Variance-Aware Sparse Linear Bandits

Yan Dai [1], Ruosong Wang [2], Simon S. Du [2]

[1] Institute for Interdisciplinary Information Sciences, Tsinghua University

[2] Paul G. Allen School of Computer Science & Engineering, University of Washington

Presented by Yan Dai

Problem Setup & Related Work

- A T-round game between an *agent* and an *environment* with **heteroscedastic noises**

Problem Setup & Related Work

- A T-round game between an *agent* and an *environment* with **heteroscedastic noises**
- In round t, agent picks *action* a_t from d-dim unit sphere $\mathbb{S}^{d-1}$ and *observes* $r(a_t) + \eta_t$:
  $r(a_t) = \langle a_t, \theta^* \rangle$, $\eta_t \sim \mathcal{N}(0, \sigma_t^2)$, both θ^* and $\{\sigma_t^2\}_{t=1}^T$ are **unknown**

Problem Setup & Related Work

- A T-round game between an *agent* and an *environment* with **heteroscedastic noises**
- In round t, agent picks *action* a_t from d-dim unit sphere $\mathbb{S}^{d-1}$ and *observes* $r(a_t) + \eta_t$:
  $r(a_t) = \langle a_t, \theta^* \rangle$, $\eta_t \sim \mathcal{N}(0, \sigma_t^2)$, both θ^* and $\{\sigma_t^2\}_{t=1}^T$ are **unknown**
- **Assumption** (Sparsity): $\|\theta^*\|_0 = s \ll d$. However, s is also **unknown**.

Problem Setup & Related Work

- A T-round game between an *agent* and an *environment* with **heteroscedastic noises**
- In round t, agent picks *action* a_t from d-dim unit sphere $\mathbb{S}^{d-1}$ and *observes* $r(a_t) + \eta_t$:
  $r(a_t) = \langle a_t, \theta^* \rangle$, $\eta_t \sim \mathcal{N}(0, \sigma_t^2)$, both θ^* and $\{\sigma_t^2\}_{t=1}^T$ are **unknown**
- **Assumption** (Sparsity): $\|\theta^*\|_0 = s \ll d$. However, s is also **unknown**.
- **Objective**: Minimize *regret* $\mathfrak{R}_T = \mathbb{E}[\sum_{t=1}^T \langle \theta^* - a_t, \theta^* \rangle]$ (assuming $\theta^* \in \mathbb{S}^{d-1}$)

Problem Setup & Related Work

- A T-round game between an *agent* and an *environment* with **heteroscedastic noises**
- In round t, agent picks *action* a_t from d-dim unit sphere $\mathbb{S}^{d-1}$ and *observes* $r(a_t) + \eta_t$:
  $r(a_t) = \langle a_t, \theta^* \rangle$, $\eta_t \sim \mathcal{N}(0, \sigma_t^2)$, both θ^* and $\{\sigma_t^2\}_{t=1}^T$ are **unknown**
- **Assumption** (Sparsity): $\|\theta^*\|_0 = s \ll d$. However, s is also **unknown**.
- **Objective**: Minimize *regret* $\mathfrak{R}_T = \mathbb{E}[\sum_{t=1}^T \langle \theta^* - a_t, \theta^* \rangle]$ (assuming $\theta^* \in \mathbb{S}^{d-1}$)

Table 1: Comparison with Related Work; $\tilde{O}$ hides all logarithmic factors in $s, d,$ and T

	Setting	Regret
Abbasi-Yadkori et al. (2012)	Sparse Linear Bandit with $\boldsymbol{\sigma_t \equiv 1}$	$\tilde{O}(\sqrt{sdT})$
This paper	Sparse Linear Bandit with varying σ_t	$\tilde{O}\left((s^{2.5} + s\sqrt{d})\sqrt{\sum_{t=1}^T \sigma_t^2} + s^3\right)$

Problem Setup & Related Work

- A T-round game between an *agent* and an *environment* with **heteroscedastic noises**
- In round t, agent picks *action* a_t from d-dim unit sphere $\mathbb{S}^{d-1}$ and *observes* $r(a_t) + \eta_t$:
  $r(a_t) = \langle a_t, \theta^* \rangle$, $\eta_t \sim \mathcal{N}(0, \sigma_t^2)$, both θ^* and $\{\sigma_t^2\}_{t=1}^T$ are **unknown**
- **Assumption** (Sparsity): $\|\theta^*\|_0 = s \ll d$. However, s is also **unknown**.
- **Objective**: Minimize *regret* $\mathfrak{R}_T = \mathbb{E}[\sum_{t=1}^T \langle \theta^* - a_t, \theta^* \rangle]$ (assuming $\theta^* \in \mathbb{S}^{d-1}$)

Table 1: Comparison with Related Work; $\tilde{O}$ hides all logarithmic factors in $s, d,$ and T

