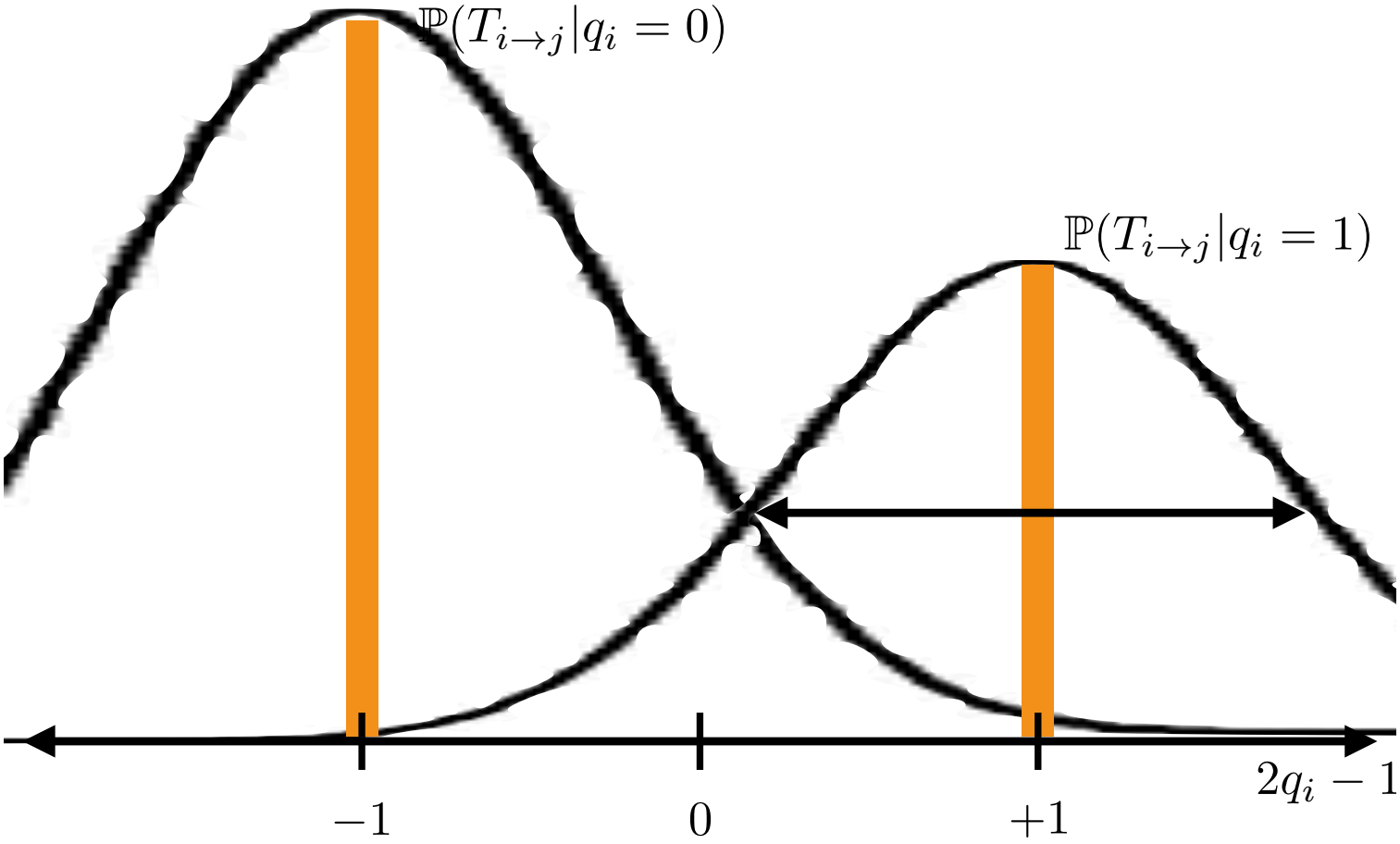
$$\underbrace{W_{j\to i}}_{\textit{reliability}} = \sum_{i'\neq i} \underbrace{T_{i'\to j}}_{\textit{likelihood}} A_{i'j}$$

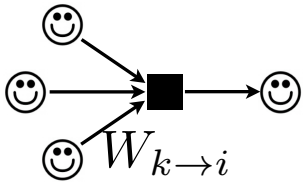
A worker is reliable if the worker agreed with our belief on other tasks

