ACM SIGMETRICS
special interest group on performance evaluation

# Guiding the Recommender: Information-Aware Auto-Bidding for Content Promotion

**Yumou Liu, Zhenzhe Zheng, Jiang Rong, Yao Hu, Fan Wu, Guihai Chen**

Shanghai Jiao Tong University, Xiaohongshu

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

小红书

# Outline

- **Background & Motivation**
- Problem & Challenges
- Methods
- Theoretical Analysis
- Experimental Evaluation

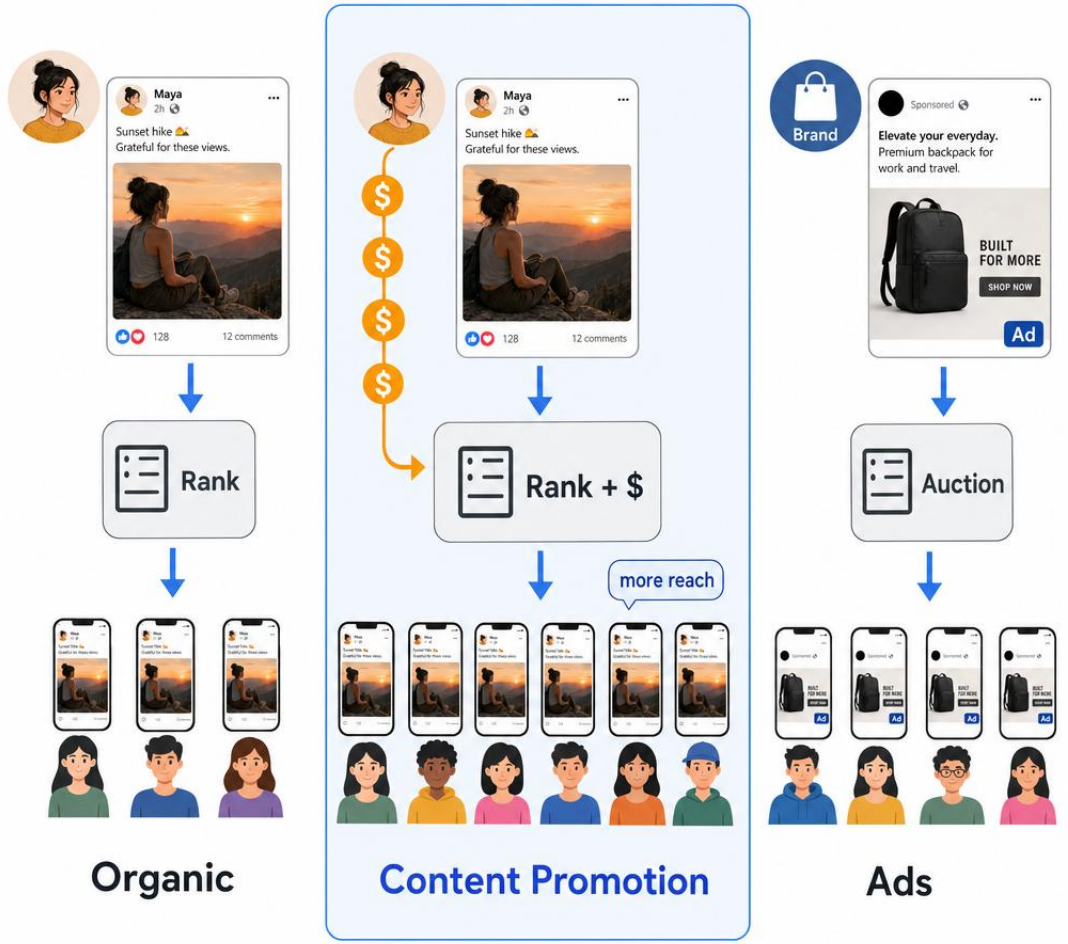
# Online Content Platforms

- Facts:
  - **Million-level** new notes are posted daily in Xiaohongshu (RedNote).
  - Most of the notes receive near zero impressions **after 24 hours**.

- Cold-Start Matters
  - Platform only guarantees 100 impressions for each new note.
  - Recommendation model learns from the impressions.
    - Determines if there are more impressions later.

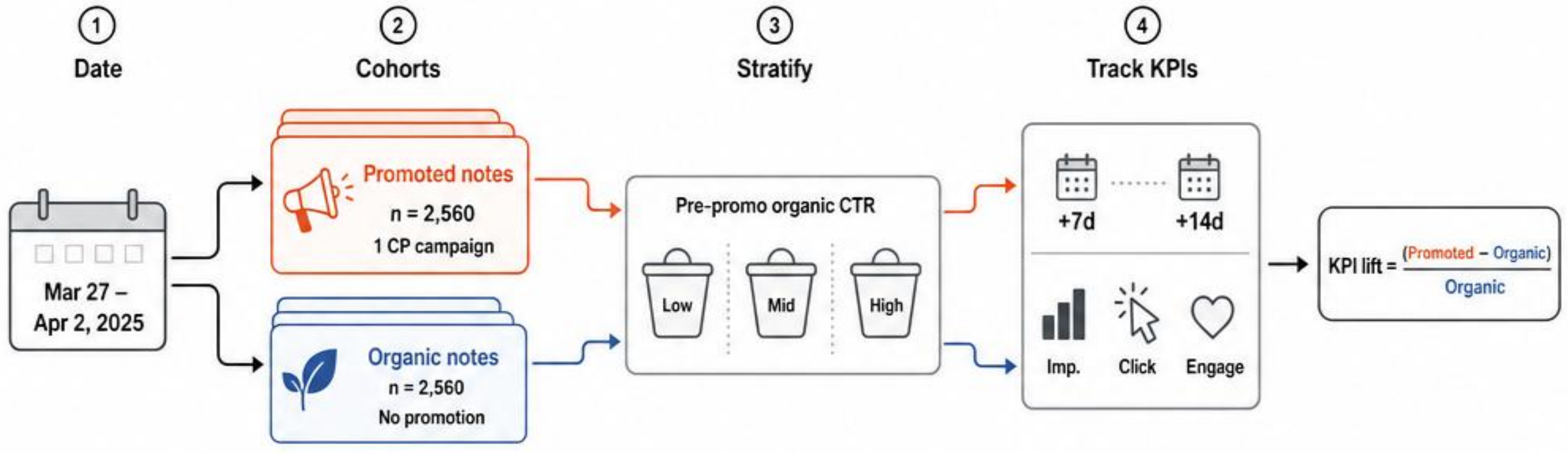
# Content Promotion

- Allows creators to pay for additional impressions.
- Money enters the ranking process.
- Techniques in online advertising are inherited.
  - Auctions.
  - Auto-Bidding.