	Setting	Regret
Abbasi-Yadkori et al. (2012)	Sparse Linear Bandit with $\boldsymbol{\sigma_t \equiv 1}$	$\tilde{O}(\sqrt{sdT})$
Dong et al. (2021)	Sparse Linear Bandit with $\boldsymbol{\sigma_t \equiv 0}$	$\tilde{O}(s)$
This paper	**Sparse Linear Bandit with varying $\boldsymbol{\sigma_t}$**	$\tilde{O}\left((s^{2.5} + s\sqrt{d})\sqrt{\sum_{t=1}^T \sigma_t^2} + s^3\right)$

Problem Setup & Related Work

- A T-round game between an *agent* and an *environment* with **heteroscedastic noises**
- In round t, agent picks *action* a_t from d-dim unit sphere $\mathbb{S}^{d-1}$ and *observes* $r(a_t)+\eta_t$:
  $r(a_t)=\langle a_t,\theta^*\rangle$, $\eta_t\sim\mathcal{N}(0,\sigma_t^2)$, both θ^* and $\{\sigma_t^2\}_{t=1}^T$ are **unknown**
- **Assumption** (Sparsity): $\|\theta^*\|_0=s\ll d$. However, s is also **unknown**.
- **Objective**: Minimize *regret* $\mathfrak{R}_T=\mathbb{E}[\sum_{t=1}^T\langle\theta^*-a_t,\theta^*\rangle]$ (assuming $\theta^*\in\mathbb{S}^{d-1}$)

Table 1: Comparison with Related Work; $\tilde{O}$ hides all logarithmic factors in $s, d,$ and T

	Setting	Regret
Abbasi-Yadkori et al. (2012)	Sparse Linear Bandit with $\boldsymbol{\sigma_t\equiv 1}$	$\tilde{O}(\sqrt{sdT})$
Dong et al. (2021)	Sparse Linear Bandit with $\boldsymbol{\sigma_t\equiv 0}$	$\tilde{O}(s)$
This paper	**Sparse Linear Bandit with varying $\boldsymbol{\sigma_t}$**	$\tilde{O}\left((s^{2.5}+s\sqrt{d})\sqrt{\sum_{t=1}^T\sigma_t^2}+s^3\right)$

Problem Setup & Related Work

- A T-round game between an *agent* and an *environment* with **heteroscedastic noises**
- In round t, agent picks *action* a_t from d-dim unit sphere $\mathbb{S}^{d-1}$ and *observes* $r(a_t) + \eta_t$:
  $r(a_t) = \langle a_t, \theta^* \rangle$, $\eta_t \sim \mathcal{N}(0, \sigma_t^2)$, both θ^* and $\{\sigma_t^2\}_{t=1}^T$ are **unknown**
- **Assumption** (Sparsity): $\|\theta^*\|_0 = s \ll d$. However, s is also **unknown**.
- **Objective**: Minimize *regret* $\mathfrak{R}_T = \mathbb{E}[\sum_{t=1}^T \langle \theta^* - a_t, \theta^* \rangle]$ (assuming $\theta^* \in \mathbb{S}^{d-1}$)

$\tilde{O}(\text{poly}(s)\,\boldsymbol{\sqrt{dT}})$ when $\boldsymbol{\sigma_t \equiv 1}$

$\tilde{O}(\textbf{poly}(\boldsymbol{s}))$ when $\boldsymbol{\sigma_t \equiv 0}$

Table 1: Comparison with Related Work; $\tilde{O}$ hides all logarithmic factors in s, d, and T

	Setting	Regret
Abbasi-Yadkori et al. (2012)	Sparse Linear Bandit with $\boldsymbol{\sigma_t \equiv 1}$	$\tilde{O}(\sqrt{sdT})$
Dong et al. (2021)	Sparse Linear Bandit with $\boldsymbol{\sigma_t \equiv 0}$	$\tilde{O}(s)$
This paper	**Sparse Linear Bandit with varying $\boldsymbol{\sigma_t}$**	$\tilde{O}\left((s^{2.5} + s\sqrt{d})\sqrt{\sum_{t=1}^T \sigma_t^2} + s^3\right)$

More Importantly…

$\tilde{O}(\mathrm{poly}(s)\sqrt{\boldsymbol{dT}})$ when $\boldsymbol{\sigma_t \equiv 1}$

$\tilde{O}(\mathbf{poly}(\boldsymbol{s}))$ when $\boldsymbol{\sigma_t \equiv 0}$

Table 1: Comparison with Related Work; $\tilde{O}$ hides all logarithmic factors in $s, d,$ and T

	Setting	Regret
Abbasi-Yadkori et al. (2012)	Sparse Linear Bandit with $\boldsymbol{\sigma_t \equiv 1}$	$\tilde{O}(\sqrt{sdT})$
Dong et al. (2021)	Sparse Linear Bandit with $\boldsymbol{\sigma_t \equiv 0}$	$\tilde{O}(s)$
This paper	**Sparse Linear Bandit with varying $\boldsymbol{\sigma_t}$**	$\tilde{O}\left((s^{2.5}+s\sqrt{d})\sqrt{\sum_{t=1}^T \sigma_t^2}+s^3\right)$