DS model + non-adaptive scheme



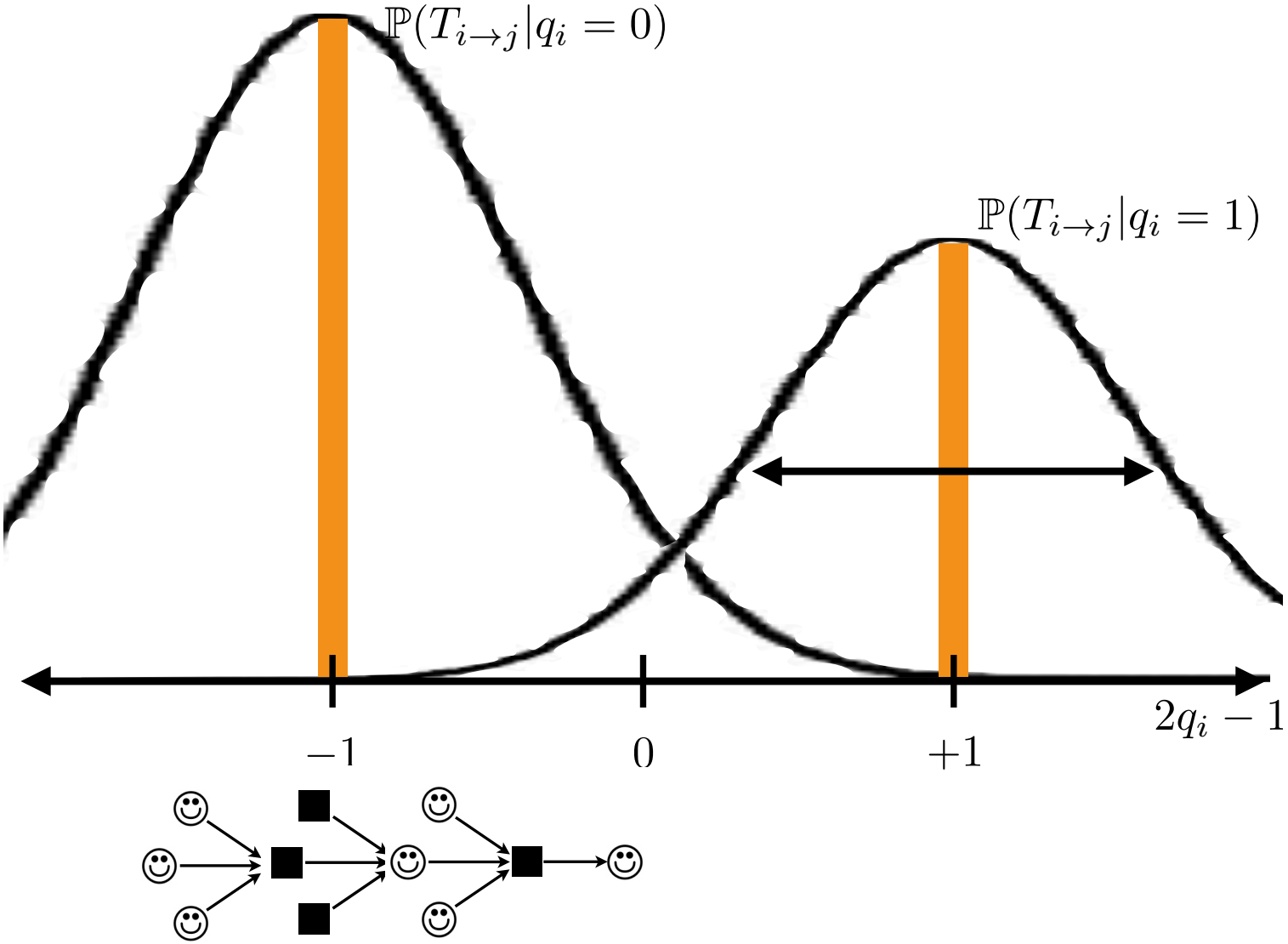
DS model + non-adaptive scheme



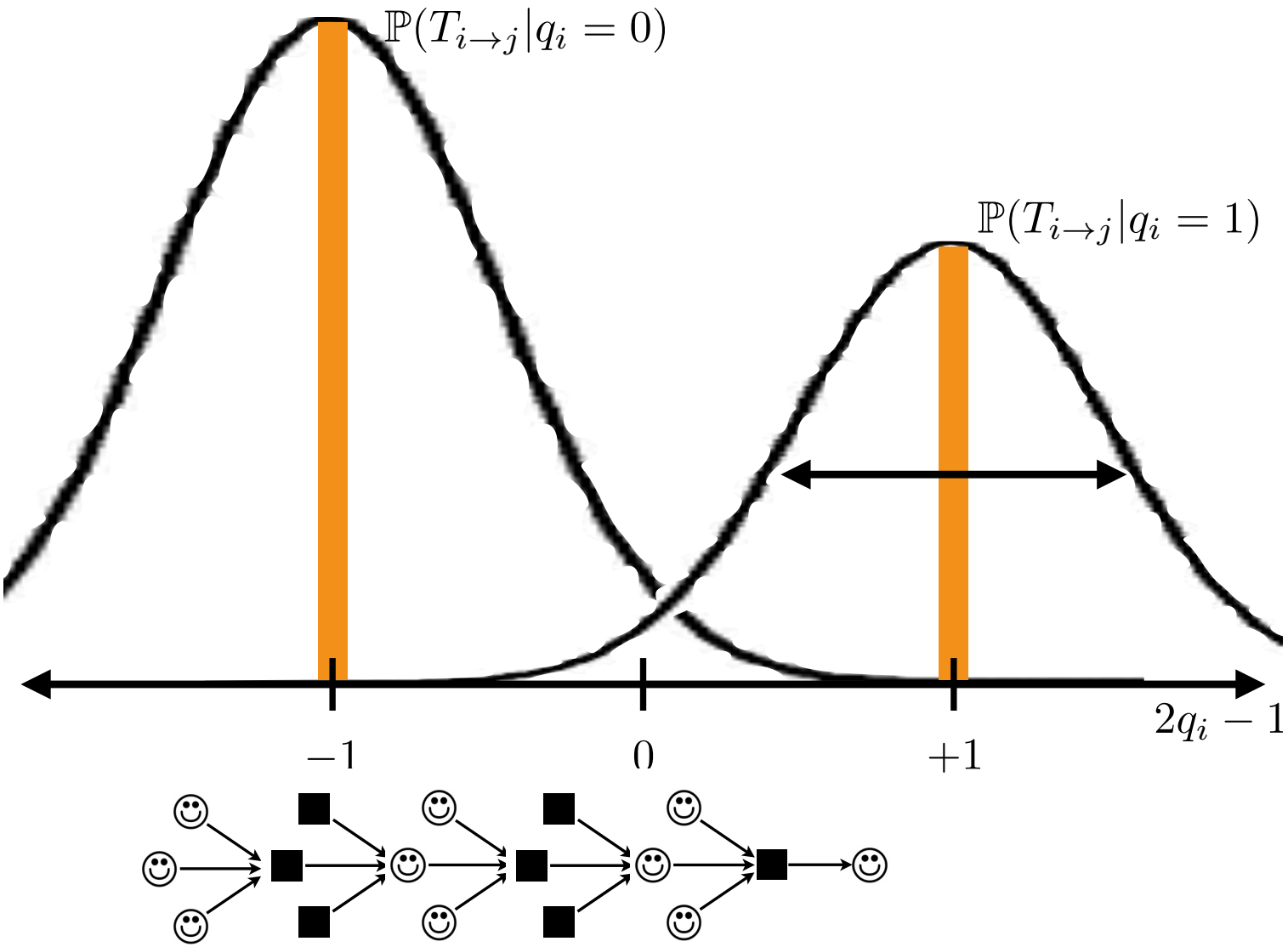


$$T_{i\to j} = \sum A_{ik} W_{k\to i}$$

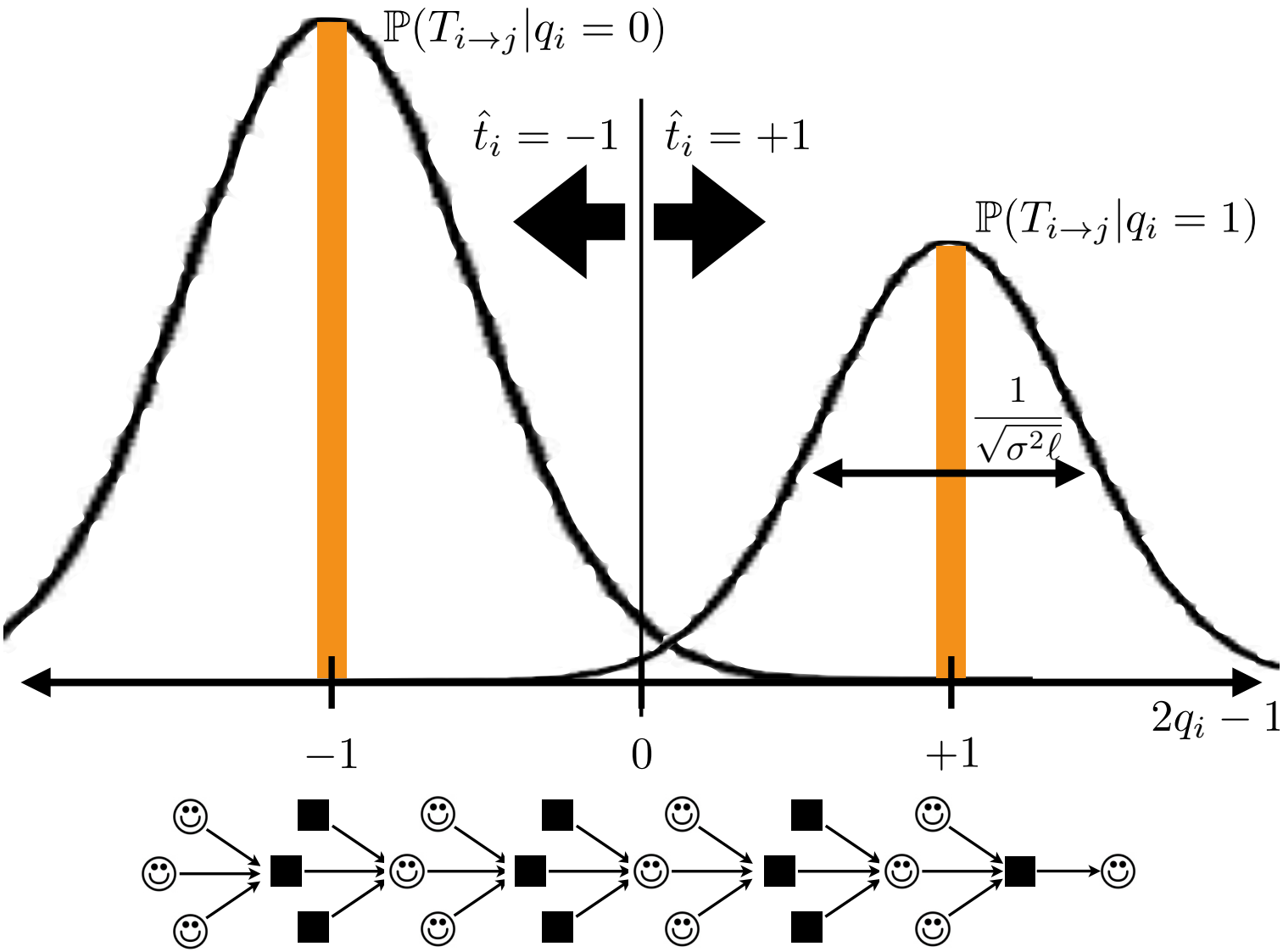
DS model + non-adaptive scheme



