Volumetric Ellipsoids: An exploration basis for learning

Elad Hazan  Zohar Karnin  Raghu Mehka
Technion  Yahoo Labs  MSR

COLT’ 14
Experiment Design

$n$ patients, each with $d$ features: $x_1, \ldots, x_n$ in $\mathbb{R}^d$

experiment = choose $x_i$, gives noisy measurement:

$$\langle w^*, x_i \rangle + \epsilon_i$$

Goal: regress on data with as few measurements as possible (learn $w^*$)
Experiment Design

$n$ patients, each with $d$ features: $x_1,...,x_n$ in $\mathbb{R}^d$

Goal: regress on data with as few measurements as possible (learn $w^*$)

Our objective: Find $S \subseteq \{x_1,...,x_n\}$ of minimal size with

$$\max_x \text{Var}[\langle x, w-w^* \rangle] \leq \epsilon$$
Low variance exploration basis

Given $K$ (discrete or continuous):

- Subset of $K$: $S = \{v_1, \ldots, v_t\} \subseteq K$
- Small cardinality – $t$
- “generalize” all $x$ in $K$ with low variance:
Low variance exploration basis

Given $K$ (discrete or continuous):

- Subset of $K$: $S = \{v_1, \ldots, v_t\} \subseteq K$
- Small cardinality – $t$
- “generalize” all $x$ in $K$ with low variance:

$$x = \sum_{i \in [t]} \alpha_i \cdot v_i$$

Volumetric Spanner

$$\sum_i \alpha_i^2 \leq 1$$
Low variance exploration basis

Given $K$ (discrete or continuous):

- Subset of $K$: $S = \{v_1, \ldots, v_t\} \subseteq K$
- Small cardinality – $t$
- "generalize" all $x$ in $K$ with low variance:

$$x = \sum_{i \in [t]} \alpha_i v_i$$

<table>
<thead>
<tr>
<th>Volumetric Spanner</th>
<th>Barycentric spanner</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sum_i \alpha_i^2 \leq 1$</td>
<td>$\max_i</td>
</tr>
</tbody>
</table>

[Awerbuch Kleinberg]
Low variance exploration basis

Given $K$ (discrete or continuous):

- Subset of $K$: $S = \{v_1, ..., v_t\} \subseteq K$
- Small cardinality – $t$
- “generalize” all $x$ in $K$ with low variance:

$$
\langle x, w^* \rangle = \sum_i \alpha_i \langle v_i, w^* \rangle
$$

$$
\text{Var}[\langle x, w^* \rangle] = \text{Var} [\sum_i \alpha_i \langle v_i, w^* \rangle] \leq \|\alpha\|^2 \sigma^2
$$
Application: BLO

• Round: player picks $x_t$, nature picks $L_t$, loss = $\ell_t$

(Loss) $\ell_t = \langle x_t, L_t \rangle$
Application: BLO

- Round: player picks $x_t$, nature picks $L_t$, loss $= \ell_t$
- Only loss is observed, $L_t$ is not

(Loss) $\ell_t = \langle x_t, L_t \rangle$
Application: BLO

Round: player picks $x_t$, nature picks $L_t$, loss $= \ell_t$

Only loss is observed, $L_t$ is not

Typical assumption: $\forall x_t \in K, L_t \in Z : |\langle x_t, L_t \rangle| \leq 1$
Application: BLO

Round: player picks $x_t$, nature picks $L_t$, loss $= \ell_t$

Only loss is observed, $L_t$ is not

Typical assumption: $\forall x_t \in K, L_t \in Z : \quad |\langle x_t, L_t \rangle| \leq 1$

Regret $= \sum_t \langle x_t, L_t \rangle - \sum_t \langle x^*, L_t \rangle$
# BLO Results

<table>
<thead>
<tr>
<th>Paper</th>
<th>Regret</th>
<th>Efficient?</th>
<th>Comments</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Awerbuch, Kleinberg]</td>
<td>$T^{2/3} \text{poly}(d)$</td>
<td>Yes</td>
<td></td>
<td>Barycentric spanner</td>
</tr>
<tr>
<td>[Abernethy, Hazan, Rakhlin]</td>
<td>$d (\theta T)^{1/2}$</td>
<td>Yes</td>
<td>$\theta =$self concordant barrier $= O(d)$, but can be $O(1)$</td>
<td>Self concordant barriers</td>
</tr>
<tr>
<td>[Bubeck, Cesa-Bianchi, Kakade]</td>
<td>$dT^{1/2}$</td>
<td>No</td>
<td>matches lower bound (for general convex bodies)</td>
<td>John’s decomposition</td>
</tr>
<tr>
<td>Here</td>
<td>$dT^{1/2}$</td>
<td>Yes</td>
<td>For some specific bodies, better regret is possible. Worst case is $dT^{1/2}$</td>
<td>Volumetric Spanners</td>
</tr>
</tbody>
</table>
Volumetric Ellipsoids