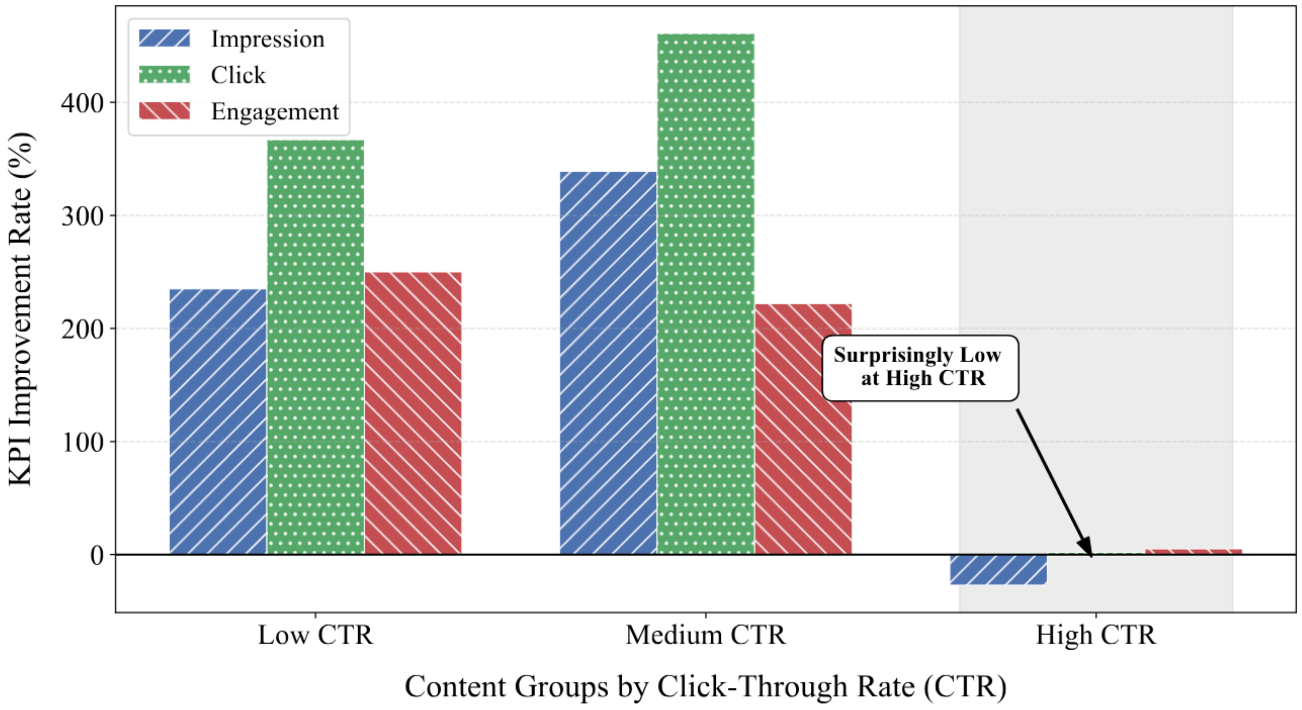
# Empirical Finding - Setup

- *Does ad techniques work well in content Promotion?*
- Study design: 2,560 promoted posts vs. 2,560 organic posts.
  - Content-similar notes selected *manually*.
- Stratified by pre-promotion organic CTR.
- Tracked KPIs at 7 days post-campaign.
  - Impressions, Clicks, Engagements.



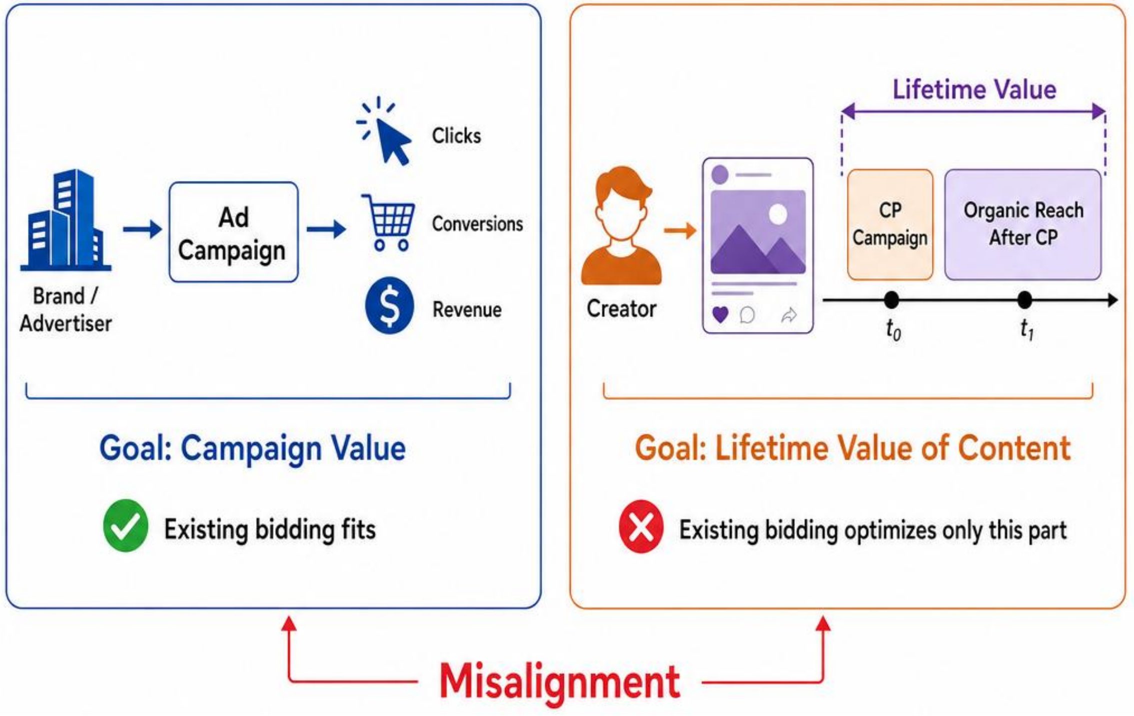
# Empirical Finding - Insight

- Striking dichotomy revealed.
  - Low&Medium-CTR bucket: +200% engagement lift, promotion helps.
  - High-CTR bucket: ~-25% decline in impressions, **promotion harms**.



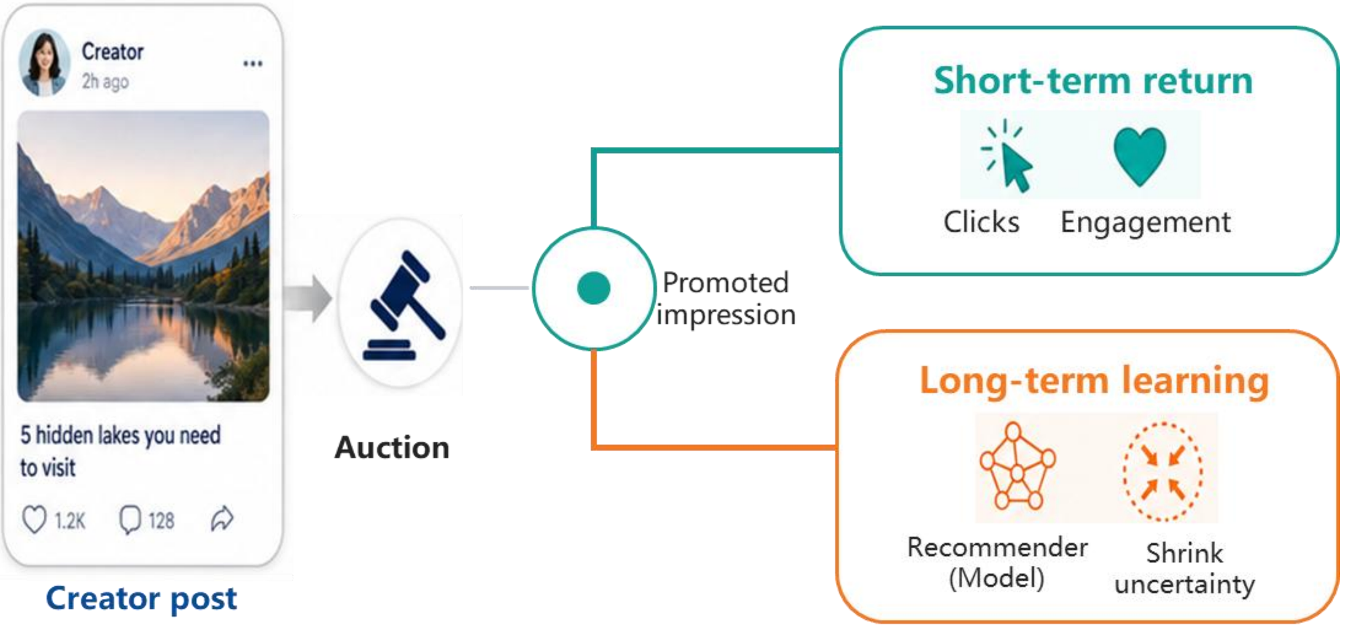
# Why existing bidding algorithms fail in CP?

- Different goals between advertisers and content creators.
- Misalignment between the objective of creators and existing bidding algorithms.



# Idea: Bidding for the Recommender

- Intuition: Bid to buy informative impressions, not only more impressions.




- The long-term goal should be **minimizing the uncertainty** of the recommender.
  - So that a high-quality note will not be submerged.

# Outline

- Background & Motivation
- **Problem & Challenges**
- Methods
- Theoretical Analysis
- Experimental Evaluation

# Model & Formulation

- Creator $i \in \mathcal{S}$ with content note features $\boldsymbol{x_i}$.
- Platform pCTR model $\mathcal{M}$: predicts $\hat{\sigma}_i = \mathcal{M}(\boldsymbol{x_i})$; true CTR $\boldsymbol{\sigma_i}$
- Auction: First price, pay-per-impression, rank by eCPM $\hat{\sigma}_i \cdot b_i$
- Creator's private value $v_i$ per click.
- Dual-objective under budget $B_i$

$$\max_{\mathcal{S}} F(\mathcal{S}) = \beta \cdot V(\mathcal{S}) + (1-\beta) \cdot U(\mathcal{S})$$

$$s.t. \sum_{t \in \mathcal{S}} b_t \leq B_i$$

- where $V(\mathcal{S}) = \sum_{\boldsymbol{z} \in \mathcal{S}} \hat{\sigma}(\boldsymbol{z})$ (total expected clicks) is the short-term objective.
- $U(\mathcal{S})$ is the long-term objective.