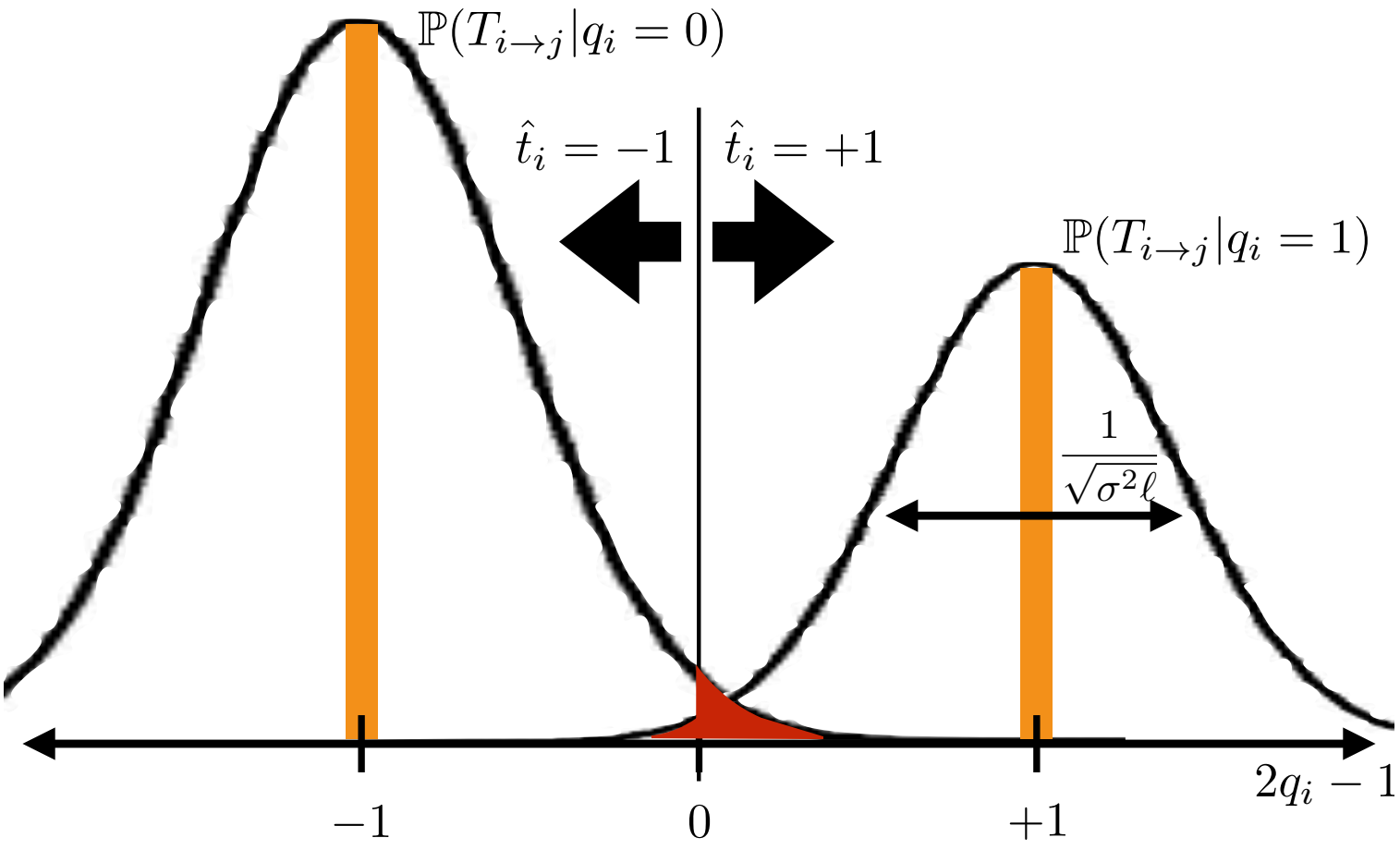
DS model + non-adaptive scheme



DS model + non-adaptive scheme

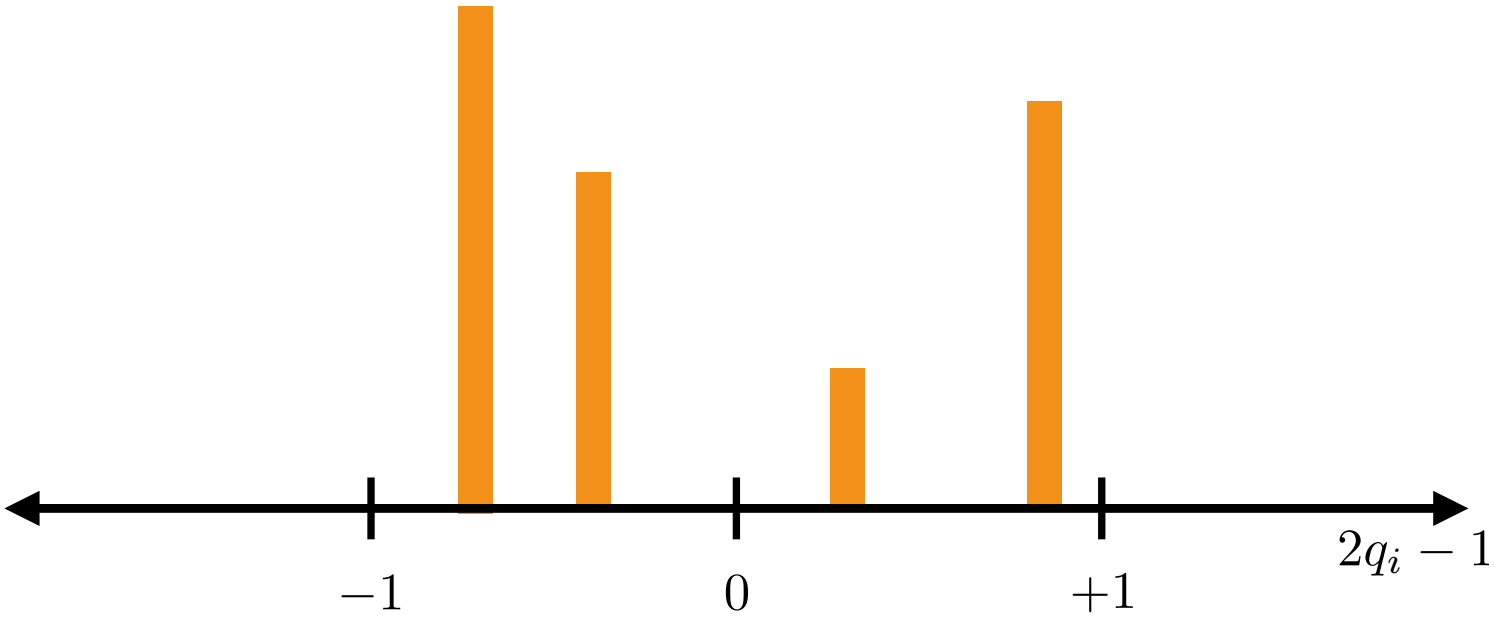


DS model + non-adaptive scheme



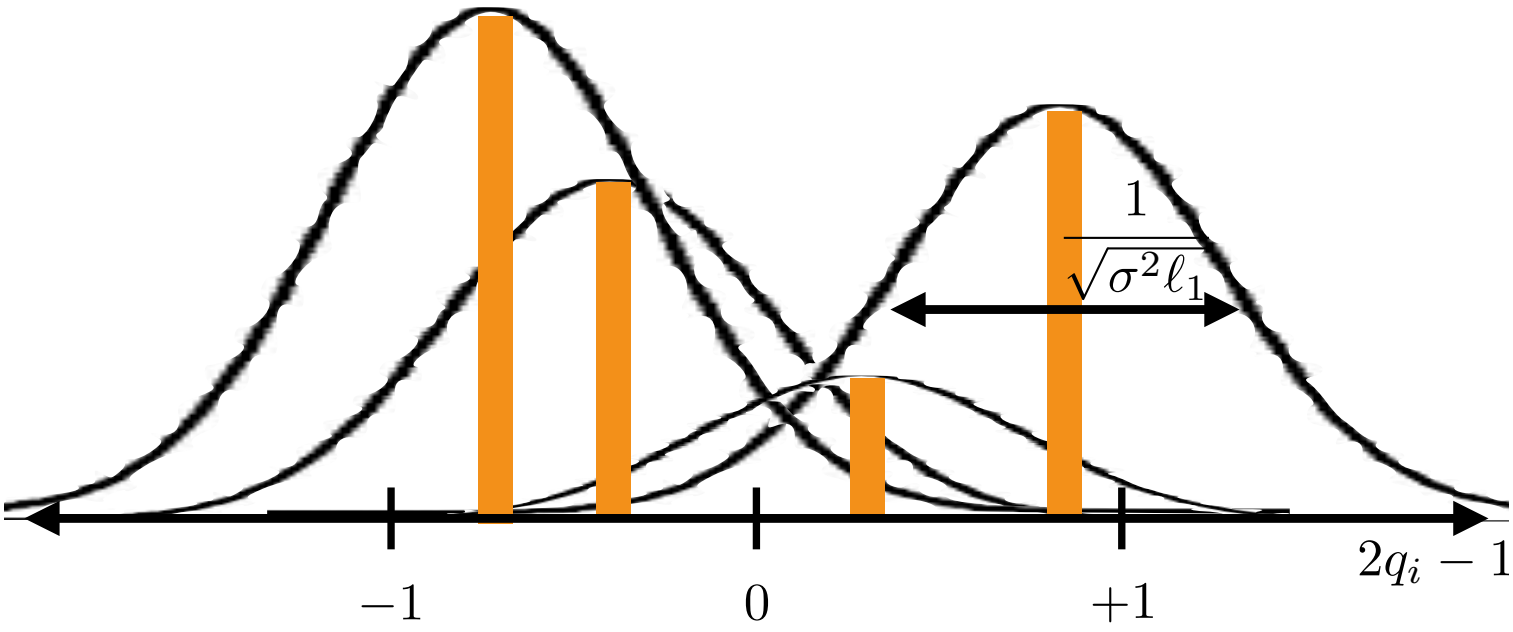
$$P_{\text{err}} \leq e^{-C\sigma^2 \ell}$$

