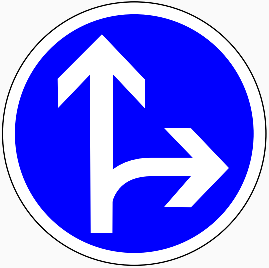
Sequential Learning of the Pareto Front in Multi-objective Bandits



Wouter M. Koolen

CWI Centrum Wiskunde & Informatica

UNIVERSITY OF TWENTE.

ELLIS ILIR Workhop
Oberwolfach
Tuesday 27th February, 2024

Team Effort

UMPA
ENS DE LYON

Élise Crepon

Aurélien Garivier

Outline

Starting Point

Almost all **optimisation** is multi-objective when you think about it.

- Vacation : sunny and tasty
- Drug trial : efficacy and toxicity
- Product dev: cost and sustainability
- ...

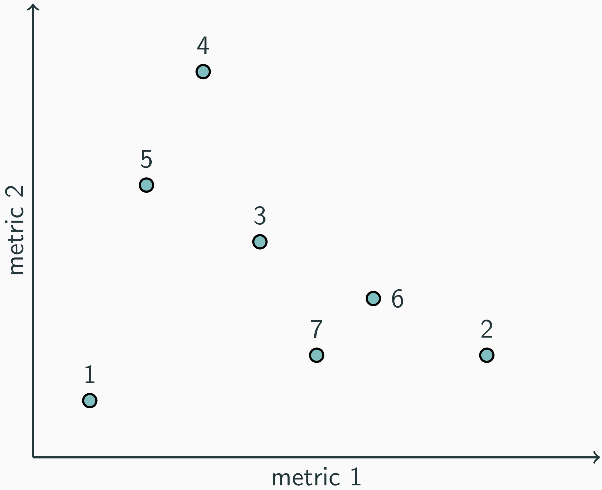
Starting Point

Almost all **optimisation** is multi-objective when you think about it.

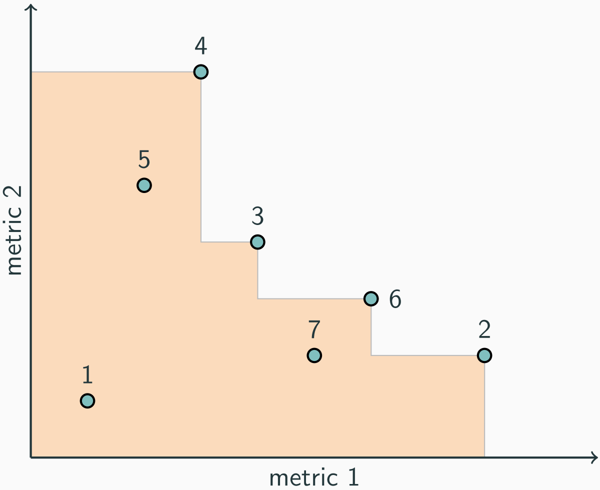
- Vacation : sunny and tasty
- Drug trial : efficacy and toxicity
- Product dev: cost and sustainability
- ...

Today: *not* in the mood to scalarise

Pareto Front



Pareto Front



Pareto front is $\{4, 3, 6, 2\}$.

Outline

1. Motivation
2. Setting
3. Our Results
4. Those Computations
5. Conclusion

Setting

K-armed multi-objective bandit $\vec{\boldsymbol{\mu}} = (\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K)$.

Setting

K-armed multi-objective bandit $\vec{\boldsymbol{\mu}} = (\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K)$.

Each arm k represented by a mean vector $\boldsymbol{\mu}_k$ in $\mathbb{R}^d$.

Setting

K-armed multi-objective bandit $\vec{\boldsymbol{\mu}} = (\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K)$.

Each arm k represented by a mean vector $\boldsymbol{\mu}_k$ in $\mathbb{R}^d$.

Observations from arm k are i.i.d. multivariate Gaussian $\mathcal{N}(\boldsymbol{\mu}_k, I)$.

Setting

K-armed multi-objective bandit $\vec{\boldsymbol{\mu}} = (\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_K)$.

Each arm k represented by a mean vector $\boldsymbol{\mu}_k$ in $\mathbb{R}^d$.

Observations from arm k are i.i.d. multivariate Gaussian $\mathcal{N}(\boldsymbol{\mu}_k, I)$.

We assume all $\boldsymbol{\mu}_k$ are different.

Setting

K-armed multi-objective bandit $\vec{\boldsymbol{\mu}} = (\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K)$.

Each arm k represented by a mean vector $\boldsymbol{\mu}_k$ in $\mathbb{R}^d$.

Observations from arm k are i.i.d. multivariate Gaussian $\mathcal{N}(\boldsymbol{\mu}_k, I)$.

We assume all $\boldsymbol{\mu}_k$ are different.