# What is the long-term objective?

- A toy model:
  - A creator updates its creating strategy $x_t$ randomly, and keeps it if better.
  - It receives a random reward $R$ in each round with variance $\xi^2$.

- Theorem 5

$$\frac{1}{T}\sum_{t=0}^{T=1} \mathbb{E}[||\nabla R(x_t)||^2] \leq \mathcal{O}\left(\frac{\xi \cdot d \cdot \Delta R}{r^2 T}\right) + \mathcal{O}(L \cdot \xi \cdot d)$$

- Reducing **model uncertainty** -> reducing reward noise ($\xi$) -> better creator performance.

# Quantifying Uncertainty: Optimal Experimental Design

- Fisher Information Matrix (FIM) quantifies informativeness of observations:

$$\mathcal{J}(\mathcal{S}) = \sum_{z \in \mathcal{S}} g_\theta(x_z, y_z) g_\theta(x_z, y_z)^\top$$

- Classical OED criteria
  - D-optimal: $\max \det(\mathcal{J})$, minimize confidence ellipsoid volume
  - A-optimal: $\min tr(\mathcal{J}^{-1})$, minimize average parameter variance
  - I-optimal: $\min \int w(x) g(x)^\top \mathcal{J}^{-1} g(x) dx$, minimize prediction variance over the region of interest

- Problem: All require $\mathcal{J}^{-1}$ -- non-decomposable, $O(d^3)$ per update.

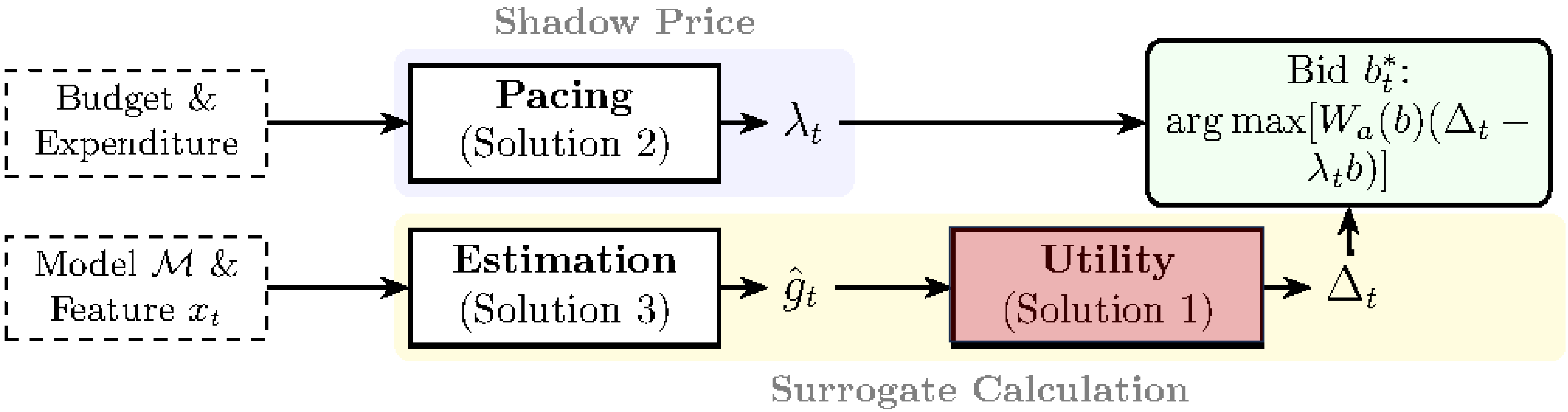
# Technical Challenges

- C1: **Tractable** long-term objective.
  - Avoid costly matrix operations for millisecond bidding latency.
- C2: Practical Budget management under uncertainty
  - Sequential requests, irrevocable decisions, no foresight.
- C3: Missing label at bid time.
  - Engagement happens after bidding.

# Outline

- Background & Motivation
- Problem & Challenges
- **Methods**
- Theoretical Analysis
- Experimental Evaluation

# Solution Overview



# Component 1: Gradient Coverage

- Recall C1: *Tractable long-term objective (without matrix inverse).*
- Key idea: max gradient similarity between request and a validation set

$$U_\lambda(\mathcal{S}) = \sum_{\boldsymbol{x} \in \mathcal{D}_{val}} \max_{\boldsymbol{z} \in \mathcal{S}} \exp\left(-\lambda \left\| g_{\theta_0}(\boldsymbol{x}) - g_{\theta_0}(\boldsymbol{z}) \right\|^2\right)$$

- Intuition breakdown:
  - $g_{\theta_0}(\boldsymbol{x})$ : loss gradient of validation sample
  - $\exp(-\lambda\|\cdot\|^2)$: Gaussian kernel measuring gradient similarity
- Facility-location-style objective.
- Computational cost: $\tilde{O}(|\mathcal{D}_{val}|)$ per impression, no matrix inversion.

# Theoretical Justification

- Theorem 1 (Regularized Fisher-Coverage Relationship)
  - Let $G_\gamma(\mathcal{S}) = \sum_{\boldsymbol{x}\in\mathcal{D}_{val}} g(\boldsymbol{x})^\top \mathcal{J}_\gamma(\mathcal{S})^{-1} g(\boldsymbol{x})$ be the regularized total uncertainty.
    - $\mathcal{J}_\gamma = \sum_{z\in\mathcal{S}} g(z)g(z)^\top + \gamma I$
  - The increasing $U_\lambda(\mathcal{S})$ tightens the upper bound on $G_\gamma(\mathcal{S})$.

$$G_\gamma(\mathcal{S}) \leq \tilde{O}\left(-U_\lambda(\mathcal{S})\right)$$

- Minimizing $G_\gamma(\mathcal{S})$ is I-optimal design.
- Implication: $U(\mathcal{S})$ is a monotone surrogate for I-optimal design.

# Proof Sketch of Theorem 1

- **Step 1 (Nearest-neighborhood bound):** For each $x \in \mathcal{D}_{val}$, let $z \in \mathcal{S}$ be its closest gradient neighbor with distance $d_x = \|g(x) - g(z_x)\|^2$. By PSD ordering:
$$\mathcal{J}(\mathcal{S})^{-1} \preccurlyeq (\gamma I_d + g(z_x)g(z_x)^\top)^{-1}.$$
- **Step 2 (Sherman-Morrison inversion):** Apply the rank-1 update formula:
$$g(x)^\top \mathcal{J}_\gamma^{-1} g(x) \le \frac{\|g(x)\|^2}{\gamma} - \frac{1}{\gamma^2} \frac{\langle g(x), g(z_x)\rangle^2}{1 + \frac{\|g(z_x)\|^2}{\gamma}}.$$

# Proof Sketch of Theorem 1 (continued)

- **Step 3 (Relate inner product to distance):** Using $\|g(x) - g(z_x)\|^2 = \|g(x)\|^2 + \|g(z_x)\|^2 - 2\langle g(x), g(z_x)\rangle$, and bounds $\|g\| \leq L$, $\|g(z_x)\| \geq m$:

$$On\ A_\tau = \{x: d_x \leq \tau\}: \langle g(x), g(x_z)\rangle^2 \geq \frac{\left(2m^2 - \tau\right)^2}{4}$$

- **Step 4 (Sum and connect to $U_\lambda$):** Summing over $\mathcal{D}_{val}$ and counting $|\mathcal{A}_\tau|$ via the kernel $U_\lambda \leq |A_\tau| + (k - |A_\tau|)e^{-\lambda\tau}$, so$|A_\tau| \geq \frac{U_\lambda - ke^{-\lambda\tau}}{1 - e^{-\lambda\tau}}$. Substituting yields the bound.