Generalized DS model + adaptive scheme



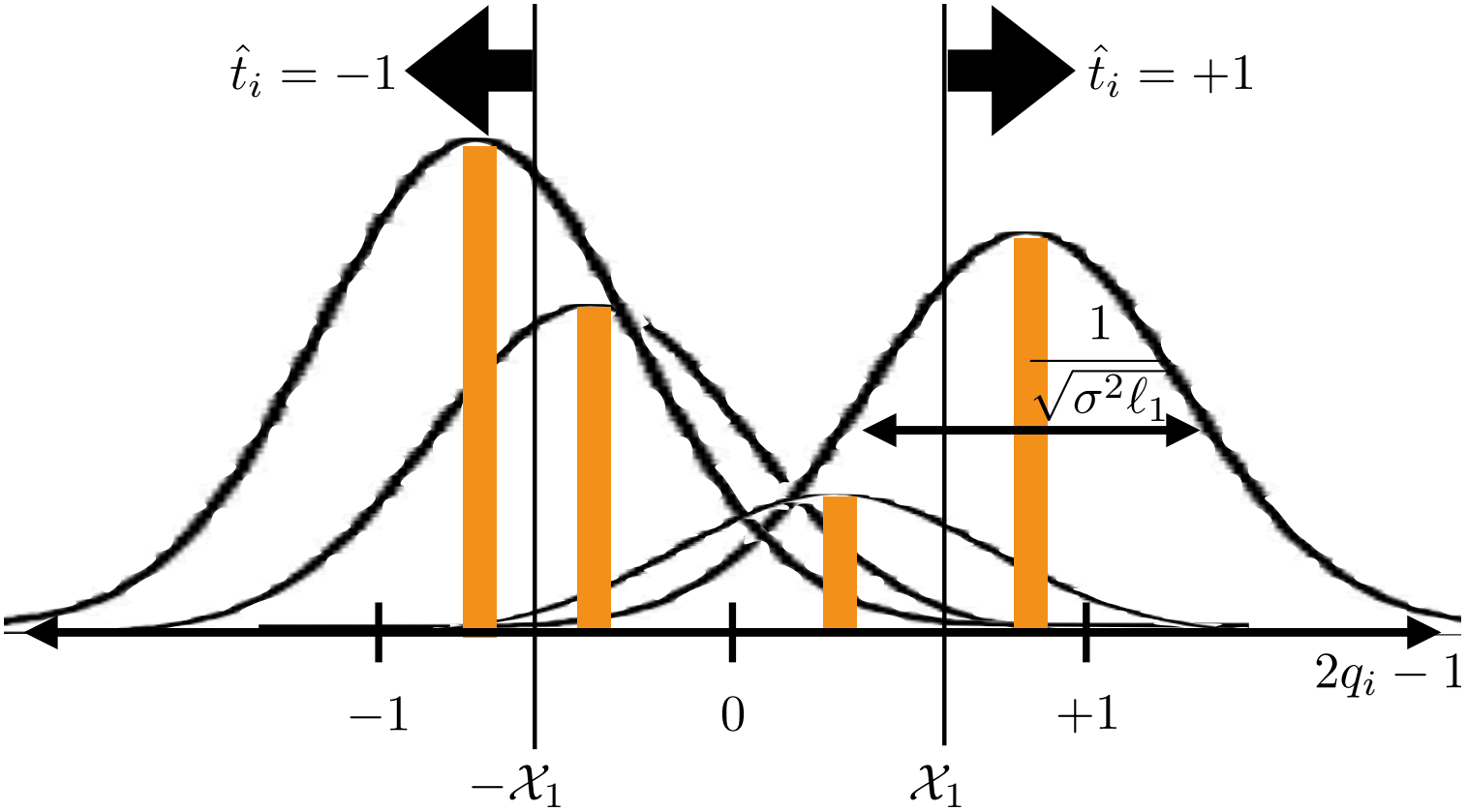
- Repeat rounds $t \in \{1, 2, \ldots\}$
  - Assign tasks with random (ℓ_t, r)-regular random graph
  - Run Spectral Algorithm
  - classify high-confidence tasks with threshold $\mathcal{X}_t$