We say arm k dominates arm i, denoted $\boldsymbol{\mu}_k \succeq \boldsymbol{\mu}_i$, if $\mu_k^j \geq \mu_i^j$ in every dimension $j = 1, \ldots, d$.

Setting

K-armed multi-objective bandit $\vec{\boldsymbol{\mu}} = (\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K)$.

Each arm k represented by a mean vector $\boldsymbol{\mu}_k$ in $\mathbb{R}^d$.

Observations from arm k are i.i.d. multivariate Gaussian $\mathcal{N}(\boldsymbol{\mu}_k, I)$.

We assume all $\boldsymbol{\mu}_k$ are different.

We say arm k dominates arm i, denoted $\boldsymbol{\mu}_k \succeq \boldsymbol{\mu}_i$, if $\mu_k^j \geq \mu_i^j$ in every dimension $j = 1, \ldots, d$.

The **Pareto front** is the set of non-dominated arms:

$$S^*(\vec{\boldsymbol{\mu}}) \;:=\; \left\{ k \;\middle|\; \forall i \neq k : \boldsymbol{\mu}_i \not\succeq \boldsymbol{\mu}_k \right\}$$

Protocol

We work in the setting of **fixed-confidence** $\delta \in (0, 1)$.

Protocol

For $t = 1, 2, \ldots, \tau$*:*

- *Learner picks an arm* $I_t \in [K]$.
- *Learner sees* $X_t \sim \mathcal{N}(\boldsymbol{\mu}_{I_t}, I)$

Learner recommends Pareto front $\hat{S} \subseteq [K]$

Objectives

Learner is δ**-correct** if for any bandit instance $\vec{\boldsymbol{\mu}}$

$$\mathbb{P}_{\vec{\boldsymbol{\mu}}}\left\{\tau < \infty \wedge \hat{S} \neq S^*(\vec{\boldsymbol{\mu}})\right\} \;\leq\; \delta$$

Goal: minimise sample complexity $\mathbb{E}_{\vec{\boldsymbol{\mu}}}[\tau]$ over all δ-correct strategies.

Background Theory: Lower Bound

Define the *alternatives* to $\vec{\boldsymbol{\mu}}$ by

$$\mathrm{Alt}(\vec{\boldsymbol{\mu}}) \;\coloneqq\; \left\{\vec{\boldsymbol{\lambda}} \in \mathbb{R}^{K \times d} \;\middle|\; S^*(\vec{\boldsymbol{\lambda}}) \neq S^*(\vec{\boldsymbol{\mu}})\right\}.$$

NB recall S^* is Pareto front

Background Theory: Lower Bound

Define the *alternatives* to $\vec{\boldsymbol{\mu}}$ by

$$\mathrm{Alt}(\vec{\boldsymbol{\mu}}) \;:=\; \left\{\vec{\boldsymbol{\lambda}} \in \mathbb{R}^{K\times d} \;\middle|\; S^*(\vec{\boldsymbol{\lambda}}) \neq S^*(\vec{\boldsymbol{\mu}})\right\}.$$

NB recall S^* is Pareto front

Theorem (Garivier and Kaufmann 2016)

Fix a δ-correct strategy. Then for every bandit model $\vec{\boldsymbol{\mu}}$

$$\mathbb{E}_{\vec{\boldsymbol{\mu}}}[\tau] \;\geq\; T^*(\vec{\boldsymbol{\mu}}) \ln \frac{1}{\delta}$$

where the characteristic time $T^*(\vec{\boldsymbol{\mu}})$ is given by

$$\frac{1}{T^*(\vec{\boldsymbol{\mu}})} \;=\; \max_{\boldsymbol{w} \in \triangle_K} \; \min_{\vec{\boldsymbol{\lambda}} \in \mathrm{Alt}(\vec{\boldsymbol{\mu}})} \; \frac{1}{2} \sum_{k=1}^{K} w_k \left\|\boldsymbol{\mu}_k - \boldsymbol{\lambda}_k\right\|^2.$$

Background Theory II: Algorithm

Idea is consider the oracle weight map

$$\boldsymbol{w}^*(\vec{\boldsymbol{\mu}}) \;:=\; \operatorname*{arg\,max}_{\boldsymbol{w} \in \triangle_K} \; \min_{\vec{\boldsymbol{\lambda}} \in \mathrm{Alt}(\vec{\boldsymbol{\mu}})} \; \frac{1}{2} \sum_{k=1}^{K} w_k \left\| \boldsymbol{\mu}_k - \boldsymbol{\lambda}_k \right\|^2$$

and track the plug-in estimate: sample arm $I_t \sim \boldsymbol{w}^* \left(\hat{\vec{\boldsymbol{\mu}}}(t-1) \right)$.