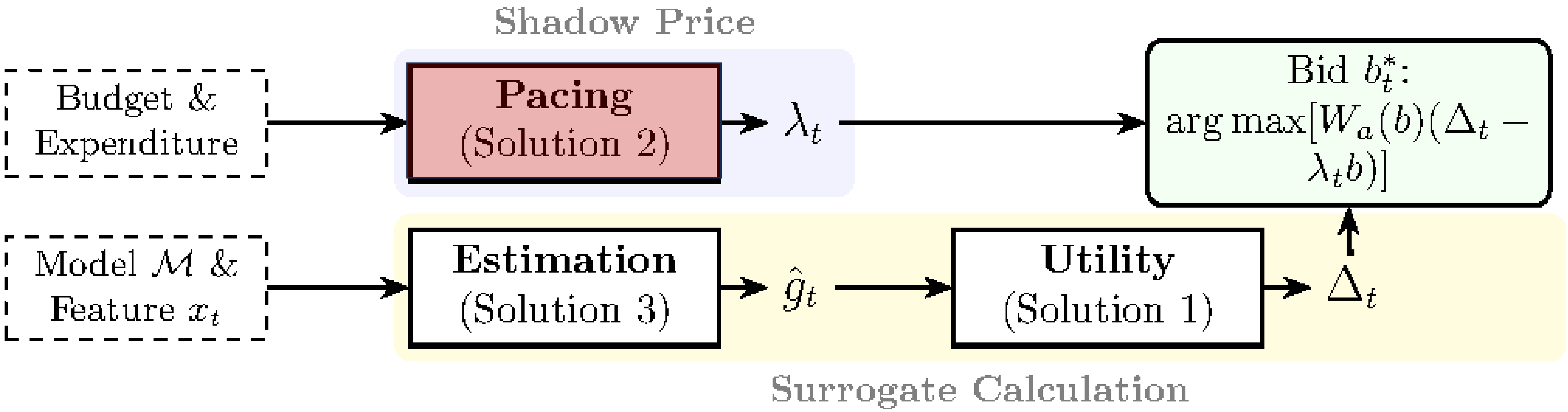
# Properties: Submodularity

- Theorem 2: $F(\mathcal{S})$ is monotone submodular.
  - $F(\mathcal{S}) = \beta \cdot V(\mathcal{S}) + (1-\beta) \cdot U(\mathcal{S})$

  - For $U(\mathcal{S})$: a novel gradient -> much coverage when $\mathcal{S}$ is small; little when $\mathcal{S}$ is large.
  - For $V(\mathcal{S})$: additive (trivially submodular)

- Why it matters?
  - Budget-constrained maximization admits $(1-1/e)$ greedy approximation.
  - Enable per-impression marginal utility decomposition
    - $\Delta_t = F(\mathcal{S}_{t-1} \cup \{x_t\}) - F(\mathcal{S}_{t-1})$

# Solution Overview



# Component 2: Two-Stage Bidding

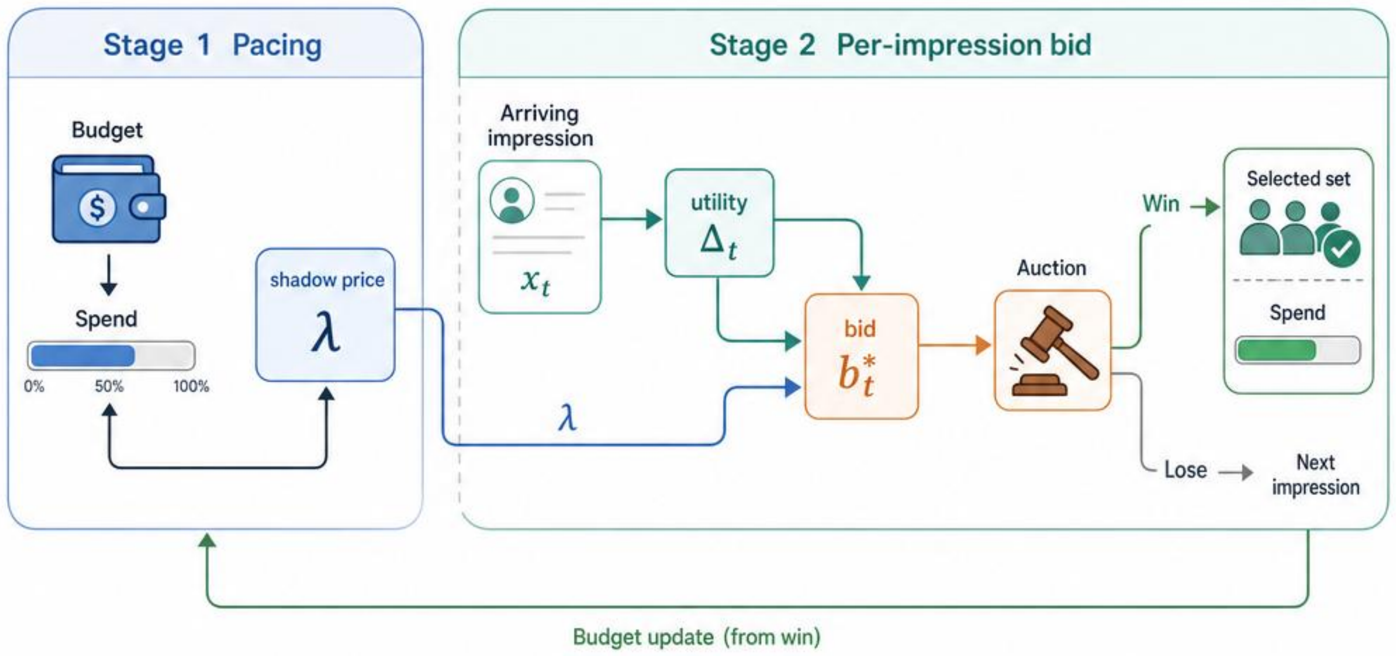
- **Lagrangian relaxation** of budget constraint

$$\mathcal{L}(\mathcal{S},\lambda) = F(\mathcal{S}) - \lambda\left(\sum_{t\in\mathcal{S}} b_t - B_i\right)$$

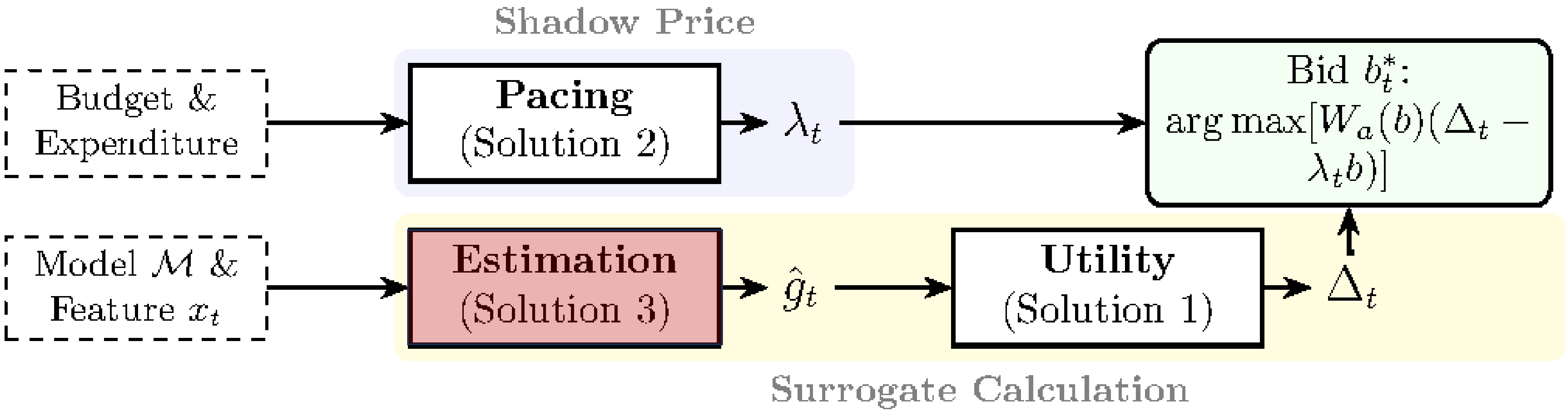
- Stage 1: **Campaign-Level Pacing** (Shadow Price $\lambda$)
  - $\lambda_k = \lambda_{k-1} \cdot \exp\left(\eta \cdot \frac{Cost_{k-1} - Paced\_Budget_k}{B_i}\right)$
  - Linear pacing schedule: $Paced\_Budget_k = B \cdot k/K$.
- Intuition
  - Overspend→ $\lambda$ increases→bids become more conservative.
  - Underspend→ $\lambda$ decreases→bids become more aggressive.

# Two-Stage Bidding (Continued)

- Stage 2: **Impression-Level Bid Optimization**
  - For each auction, compute marginal utility $\Delta_t$ and solve:
$$b_t^* = \text{argmax}_{b_t \geq 0}[W_a(b_t) \cdot (\Delta_t - \lambda b_t)]$$
  - $W_a(b_t)$ is the winning probability.