Generalized DS model + adaptive scheme



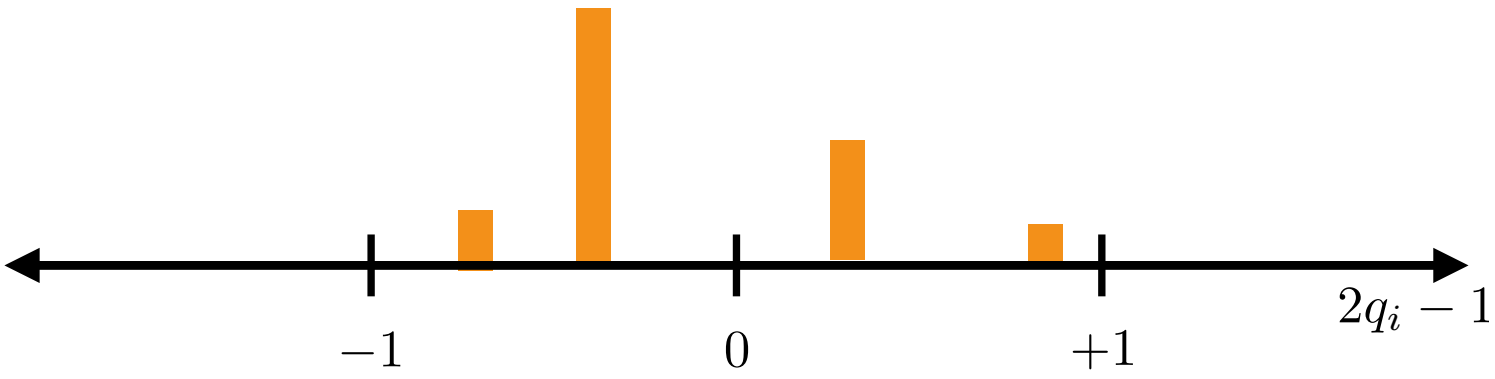
- Repeat rounds $t \in \{1, 2, \ldots\}$
  - Assign tasks with random (ℓ_t, r)-regular random graph
  - Run Spectral Algorithm
  - classify high-confidence tasks with threshold $\mathcal{X}_t$

Generalized DS model + adaptive scheme



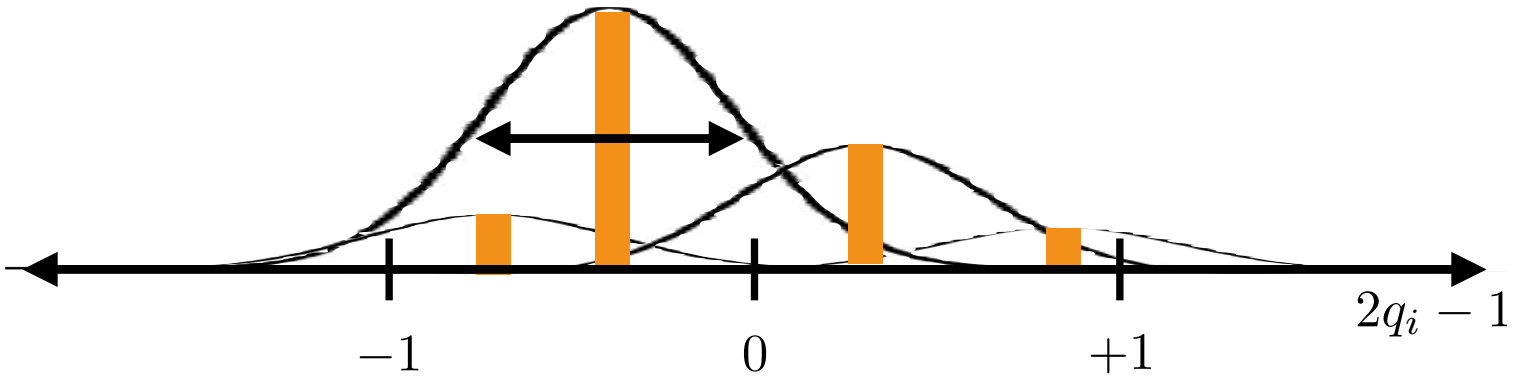
- Repeat rounds $t \in \{1, 2, \ldots\}$
  - Assign tasks with random (ℓ_t, r)-regular random graph
  - Run Spectral Algorithm
  - classify high-confidence tasks with threshold $\mathcal{X}_t$

Generalized DS model + adaptive scheme



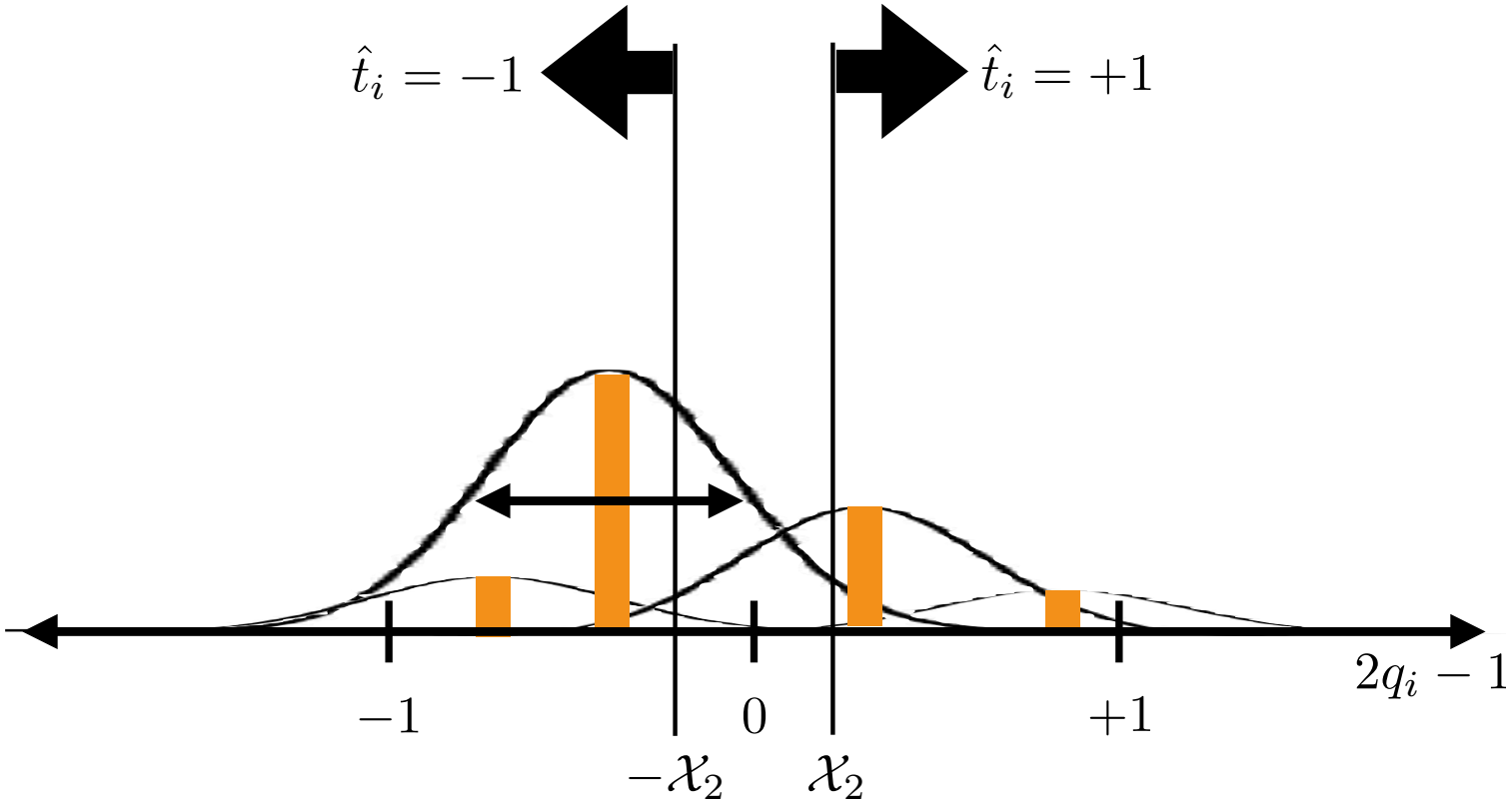
- Repeat rounds $t \in \{1, 2, \ldots\}$
  - Assign tasks with random (ℓ_t, r)-regular random graph
  - Run Spectral Algorithm
  - classify high-confidence tasks with threshold $\mathcal{X}_t$