Background Theory II: Algorithm

Idea is consider the oracle weight map

$$\boldsymbol{w}^*(\vec{\boldsymbol{\mu}}) \;:=\; \operatorname*{arg\,max}_{\boldsymbol{w} \in \triangle_K} \min_{\vec{\boldsymbol{\lambda}} \in \mathrm{Alt}(\vec{\boldsymbol{\mu}})} \frac{1}{2} \sum_{k=1}^{K} w_k \left\| \boldsymbol{\mu}_k - \boldsymbol{\lambda}_k \right\|^2$$

and track the plug-in estimate: sample arm $I_t \sim \boldsymbol{w}^* \left(\hat{\vec{\boldsymbol{\mu}}}(t-1) \right)$.

Theorem (Degenne and Koolen, 2019)

Take set-valued interpretation of $\arg\max$ *defining* $\boldsymbol{w}^*$*. Then* $\vec{\boldsymbol{\mu}} \mapsto \boldsymbol{w}^*(\vec{\boldsymbol{\mu}})$ *is upper-hemicontinuous and convex-valued. Suitable tracking ensures that as* $\hat{\vec{\boldsymbol{\mu}}}(t) \to \vec{\boldsymbol{\mu}}$*, any choice* $\boldsymbol{w}_t \in \boldsymbol{w}^*(\hat{\vec{\boldsymbol{\mu}}}(t-1))$ *have*

$$\min_{\boldsymbol{w} \in \boldsymbol{w}^*(\vec{\boldsymbol{\mu}})} \left\| \boldsymbol{w}_t - \boldsymbol{w} \right\|_\infty \to 0$$

Track-and-Stop is asymptotically optimal: $\limsup_{\delta \to 0} \frac{\mathbb{E}_{\vec{\mu}}[\tau]}{\ln \frac{1}{\delta}} = T^*(\vec{\boldsymbol{\mu}})$.

Outline

1. Motivation
2. Setting
3. Our Results
4. Those Computations
5. Conclusion

Contribution

Kone, Kaufmann, and Richert (2023) consider identifying the Pareto Front among K arms in d dimensions.

- Asymptotically optimal algorithm for Pareto Front Identification.
- Computations in exponential $O(d^K)$ time per round.

Our Contribution

- Computations in polynomial $O(K^d)$ time per round.

Outline

1. Motivation
2. Setting
3. Our Results
4. Those Computations
5. Conclusion

What do we need to calculate

Degenne, Koolen, and Ménard (2019): sufficient to implement best-response oracle (= gradient)

$$\vec{\boldsymbol{\mu}}, \boldsymbol{w} \;\mapsto\; \min_{\vec{\boldsymbol{\lambda}} \in \mathrm{Alt}(\vec{\boldsymbol{\mu}})} \frac{1}{2} \sum_{k=1}^{K} w_k \left\| \boldsymbol{\mu}_k - \boldsymbol{\lambda}_k \right\|^2$$

What do we need to calculate

Degenne, Koolen, and Ménard (2019): sufficient to implement best-response oracle (= gradient)

$$\vec{\boldsymbol{\mu}}, \boldsymbol{w} \;\mapsto\; \min_{\vec{\boldsymbol{\lambda}} \in \mathrm{Alt}(\vec{\boldsymbol{\mu}})} \; \frac{1}{2} \sum_{k=1}^{K} w_k \left\| \boldsymbol{\mu}_k - \boldsymbol{\lambda}_k \right\|^2$$

Objective is convex, but domain $\mathrm{Alt}(\vec{\boldsymbol{\mu}})$ is not.

What do we need to calculate

Degenne, Koolen, and Ménard (2019): sufficient to implement best-response oracle (= gradient)

$$\vec{\boldsymbol{\mu}}, \boldsymbol{w} \ \mapsto \ \min_{\vec{\boldsymbol{\lambda}} \in \mathrm{Alt}(\vec{\boldsymbol{\mu}})} \ \frac{1}{2} \sum_{k=1}^{K} w_k \left\| \boldsymbol{\mu}_k - \boldsymbol{\lambda}_k \right\|^2$$

Objective is convex, but domain $\mathrm{Alt}(\vec{\boldsymbol{\mu}})$ is not.