# Solution Overview



# Component 3: Missing Labels at Bid Time

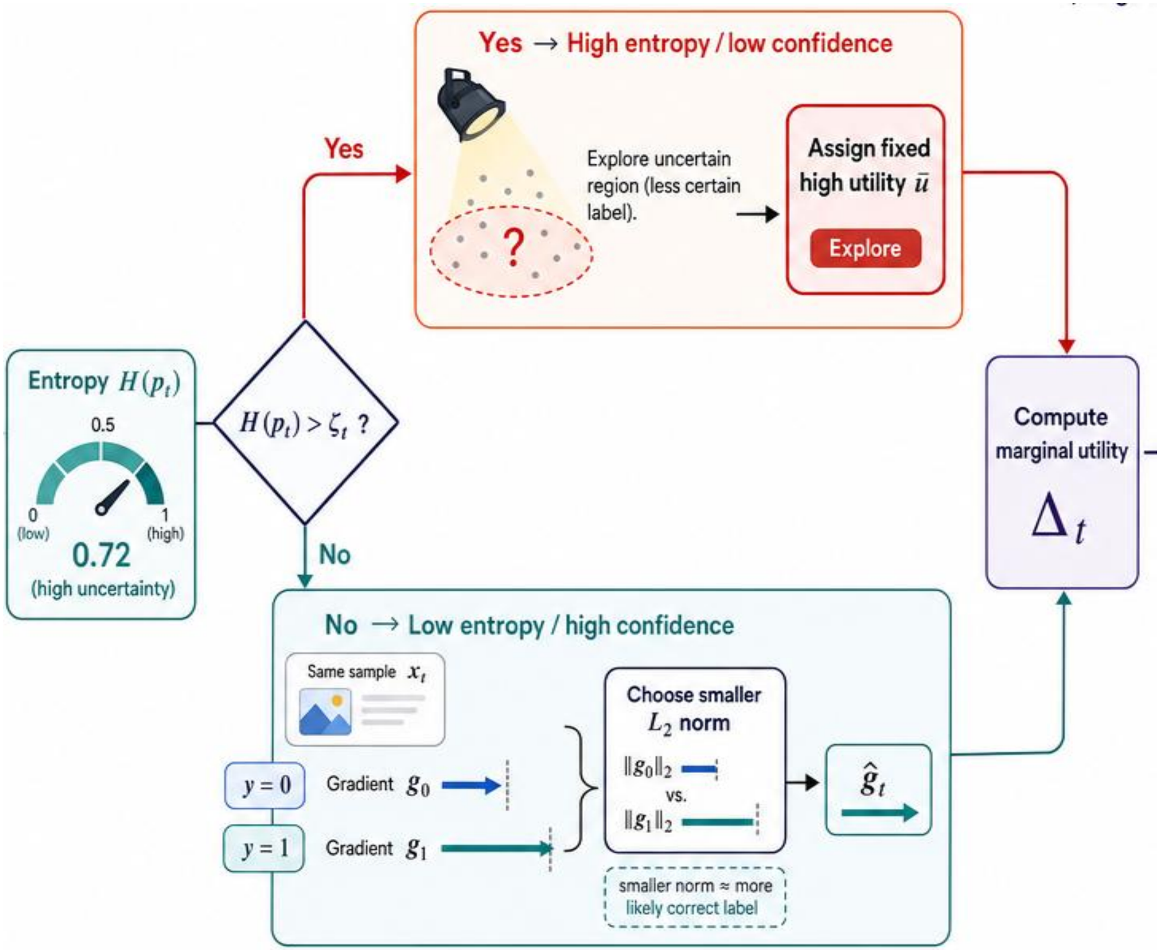
- Recall: $\Delta_t$ depends on $\boldsymbol{g}(\boldsymbol{x}_t, \boldsymbol{y}_t)$ -- the loss gradient
  - $\boldsymbol{y}_t$ is **unknown at bidding time**.

- Naïve approaches that fail:
  - Expected gradient $\mathbb{E}_y[g]$ under predicted distribution: collapses in confident-but-wrong regime.
  - Random guess: no better than chance.

- Need a robust heuristic that works without labels.

# Confidence-Gated Gradient Estimation

- Core idea: Trust the trained model and use its prediction entropy.



# Extension: Zeroth-Order Gradient Estimation

- Motivation:
  - Computing gradient at bid time is not scalable/feasible for online service.
- Two-point ZO estimator

$$\hat{g}_{ZO} \approx \frac{1}{2\mu}[\mathcal{L}(\theta + \mu\boldsymbol{u}) - \mathcal{L}(\theta - \mu\boldsymbol{u})]\boldsymbol{u}$$

- $u \sim Unif\left(\mathbb{S}^{d-1}\right)$
- Trade-off: ZO introduces variance but preserves effectiveness.
- Enables information-aware bidding for any model architecture.

# Outline

- Background & Motivation
- Problem & Challenges
- Methods
- **Theoretical Analysis**
- Experimental Evaluation

# Regret Bound

- **Theorem 3 (Regret Bound for FP CPM Dual Pacing)**
  - Assumptions:
    - $b_t \in [0, b_{max}]$, $\|\Delta_t\| \le \Delta_{max}$, $\lambda_t \in (0, \lambda_{max}]$
  - Algorithm's total expected utility, by choosing $\eta = \frac{1}{C}\sqrt{\frac{\log\left(\frac{\lambda_{max}}{\lambda_0}\right)}{T}}$,

$$ALG \ge OPT - O(C\sqrt{T \log(\lambda_{\max}/\lambda_0)})$$

    - $ALG := \sum_{t=1}^{T} \mathbb{E}[W_t(b_t)\Delta_t]$, $OPT := \max_{\mathcal{S}^*, \text{Spend}(\mathcal{S}^*) \le B} \sum_{t=1}^{T} \mathbb{E}[\Delta_t^{(\mathcal{S}^*)}]$

  - Sublinear regret in number of auctions $T$.
  - Proves that online dual-variable pacing is near-optimal against offline opt.

# Proof Sketch of Theorem 3

- **Step 1 (Define per-round loss):** let $\ell_t(\lambda) = -f_t(\lambda)$ where $f_t(\lambda) = \max_b W_t(b)(\Delta_t - \lambda b)$. $\ell_t$ is convex; its subgradient: $\partial\ell_t(\lambda_{t-1}) \ni h_t/C$, where $h_t = W_t(b_t)b_t$ is the expected spend.
- **Step 2 (Mirror descent regret):** The update

$$\lambda_t = \lambda_{t-1}\exp(\eta h_t/C)$$

is OMD with negative entropy regularizer.

$$\sum_{t=1}^{T} f_t(\lambda_{t-1}) \geq \sum_{t=1}^{T} f_t(\lambda^*) - \frac{\log\left(\frac{\lambda_{max}}{\lambda_0}\right)}{\eta} - \frac{\eta T}{2}$$

# Proof Sketch of Theorem 3 (Continued)

- **Step 3 (Relate utility to dual obj):** By envelope theorem: $f_t(\lambda_{t-1}) = W_t(b_t)(\Delta_t - \lambda_{t-1}b_t)$. Therefore
$$\sum_t W_t(b_t)\Delta_t = \sum_t f_t(\lambda_{t-1}) + \sum_t \lambda_{t-1}h_t\,.$$
Another MWU bound gives
$$\sum_t \lambda_{t-1}h_t \geq \lambda^* \sum_t h_t - O\left(\frac{1}{\eta} + \eta T C^2\right).$$
- **Step 4 (Combine with weak duality):** $OPT \leq \sum_t f_t(\lambda^*) + \lambda^* B$. Setting $\eta = \frac{1}{C}\sqrt{\frac{\log\left(\frac{\lambda_{max}}{\lambda_0}\right)}{T}}$ yields the $O(C\sqrt{T})$ regret.

# Budget Feasibility

- **Theorem 4 (Budget Feasibility Guarantee)**

$$\mathbb{E}\left[\sum_{t=1}^{T} b_t \cdot 1_{win_t}\right] \leq B + \frac{\log(\lambda_{max} / \lambda_0)}{\eta}$$

- The overshoot term can be made small by tuning $\eta$.