More Importantly…

Main Theorem (informal). For **any** *variance-aware linear bandit algorithm* $\mathcal{A}$ whose regret $\mathfrak{R}_T^{\mathcal{A}}$ satisfies

$$\mathfrak{R}_T^{\mathcal{A}} = \tilde{O}\left(f(d)\sqrt{\sum \sigma_t^2} + g(d)\right) \text{ for some functions } f, g,$$

our framework gives a *variance-aware sparse linear bandit algorithm* $\mathcal{B}$ whose regret $\mathfrak{R}_T^{\mathcal{B}}$ satisfies

$$\mathfrak{R}_T^{\mathcal{B}} = \tilde{O}\left((sf(s) + s\sqrt{d})\sqrt{\sum \sigma_t^2} + sg(s)\right), \text{giving } \tilde{O}(\text{poly}(s)\,\sqrt{dT}) \text{ when } \sigma_t \equiv 1 \text{ and } \tilde{O}(\textbf{poly}(s)) \text{ when } \sigma_t \equiv 0.$$

$\tilde{O}(\text{poly}(s)\,\sqrt{\boldsymbol{dT}})$ when $\boldsymbol{\sigma_t \equiv 1}$

$\tilde{O}(\textbf{poly}(\boldsymbol{s}))$ when $\boldsymbol{\sigma_t \equiv 0}$

Table 1: Comparison with Related Work; $\tilde{O}$ hides all logarithmic factors in $s, d,$ and T

	Setting	Regret
Abbasi-Yadkori et al. (2012)	Sparse Linear Bandit with $\boldsymbol{\sigma_t \equiv 1}$	$\tilde{O}(\sqrt{sdT})$
Dong et al. (2021)	Sparse Linear Bandit with $\boldsymbol{\sigma_t \equiv 0}$	$\tilde{O}(s)$
This paper	**Sparse Linear Bandit with varying $\boldsymbol{\sigma_t}$**	$\tilde{O}\left((s^{2.5} + s\sqrt{d})\sqrt{\sum_{t=1}^T \sigma_t^2} + s^3\right)$

More Importantly…

Main Theorem (informal). For **any** *variance-aware linear bandit algorithm* $\mathcal{A}$ whose regret $\mathfrak{R}_T^{\mathcal{A}}$ satisfies

$$\mathfrak{R}_T^{\mathcal{A}} = \tilde{O}\left(f(d)\sqrt{\sum \sigma_t^2} + g(d)\right) \text{ for some functions } f, g,$$

our framework gives a *variance-aware sparse linear bandit algorithm* $\mathcal{B}$ whose regret $\mathfrak{R}_T^{\mathcal{B}}$ satisfies

$$\mathfrak{R}_T^{\mathcal{B}} = \tilde{O}\left((sf(s) + s\sqrt{d})\sqrt{\sum \sigma_t^2} + sg(s)\right), \text{giving } \tilde{O}(\mathrm{poly}(s)\sqrt{dT}) \text{ when } \sigma_t \equiv 1 \text{ and } \tilde{O}(\mathbf{poly}(s)) \text{ when } \sigma_t \equiv 0.$$

$\tilde{O}(\mathrm{poly}(s)\sqrt{dT})$ when $\sigma_t \equiv 1$

$\tilde{O}(\mathbf{poly}(s))$ when $\sigma_t \equiv 0$

Table 1: Comparison with Related Work; $\tilde{O}$ hides all logarithmic factors in $s, d,$ and T

	Setting	Regret
Abbasi-Yadkori et al. (2012)	Sparse Linear Bandit with $\sigma_t \equiv 1$	$\tilde{O}(\sqrt{sdT})$
Dong et al. (2021)	Sparse Linear Bandit with $\sigma_t \equiv 0$	$\tilde{O}(s)$
This paper **(Using Kim et al. (2022))**	**Sparse Linear Bandit with varying σ_t**	$\tilde{O}\left((s^{2.5} + s\sqrt{d})\sqrt{\sum_{t=1}^T \sigma_t^2} + s^3\right)$
Kim et al. (2022)	**Linear Bandit** (i.e., $s = d$) with varying σ_t	$\tilde{O}\left(d^{1.5}\sqrt{\sum_{t=1}^T \sigma_t^2} + d^2\right)$

Thank You for Listening!

Email: yan-dai20@mails.tsinghua.edu.cn

References

- Yasin Abbasi-Yadkori, David Pal, and Csaba Szepesvari. Online-to-Confidence-Set Conversions and Application to Sparse Stochastic Bandits. In Artificial Intelligence and Statistics, pp. 1–9. PMLR, 2012.
- Kefan Dong, Jiaqi Yang, and Tengyu Ma. Provable Model-based Nonlinear Bandit and Reinforcement Learning: Shelve Optimism, Embrace Virtual Curvature. Advances in Neural Information Processing Systems, 34, 2021.
- Yeoneung Kim, Insoon Yang, and Kwang-Sung Jun. Improved Regret Analysis for Variance-Adaptive Linear Bandits and Horizon-Free Linear Mixture MDPs. Advances in Neural Information Processing Systems, 2022, 35: 1060-1072. 2022.