Generalized DS model + adaptive scheme



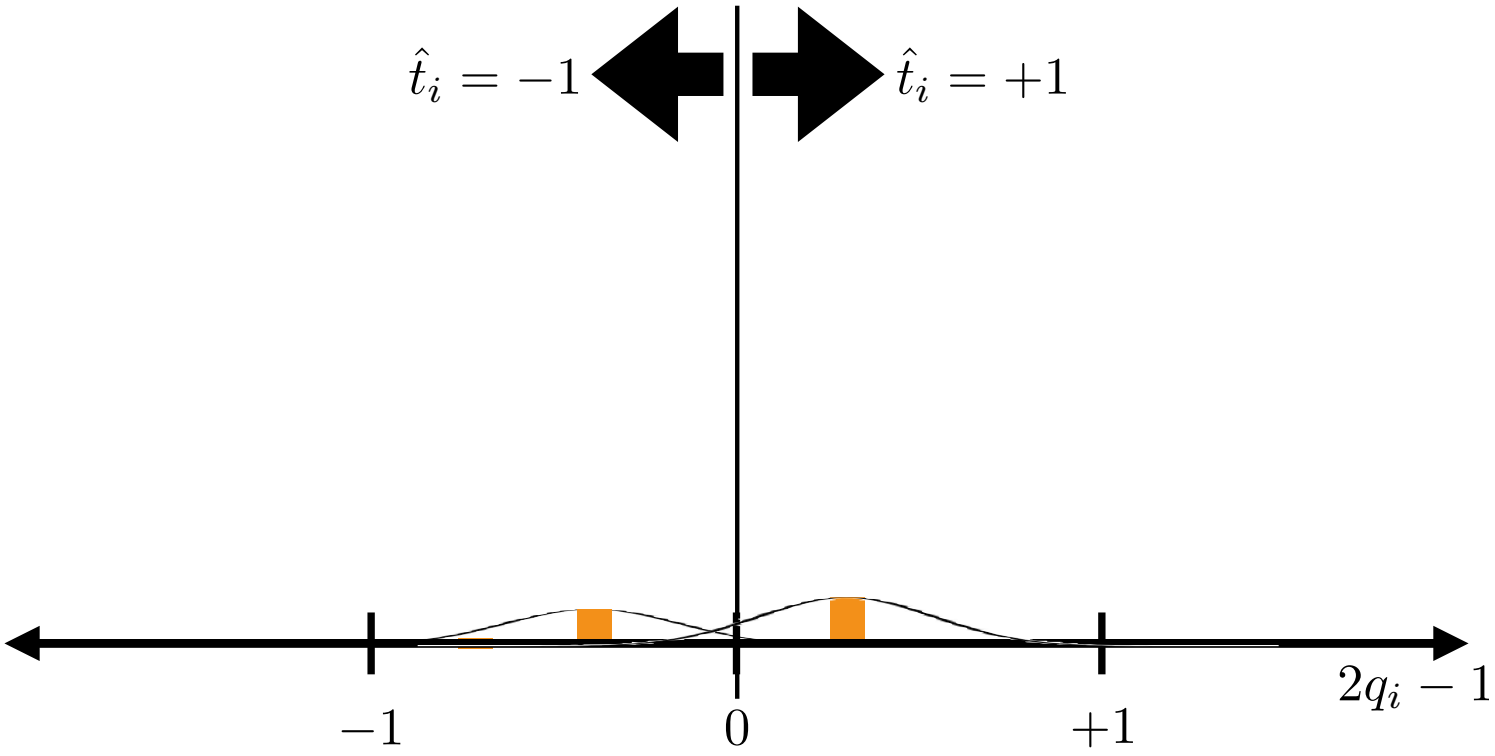
- Repeat rounds $t \in \{1, 2, \ldots\}$
  - Assign tasks with random (ℓ_t, r)-regular random graph
  - Run Spectral Algorithm
  - classify high-confidence tasks with threshold $\mathcal{X}_t$

Generalized DS model + adaptive scheme



- Repeat rounds $t \in \{1, 2, \ldots\}$
  - Assign tasks with random (ℓ_t, r)-regular random graph
  - Run Spectral Algorithm
  - classify high-confidence tasks with threshold $\mathcal{X}_t$

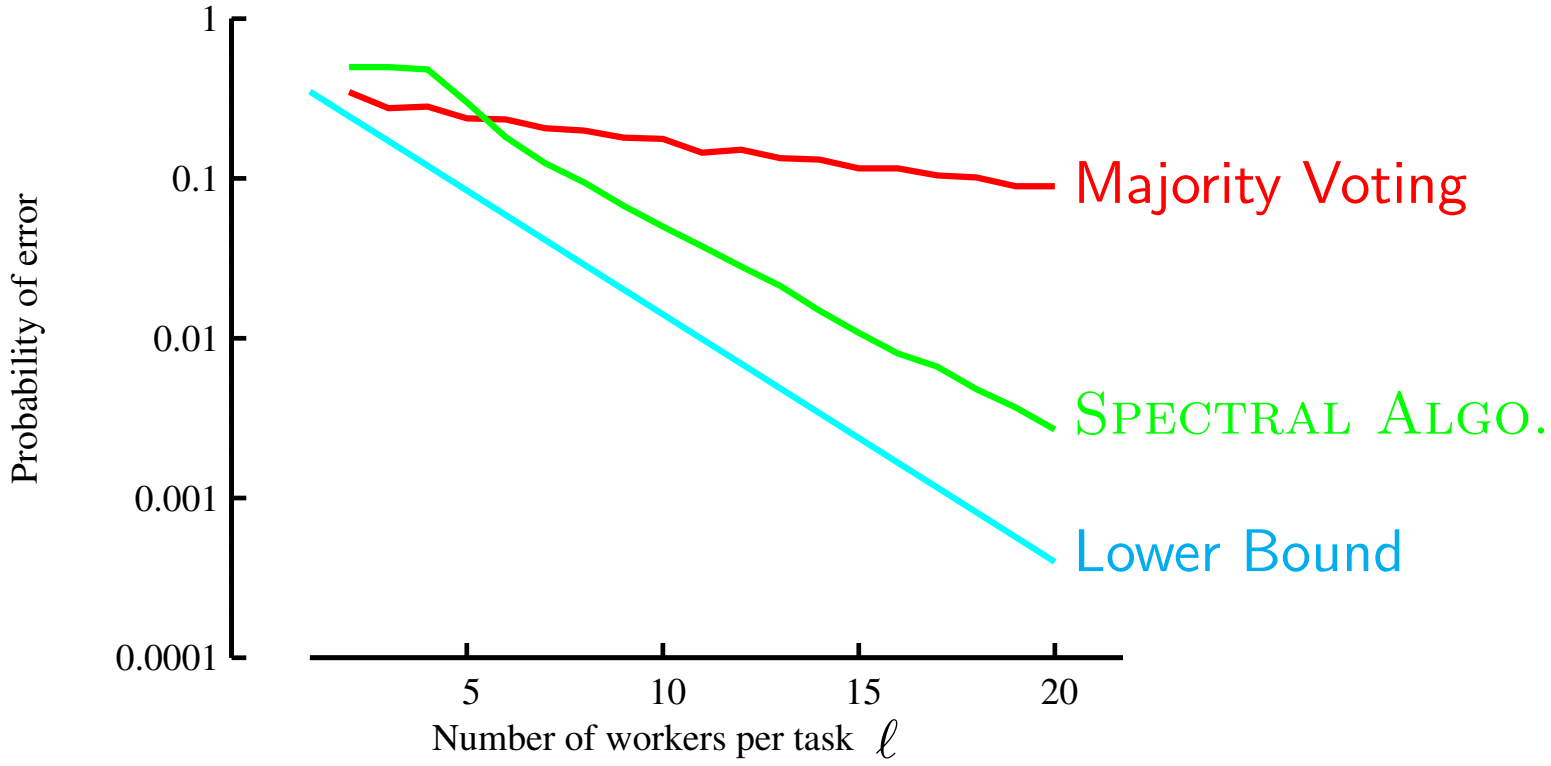
Generalized DS model + adaptive scheme



- Repeat rounds $t \in \{1, 2, \ldots\}$
  - Assign tasks with random (ℓ_t, r)-regular random graph
  - Run Spectral Algorithm
  - classify high-confidence tasks with threshold $\mathcal{X}_t$

