Dispersion for Data-Driven Algorithm Design, Online Learning, and Private Optimization
Ellen Vitercik
Joint work with Nina Balcan and Travis Dick. Appeared in FOCS 2018.

Data-driven algorithm design

Some problems have optimal, fast algorithms. Some don’t:
• Many methods – which is best for our domain?
• E.g., clustering and subset selection problems

Use ML to automate algorithm design!

This work: formal guarantees for this approach

Learning setup

• Fix large family of parameterized algorithms
  • E.g., knapsack: until knapsack full, add items in decreasing order of
    \( \frac{\text{value}}{\text{size}} \)
  
• Learner sees stream of \( T \) problem instances
  • At timestep \( t \), choose parameters \( \rho_t \)

\[ \rho_1 = 1 \quad \rho_2 = \frac{1}{4} \quad \rho_3 = \frac{3}{4} \]

• Want to minimize regret:
  • Difference between cumulative performance of those parameters and optimal parameters in hindsight:
    \[ \max_{\rho} \sum_t u_t(\rho) - \sum_t u_t(\rho_t) \]
  where \( u_t(\rho) \) measures performance of algorithm parameterized by \( \rho \) on \( t \)th problem

Main challenge

• Algorithm’s performance on an instance as function of parameters is often piecewise Lipschitz
• In general, optimizing piecewise Lipschitz functions is impossible!

Approach

Exponentially-weighted forecaster (EWF): On round \( t \), choose parameters \( \rho \) w.p. \( \propto \exp(\ell_{t-1} u_t(\rho)) \)

Dispersion

\[ (u_0, \ldots, u_T: \mathbb{R}^d \rightarrow [0, 1]) \text{ is (w,k)-dispersed at } \rho \text{ if } \ell_2 \text{-ball } B(\rho, w) \text{ contains discontinuities for } \leq k \text{ functions} \]

Lemma: If \( u_0, \ldots, u_T \) are piecewise \( L \)-Lipschitz and \( (w, k) \)-dispersed at \( \rho \), for every \( \rho \in B(\rho^*, w) \),

\[ \sum_{t=1}^T u_t(\rho) \geq OPT - TL + k \]

Proof: \( u_1, \ldots, u_T \)

Is \( u_1 \) \( L \)-Lipschitz on \( B(\rho^*, w) \)?

\[ u_1(\rho) - u_1(\rho^*) \leq \frac{1}{L} \]

\[ u_1(\rho) - u_1(\rho^*) \leq Lw \]

No (\( \leq k \) functions)

Yes (\( \leq T \) functions)

Our guarantees

Upper bound: If \( u_0, \ldots, u_T \) are piecewise \( L \)-Lipschitz and \( (w, k) \)-dispersed at \( \rho^* \), EWF has regret

\[ O \left( \sqrt{T d \log \frac{1}{w} + TL w + k} \right) \]

When is this a good bound? For \( w = \frac{1}{\sqrt{T}} \) and \( k = \hat{O} \left( \sqrt{d} \right) \) regret is \( O(\sqrt{d}) \).

Matching lower bound: For any algorithm \( A \), there are piecewise constant functions \( u_0, \ldots, u_T \) so that \( A \) has expected regret

\[ \Omega \left( \inf_{\{u_0, \ldots, u_T\}} \frac{T d \log \frac{1}{w} + k}{w} \right) \]

Infimum is over all \( (w, k) \)-dispersion parameters satisfied by \( u_0, \ldots, u_T \) at \( \rho^* \).

Applications

Prove dispersion for:
• Pricing and auction design
• Algorithm configuration for subset selection problems (e.g., knapsack and maximum weight independent set) and integer quadratic programming

Dispersion also implies differentially private optimization guarantees (tight lower bounds!)

My related work


Smooth adversaries

Adversary chooses thresholds \( u_t: [0, 1] \rightarrow \{0, 1\} \)

Discontinuity \( \tau \) “smoothened” by adding \( Z \sim N(0, \sigma^2) \).

Lemma: W.h.p, for any \( w > 0 \), \( \{u_0, \ldots, u_T\} \) is \( (w, k) \)-dispersed for \( k = O \left( \frac{d + \sqrt{d}}{\sigma^2} \right) \).

Proof: Density of \( r + Z \) is \( O \left( \frac{1}{w} \right) \).

• For any width-\( w \) interval, expected number discontinuities is \( O \left( \frac{w}{\sigma^2} \right) \).
• Intervals have VC-dim \( 2 \) \( \Rightarrow \) W.h.p., every interval contains \( \hat{O} \left( \frac{d}{\sigma^2} + \sqrt{d} \right) \) discontinuities.

\[ w = \frac{1}{\sigma^2} \Rightarrow \text{Regret} = O \left( \sqrt{\frac{T d \log \frac{2}{\sigma^2}}{\sigma^2}} \right) \]