# Proof Sketch of Theorem 4

- **Step 1 (Telescoping the dual update):** The multiplicative weights rule is $\lambda_t = \lambda_{t-1} \exp\left(\eta \frac{W_a(b_t)b_t}{B}\right)$. Taking logs and telescoping from $t = 1 \dots T$:

$$\log \lambda_T - \log \lambda_0 = \frac{\eta}{B} \sum_{t=1}^{T} W_a(b_t) b_t \, .$$

- **Step 2 (Rearrange and bound $\lambda_T$):** Since $\lambda_T \leq \lambda_{max}$

$$\sum_{t=1}^{T} W_a(b_t) b_t = \frac{B}{\eta} \log \frac{\lambda_T}{\lambda_0} \leq \frac{B}{\eta} \log \frac{\lambda_{max}}{\lambda_0} .$$

# Proof Sketch of Theorem 4 (Continued)

- **Step 3 (Take expectation):** By the law of total expectation: $\mathbb{E}\left[\mathrm{b_t} \cdot 1_{win_t} \middle| b_t\right] = \mathrm{W_a}(\mathrm{b_t})\mathrm{b_t}$. Therefore:

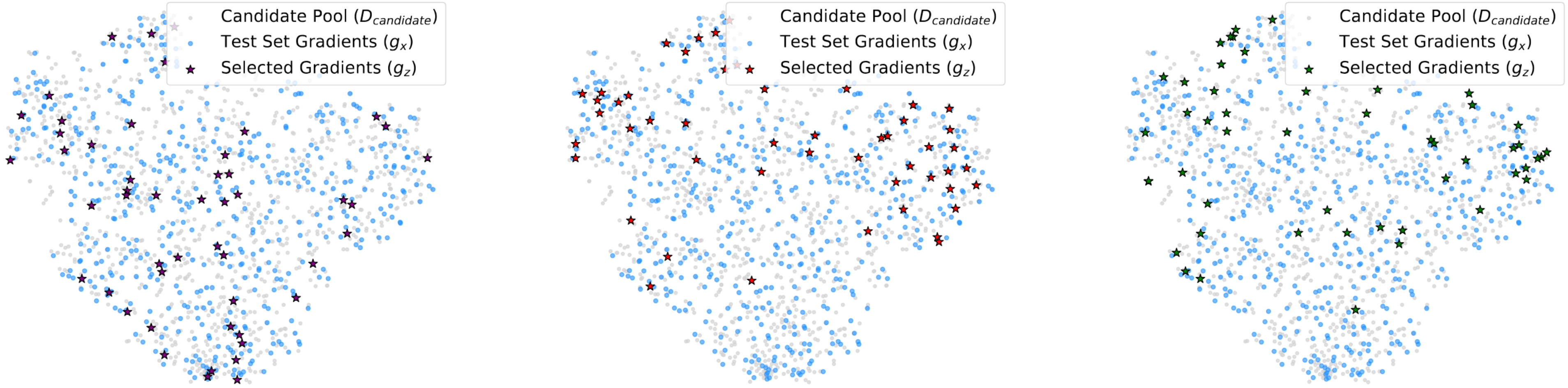
$$\mathbb{E}\left[\sum_{t=1}^{T} b_t \cdot 1_{win_t}\right] = \sum_{t=1}^{T} \mathbb{E}[W_a(b_t)b_t] \leq B + \frac{\log\left(\frac{\lambda_{max}}{\lambda_0}\right)}{\eta}.$$

The $\frac{\log\left(\frac{\lambda_{max}}{\lambda_0}\right)}{\eta}$ term is the "cost of online learning", controlled by $\eta$ and vanishing as $\eta \to \infty$ (but larger $\eta$ increases instability).

# Outline

- Background & Motivation
- Problem & Challenges
- Methods
- Theoretical Analysis
- **Experimental Evaluation**

# Exp. 1: Surrogate Relationship Validation

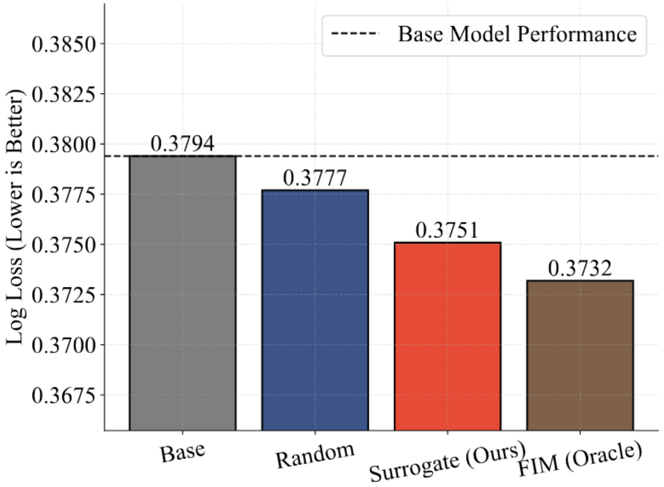



(a) Random. (b) Greedy-Surrogate (Proposed). (c) Greedy-FIM (Oracle).

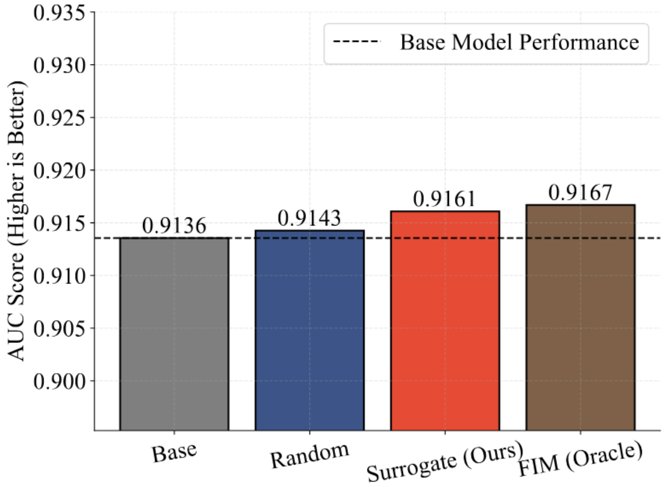
- Setup: synthetic logistic regression
- Results: t-SNE visualization of selected gradients
  - Random: clustered in dense regions, poor coverage
  - Greedy-Surrogate: diverse, well-distributed — visually matches Oracle

# Exp. 1: Surrogate Relationship Validation

- Setup: synthetic logistic regression




(a) Log Loss.

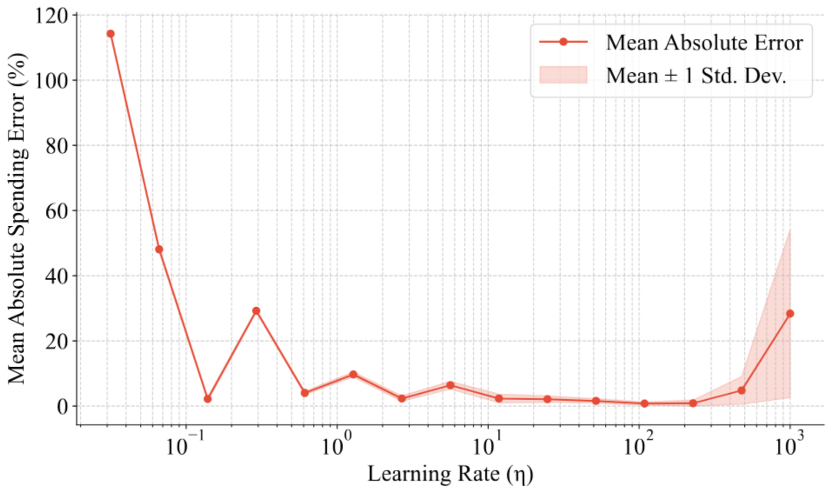



(b) AUC.

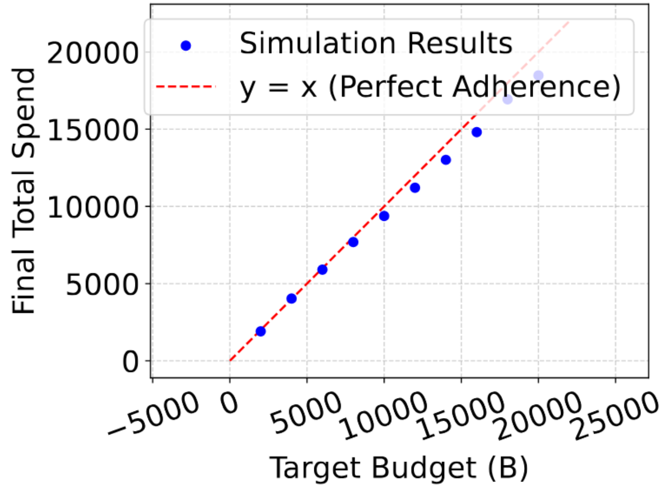
- Results: **continual training** on selected gradients
  - Greedy-Surrogate: Performance close to FIM oracle.
  - Both significantly outperform random baseline.

# Exp 2: Budget Feasibility

- Simulated auction stream ($T = 5000$), varying pacing rate $\eta$ and budget $B$.




(a) Mean-absolute error v.s. $\eta$.