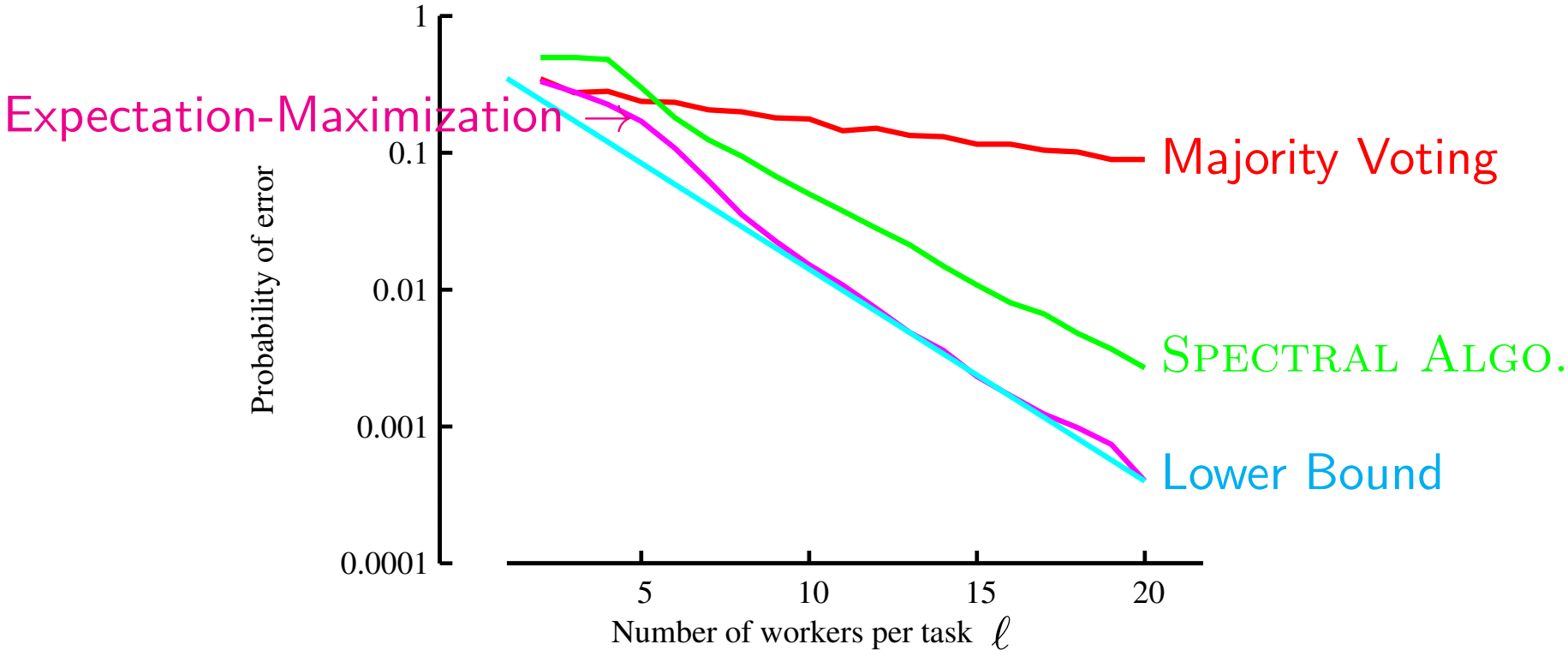
Why use only spectral?

Two 2-phase approaches



Why use only spectral?

Two 2-phase approaches



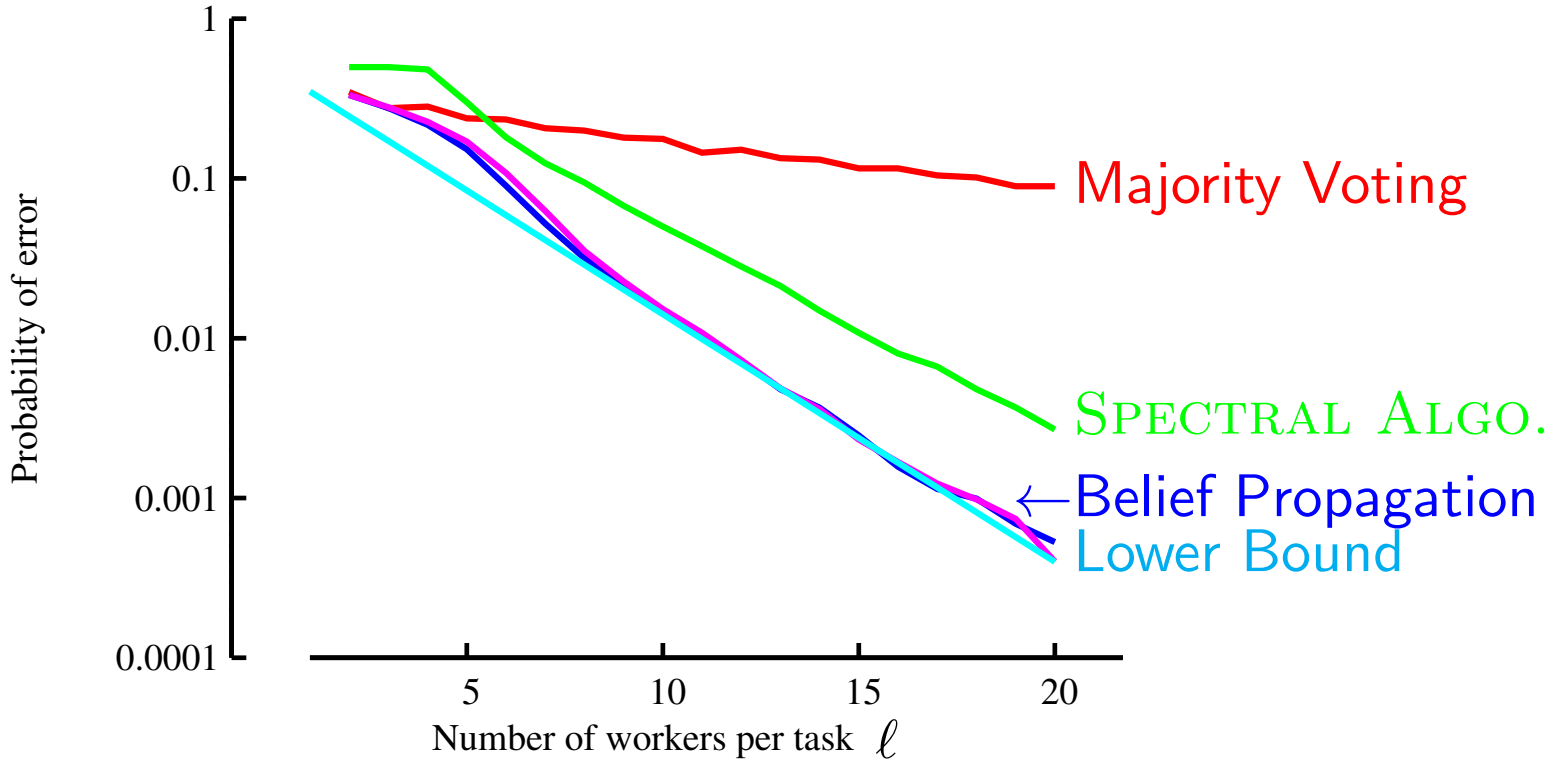
Tensor Decomposition + EM [ZhangChenZhouJordan14]

For general *k*-ary classification and general random sampling,

$$P_{\text{err}}^{\text{EM}} \leq e^{-C\sigma^2\ell}$$

Why use only spectral?