For $S = \{v_1, ..., v_t\} \subseteq K$

\[
\|x\|_S^2 = x^T (\sum_i v_i v_i^T)^{-1} x
\]
Volumetric Ellipsoids

For $S = \{v_1, ..., v_t\} \subseteq K$

- $\|x\|_S^2 = x^T (\sum_i v_i v_i^T)^{-1} x$
- $\mathcal{E}(S) = \{x \mid \|x\|_S \leq 1\}$
$S$

$\mathcal{E}(S)$
Volumetric Ellipsoids

For $S = \{v_1,\ldots,v_t\} \subseteq K$

- $\|x\|_S^2 = x^T (\sum_i v_i v_i^T)^{-1} x$
- $\mathcal{E}(S) = \{x \mid \|x\|_S \leq 1\}$

**Obs:** if $S \subseteq K \subseteq \mathcal{E}(S)$ then $\max_x \text{Var}[\langle x, w-w^* \rangle] \leq 1$
Volumetric Ellipsoids

For $S = \{v_1, \ldots, v_t\} \subseteq K$

- $\|x\|_S^2 = x^T \left( \sum_i v_i v_i^T \right)^{-1} x$
- $\mathcal{E}(S) = \{x \mid \|x\|_S \leq 1\}$

**Obs:** if $S \subseteq K \subseteq \mathcal{E}(S)$ then $\max_x \text{Var}[\langle x, w-w^* \rangle] \leq 1$

**Objective:** find $\min |S|$ s.t. $S \subseteq K \subseteq \mathcal{E}(S)$
Questions

• Given K, how small is $|S|$ of volumetric ellipsoid? “order(K)”
• Can we find it efficiently?
• What bounds does it give for experiment design?
• Can it be used in other learning problems?
What is order(K) ?

Thm: $d < \text{order}(K) \leq 12d$
What is order(K)?

Thm: \( d \leq \text{order}(K) \leq 12d \)
Order(K) > d
What is order(K) ?

Thm: $d < \text{order}(K) \leq 12d$

Proof:
John’s decomposition provide a “broken solution”; can be seen as $S$ of size $d^2$. Twice-Ramanujan sparsifiers (BSS’ 12) provide method to reduce to $12d$. 
What is order(K) ?

Thm: \( d < \text{order}(K) \leq 12d \)

Application in experiment design:
Test at most \( O(d / \epsilon^2) \) patients to regress up to precision \( \epsilon \) over all patients!
Constructive Results for Finite K

- John’s decomposition can be computed via convex programming $\text{poly}(|K|,d)$
Constructive Results for Finite $K$

- John’s decomposition can be computed via convex programming $\text{poly}(|K|, d)$
- **Lem**: For large $|K|$, we give a simpler $|K|d^2$ construction giving size $O(d \log(d) \log(|K|))$
Constructive Results for Convex \( K \)

Two types of approximation:

1) Ratio-Spanner: \( K \subseteq \rho \cdot \mathcal{E}(S) \)

2) Exp-Spanner: for \( x \in K \), \( \Pr[x \in \theta \cdot \mathcal{E}(S)] \geq 1 - \varepsilon^\theta \)
Constructive Results for Convex K

Two types of approximation:

1) Ratio-Spanner: \( K \subseteq \rho \cdot \mathcal{E}(S) \)

2) Exp-Spanner: for \( x \in K \), \( \Pr[x \in \theta \cdot \mathcal{E}(S)] \geq 1-\epsilon^\theta \)

Application to Bandit Linear Optimization: First efficient and optimal-regret\(^*\) algorithm

- \(^*\) convex sets, L2 assumption, can do better for some specific convex bodies
- Geometric hedge framework: Use approximate volumetric spanner as exploration basis
Conclusion

- Experiment design: test at most $O(d / \epsilon^2)$ patients to regress up to precision $\epsilon$ over all patients!

- Bandit Linear Optimization: first efficient and optimal-regret algorithm for general convex sets

- Active learning applications?