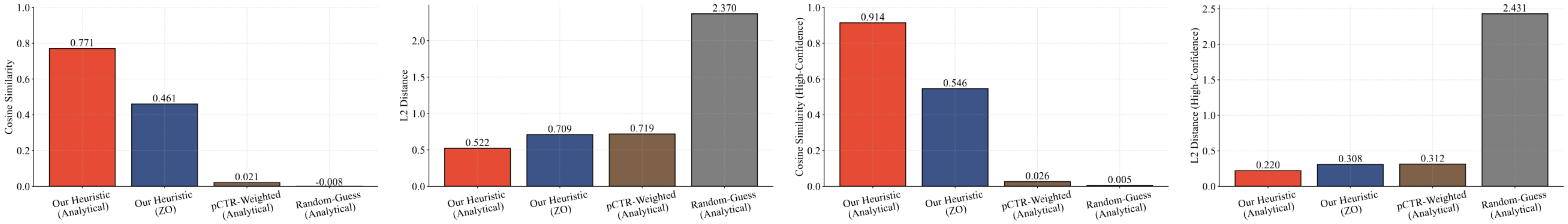



(b) Budget spend wrt. budget.

- Findings:
  - (a) Budget pacing rate matters, and optimal pacing rate exists.
  - (b) Budget feasibility holds under two-stage bidding with appropriate $\eta$.

# Exp 3: Gradient Estimation Accuracy

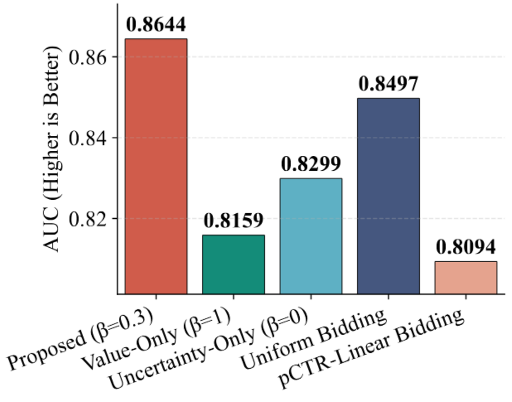
- Setup: Pretrained LR model, compare proxy gradients to ground truth.
- Methods: Our heuristic (Analytical & ZO), pCTR-weighted, random-guess
- Results
  - Our heuristics show smallest distances to the ground truth.
  - Our heuristic is robust even with noisy ZO gradients.



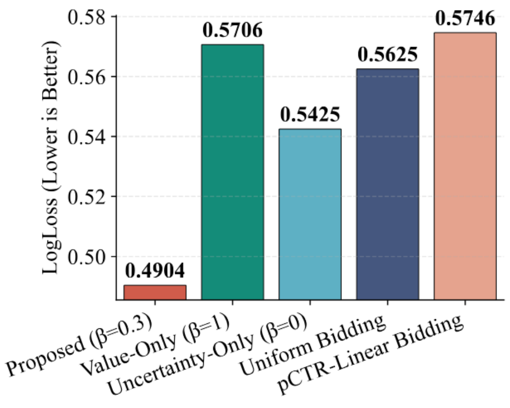

(a) Cosine Similarity on all samples. (b) L-2 distance on all samples. (c) Cosine Similarity on high-confidence samples. (d) L-2 distance on high-confidence samples.

# Exp 4: End-to-End Offline Simulation

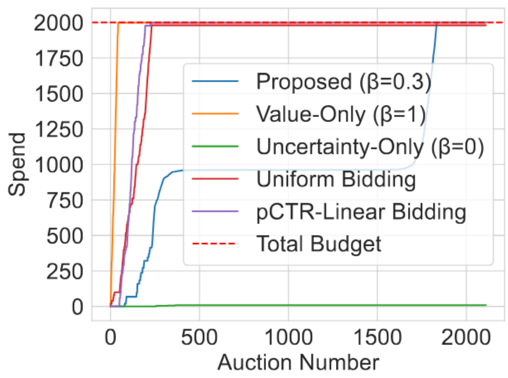
- Setup
  - 2000 sequential auctions;
  - MLP pCTR model;
  - First-price auctions;
  - Five competing bidders.
- Results
  - Best AUC and lowest Logloss.
  - Smooth spending curves.
  - Self-regulating $\lambda$.



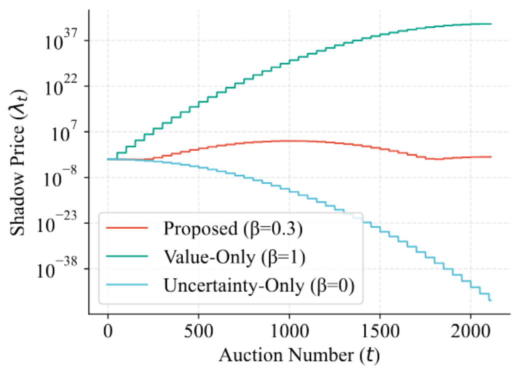

(a) AUC.




(b) Log loss.




(c) Spending vs. time.




(d) Dynamics of dual variable ($\lambda$).

# Takeaways

- **Naive promotion can backfire** for high-quality content.
- **Reframe:** paid promotion as strategic data acquisition.
- **Gradient coverage** is a principled, decomposable surrogate with formal link to I-optimal design.
- **Two-stage bidding** (shadow price pacing + marginal utility optimization) achieves sublinear regret and budget feasibility.
- **Confidence-gated heuristic** solves the missing-label problem; ZO variant enables black-box deployment.

# Limitation and Future Work

- Limitations
  - Requires model gradient (or accept ZO gradient errors).
  - Hyperparameters needed.
  - Fixed validation set required.
- Future Directions
  - Adaptive validation set updating during compaign

# Thank you!

Yumou Liu

liuyumou@sjtu.edu.cn

# More about the empirical findings (I)

- Setup:
  - Sampling period: March 27 – April 2, 2025.
  - Groups: 2,560 promoted notes vs. 2,560 organic notes.
  - Stratification: By **pre-promotion organic CTR** (measured before campaign start)
  - Follow-up: KPIs tracked at **7 days** and **14 days** post-campaign
  - KPI metric: $Improve\_Rate = \frac{KPI_{Promoted} - KPI_{organic}}{KPI_{organic}} \times 100\%$
  - Tracked KPI: cumulative impressions, clicks, engagements.

# More about the empirical findings (II)

- Sampling Criteria of Promoted Notes
  - Promotion history: Promoted **exactly once** via Shutiao; no other competitive bidding ad products used
  - Campaign budget: Single Shutiao campaign expenditure **≥ 30 units**
  - Pre-promotion activity: In-feed clicks **> 0** (not ignored by organic system)
  - Pre-promotion ceiling: Impressions **< 50,000**, Clicks **< 10,000**, Engagements **< 2,000**
  - Account / content type: **Not** enterprise account, product-centric post, or e-commerce affiliate link

# More about the empirical findings (III)

- Sampling Criteria of Organic Notes
  - Promotion history: **Never** promoted via Shutiao or any competitive bidding ad product
  - Pre-sampling activity: In-feed clicks **> 0**
  - Pre-sampling ceiling: Impressions **< 50,000**, Clicks **< 10,000**, Engagements **< 2,000**
  - Account / content type: **Not** enterprise account, product-centric post, or e-commerce affiliate link

# Background: Optimal Experimental Design

- Fisher Information Matrix (FIM) quantifies informativeness of observations:

$$\mathcal{J}(\mathcal{S}) = \sum_{z\in\mathcal{S}} g_\theta(x_z, y_z) g_\theta(x_z, y_z)^\top$$

- Classical OED criteria
  - D-optimal: $\max \det(\mathcal{J})$, minimize confidence ellipsoid volume
  - A-optimal: $\min tr(\mathcal{J}^{-1})$, minimize average parameter variance
  - I-optimal: $\min \int w(x) g(x)^\top \mathcal{J}^{-1} g(x) dx$, minimize prediction variance over the region of interest

- Problem: All require $\mathcal{J}^{-1}$ -- non-decomposable, $O(d^3)$ per update.

# Complexity & Deploy Tradeoff

| Method | Cost (per-impression) | Matrix Ops | Decomposable? | Real-Time? |
|---|---|---|---|---|
| FIM-based | $O(d^2)$ to $O(d^3)$ | Inversion/rk-1 update | x | x |
| Active-learning | $O(d^2)$ | Distant computations | x | x |
| Grad Coverage | $O(\|\mathcal{D}_{val}\|)$ | Kernel evaluations + max | ✓ | ✓ |

- Why $O(|\mathcal{D}_{val}|)$ is practical?
  - $|\mathcal{D}_{val}|$ is a **design choice**: a few hundred representative samples suffice.
  - **Cached partial maxima:** maintain running maximum per validation point; each new impression only compares against current max.
  - **Batching:** can batch multiple impressions and update utility jointly.

# Why we need submodularity?

- Justifies the Greedy Per-Impression Decision
  - Decompose the global problem into per-auction decisions using Lagrangian
$$b_t^* = \arg\max_b W(b) \cdot (\Delta_t - \lambda b)$$
  - This is effectively a greedy rule.