Two 2-phase approaches



Spectral + BP [OkOhShinYi16]

Belief propagation is near-optimal

$$P_{\text{err}}^{\text{BP}} - P_{\text{err}}^{\text{LB}} \leq m^{-\gamma} \qquad \text{for some } \gamma > 0$$

Modeling/Inference

[DawidSkene1979][SmythFayyadBurl1995]
[JinGhahramani03][ShengProvostIpeirotis08]
[WhitehillWuBergsmaMovellanRuvolo09]
[WelinderBransonPeronaBelongie10]
[WauthierJordan11][GhoshKaleMcAfee11]
[ErtekinHirshRudin11]**[KargerOhShah11,13]**
[LiuPengIhler12][ZhouPlattBasuMao12]
[DalviDasguptaKumarRastogi13][GaoZhou13]
[LiYu14][ZhangChenZhouJordan14]
[LeeKimLeeJung15]
[ShahBalakrishnanWainwright16]
[OkOhShinYi16][BonaldCombex16]

Online/Active Learning

[DonmezCarbonellSchneider09]
[ZhengScottDeng10][YanRosalesFungDy11]
[HoJabbariVaughan13][ChenLinZhou15]
[ChenAgarwalWiermanBarmanAndrew15]
[LiuLiu15]**[KhetanOh16]**[BonaldCombes16]
[MassoulieXu16][JunJamiesonNowakZhu16]

Incentive Mechanism

[DiPalantinoVojnovic09]
[ArchakSundararajan09][MasonWatts09]
[SingerMittal11,13]
[DuanKuboSugiyamaHuangHasegawaWalrand12][ZhangVanDerSchaar12]
[YangXueFangTang12][AzarChenMicali12]
[Koutsopoulo13][SinglaKrause13]
[ZhangYangZhouCaiChenLi14]
[ChawlaHartlineSivan15]
[ZhangXueYuYangTang15][ShahZhou15,16]
[ZhaoLiMa17]

Complex Tasks

[BernsteinBrandtMillerKarger11]
[MarcusKargerMaddenMillerOh12]
[WuFanYu12][PengQiangIhlerBerger13]
[SuDengFei-Fei12]
[LeeSteyversYoungDeMiller16]
[OkOhShinJangYi17]

Systems/Database

Collaborators on Matrix Completion and Crowdsourcing

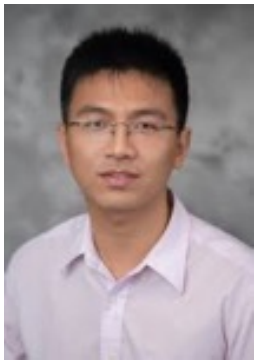

Andrea Montanari


Raghunandan Keshavan


Devavrat Shah


David R. Karger


Prateek Jain


Jiaming Xu


Ashish Khetan


Kiran Thekumparampil


Jungseul Ok


Jinwoo Shin


Yung Yi

Experiments using Amazon Mechanical Turk

- Learning similarities using Crowdsourcing
  - recommendations
  - searching

You wanted to get the tie on the top, but it was not available. Which one would be a better substitute?

- ☐ the tie the left
- ☑ the tie on the right

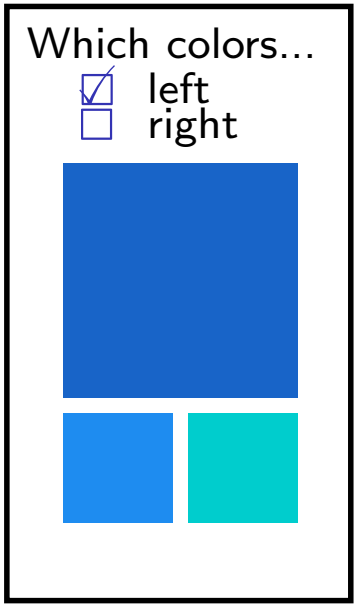
Experiments using Amazon Mechanical Turk

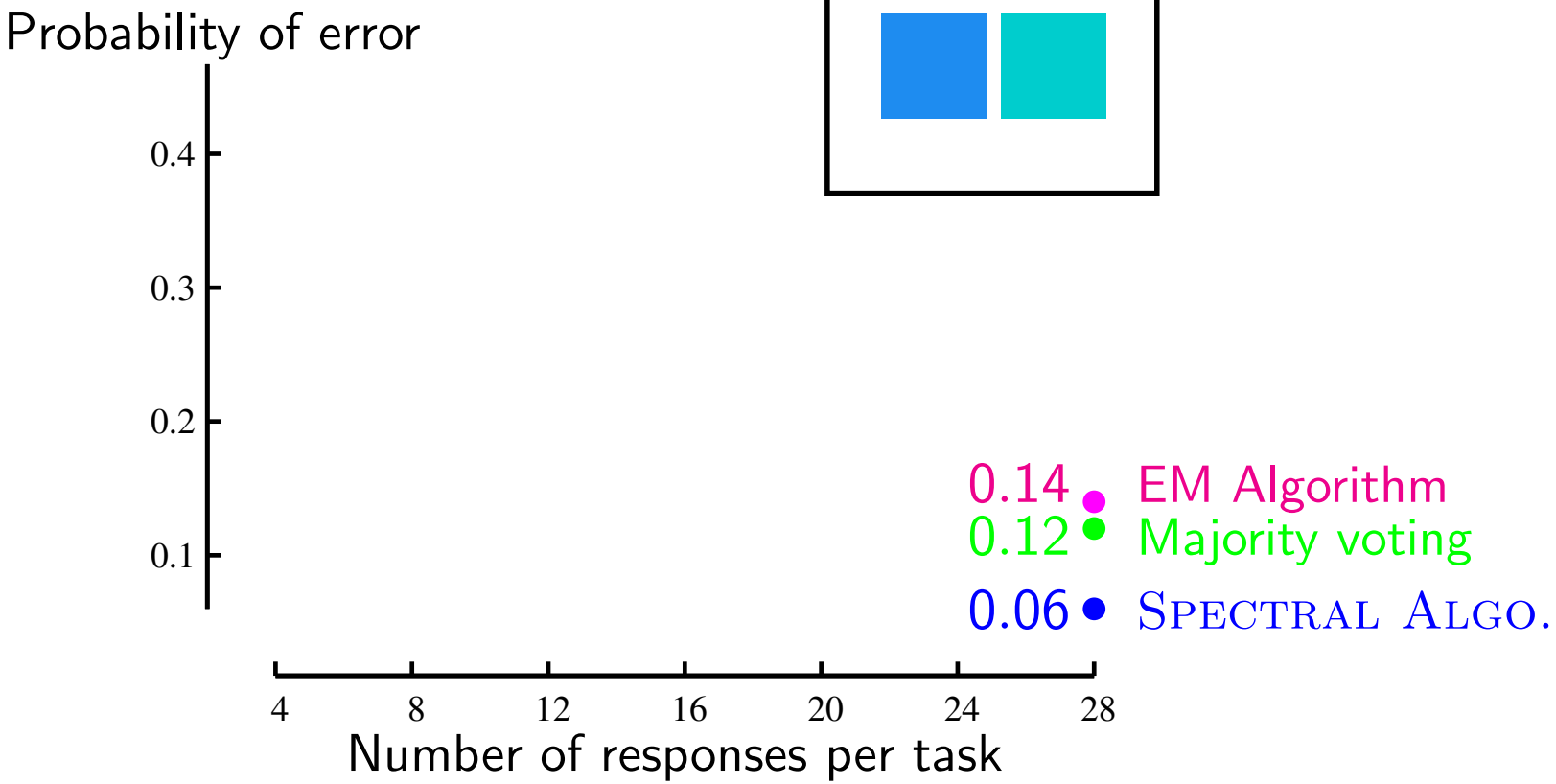
- Learning similarities using Crowdsourcing
  - recommendations
  - searching

Which tie matches my shirt better?

- ☐ the tie on the left
- ☑ the tie on the right

Experiments using Amazon Mechanical Turk





Experiments using Amazon Mechanical Turk