  - Submodularity guarantees that greedy selection achieves a known approximation ratio.
  - Otherwise $D(\lambda) = \max_{\mathcal{S}}[F(\mathcal{S}) - \lambda \sum b_t]$ would be NP-hard.

# What if SPA?

- Deviation step 1 (Lagrangian formulation):

$$\mathcal{L}(\boldsymbol{b},\lambda) = \sum_{t=1}^{T} \Delta_t \cdot x_t - \lambda \left( \sum_{t=1}^{T} z_t \cdot x_t - B \right)$$

$$= \lambda B + \sum_{t=1}^{T} (\Delta_t - \lambda z_t) \cdot x_t$$

- Step 2 (Per auction decision): $(\Delta_t - \lambda z_t) \geq 0 \xrightarrow{\text{yields}} z_t \leq \frac{\Delta_t}{\lambda}$
- Step 3 (Optimal bid): The creator pays $z_t$ and wins iff $b_t \geq z_t$, so the optimal bid is $b_t^* = \frac{\Delta_t}{\lambda}$.
- Takeaway: This work's core contribution is not changed in SPA.

# Proof Sketch for Theorem 1

- **Step 1 (Nearest-neighborhood bound):** For each $x \in \mathcal{D}_{val}$, let $z \in \mathcal{S}$ be its closest gradient neighbor with distance $d_x = \|g(x) - g(z_x)\|^2$. By PSD ordering: $\mathcal{I}(\mathcal{S})^{-1} \preccurlyeq (\gamma I_d + g(z_x)g(z_x)^\top)^{-1}$.
- **Step 2 (Sherman-Morrison inversion):** Apply the rank-1 update formula: $g(x)^\top \mathcal{I}_\gamma^{-1} g(x) \leq \frac{\|g(x)\|^2}{\gamma} - \frac{1}{\gamma^2} \frac{\langle g(x), g(z_x)\rangle^2}{1+\|g(z_x)\|^2/\gamma}$
- **Step 3 (Relate inner product to distance):** Using $\|g(x) - g(z_x)\|^2 = \|g(x)\|^2 + \|g(z_x)\|^2 - 2\langle g(x), g(z_x)\rangle$, and bounds $\|g\| \leq L$, $\|g(z_x)\| \geq m$:

$$On\ A_\tau = \{x: d_x \leq \tau\}: \langle g(x), g(x_z)\rangle^2 \geq \frac{\left(2m^2 - \tau\right)^2}{4}$$

- **Step 4 (Sum and connect to $U_\lambda$):** Summing over $\mathcal{D}_{val}$ and counting $|\mathcal{A}_\tau|$ via the kernel $U_\lambda \leq |A_\tau| + (k - |A_\tau|)e^{-\lambda\tau}$, so $|A_\tau| \geq \frac{U_\lambda - ke^{-\lambda\tau}}{1-e^{-\lambda\tau}}$. Substituting yields the bound.

# Proof Sketch of Theorem 3

- **Step 1 (Define per-round loss):** let $\ell_t(\lambda) = -f_t(\lambda)$ where $f_t(\lambda) = \max_b W_t(b)(\Delta_t - \lambda b)$. $\ell_t$ is convex; subgradient: $\partial \ell_t(\lambda_{t-1}) \ni h_t / C$, where $h_t = W_t(b_t) b_t$ is the expected spend.
- **Step 2 (Mirror descent regret):** The update $\lambda_t = \lambda_{t-1} \exp(\eta h_t / C)$ is OMD with negative entropy regularizer. $\sum_{t=1}^{T} f_t(\lambda_{t-1}) \geq \sum_{t=1}^{T} f_t(\lambda^*) - \frac{\log\left(\frac{\lambda_{max}}{\lambda_0}\right)}{\eta} - \frac{\eta T}{2}$
- **Step 3 (Relate utility to dual obj):** By envelope theorem: $f_t(\lambda_{t-1}) = W_t(b_t)(\Delta_t - \lambda_{t-1} b_t)$. Therefore $\sum_t W_t(b_t)\Delta_t = \sum_t f_t(\lambda_{t-1}) + \sum_t \lambda_{t-1} h_t$. Another MWU bound gives $\sum_t \lambda_{t-1} h_t \geq \lambda^* \sum_t h_t - O(\frac{1}{\eta} + \eta T C^2)$.
- **Step 4 (Combine with weak duality):** $OPT \leq \sum_t f_t(\lambda^*) + \lambda^* B$. Setting $\eta = \frac{1}{C}\sqrt{\frac{\log\left(\frac{\lambda_{max}}{\lambda_0}\right)}{T}}$ yields the $O(C\sqrt{T})$ regret.

# Proof sketch for Theorem 4

- **Step 1 (Telescoping the dual update):** The multiplicative weights rule is $\lambda_t = \lambda_{t-1} \exp\left(\eta \frac{W_a(b_t)b_t}{B}\right)$. Taking logs and telescoping from $t = 1 \ldots T$: $\log \lambda_T - \log \lambda_0 = \frac{\eta}{B} \sum_{t=1}^{T} W_a(b_t) b_t$.
- **Step 2 (Rearrange and bound $\lambda_T$):** Since $\lambda_T \leq \lambda_{max}$

$$\sum_{t=1}^{T} W_a(b_t) b_t = \frac{B}{\eta} \log \frac{\lambda_T}{\lambda_0} \leq \frac{B}{\eta} \log \frac{\lambda_{max}}{\lambda_0}$$

- **Step 3 (Take expectation):** By the law of total expectation: $\mathbb{E}\left[\mathrm{b_t} \cdot 1_{win_t} \middle| b_t\right] = \mathrm{W_a(b_t)b_t}$. Therefore:

$$\mathbb{E}\left[\sum_{t=1}^{T} b_t \cdot 1_{win_t}\right] = \sum_{t=1}^{T} \mathbb{E}[W_a(b_t)b_t] \leq B + \frac{\log\left(\frac{\lambda_{max}}{\lambda_0}\right)}{\eta}.$$

The $\frac{\log\left(\frac{\lambda_{max}}{\lambda_0}\right)}{\eta}$ term is the "cost of online learning", controlled by $\eta$ and vanishing as $\eta \to \infty$ (but larger $\eta$ increases instability).

# Why reduce uncertainty matters?

- Setup of a toy model:
  - A creator iteratively improves content parameters $x \in \mathbb{R}^d$.
  - At step t, propose a random perturbation $\delta_t \sim Unif(r\mathbb{S}^{d-1})$.
  - Observe noisy reward $\tilde{R}(x) = R(x) + \epsilon$, where $\epsilon \sim \mathcal{N}(0, \xi^2)$ iid.
  - Accept if improvement: $x_{t+1} = x_t + \delta_t$ if $\tilde{R}(x_t + \delta_t) > \tilde{R}(x_t)$, else $x_{t+1} = x_t$.

Fan Yao, et. al. User Welfare Optimization in Recommender Systems with Competing Content Creators. KDD 2024.

- Assumptions:
  - L-smoothness; bounded reward, Gaussian noise
  - Small step size: $r \cdot \|\nabla R(x)\| \ll \xi$.
- Result

$$\frac{1}{T}\sum_{t=0}^{T-1} \mathbb{E}[\|\nabla R(x_t)\|^2] \leq O\left(\frac{\xi \cdot d \cdot \left(R^* - R(x_0)\right)}{r^2 T}\right) + O(L \cdot \xi \cdot d)$$

# Proof sketch of Theorem 5

- **Step 1 (Compute expected improvement):** True improvement: $\Delta_t \approx \langle \nabla R(x_t), \delta_t \rangle$ (by L-smoothness, small $r$). Acceptance prob $\pi(\delta_t) = \Phi\left(\frac{\Delta_t}{\sqrt{2}\xi}\right)$. Under high-noise regime ($r \cdot \|\nabla R\| \ll \xi$), expand $\Phi(z) \approx \frac{1}{2} + \frac{z}{2\sqrt{\pi}}$:

$$\mathbb{E}[R_{t+1} - R_t] \approx \frac{1}{2}\mathbb{E}[\Delta_t] + \frac{1}{2\sqrt{\pi}\xi}\mathbb{E}[\Delta_t^2]$$

- **Step 2 (Bound using spherical expectation):** For $\delta_t$ uniform on sphere:$\mathbb{E}[\langle g, \delta \rangle^2] = \frac{r^2}{d}\|g\|^2$. Hence:

$$\mathbb{E}[R_{t+1} - R_t] \gtrsim -\frac{Lr^2}{4} + \frac{r^2}{2\sqrt{\pi}d\xi}\|g_t\|^2$$

- **Step 3 (Rearrange and telescope):** Solving for $\|g_t\|^2$ and averaging over T steps, then obtain the result.