Optimal transport problem

Being in the Alternative

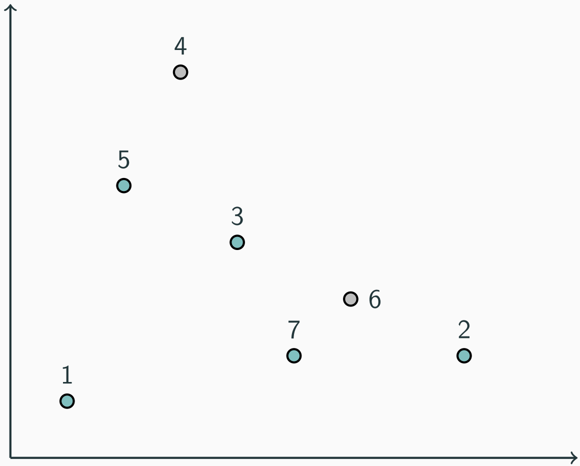
Recall

$$\vec{\boldsymbol{\lambda}} \in \mathrm{Alt}(\vec{\boldsymbol{\mu}}) \qquad \text{i.e.} \qquad S^*(\vec{\boldsymbol{\lambda}}) \neq S^*(\vec{\boldsymbol{\mu}})$$

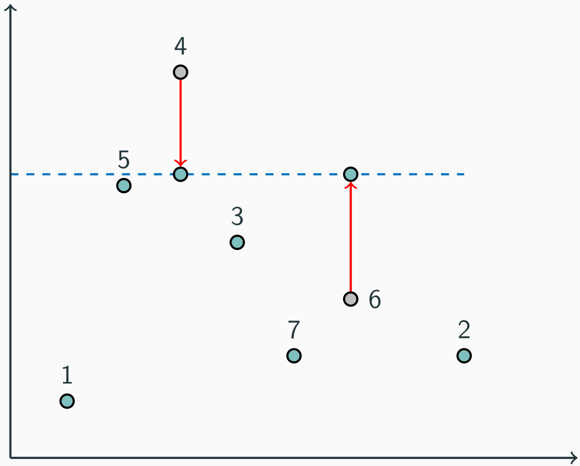
Having a different Pareto front means either

- An arm on the front in $\vec{\boldsymbol{\mu}}$ is off the front in $\vec{\boldsymbol{\lambda}}$, or
- An arm off the front in $\vec{\boldsymbol{\mu}}$ is on the front in $\vec{\boldsymbol{\lambda}}$.

Taking arm 4 off the Pareto Front

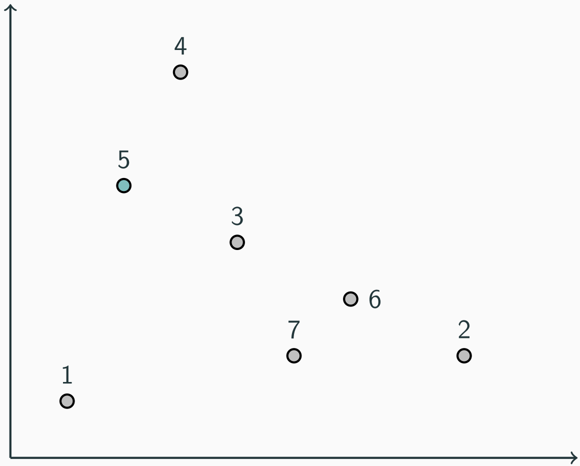


Taking arm 4 off the Pareto Front

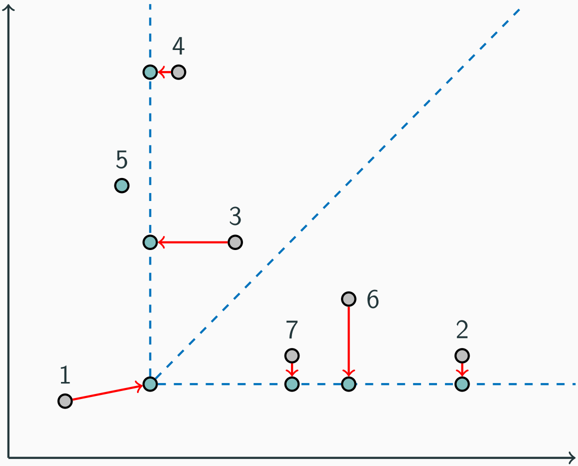


Example: we dominate arm 4 using arm 6 by moving each to the **weighted mid-point** in non-dominated coordinates.

Putting arm 1 on the Pareto Front



Putting arm 1 on the Pareto Front



Example: we make point 1 dominant by moving it north-east, and then moving all dominators **out of the way**.

The heart of the insight

The cost for moving point 1 onto the front is:

$$\min_{\boldsymbol{\lambda}_1} \frac{w_1}{2}||\boldsymbol{\mu}_1 - \boldsymbol{\lambda}_1||^2 + \sum_{k \in S^*(\vec{\boldsymbol{\mu}})} \frac{w_k}{2} \min_{j \in [d]} (\boldsymbol{\mu}_k^j - \boldsymbol{\lambda}_1^j)_+^2$$

The heart of the insight

The cost for moving point 1 onto the front is:

$$\min_{\boldsymbol{\lambda}_1} \frac{w_1}{2}||\boldsymbol{\mu}_1 - \boldsymbol{\lambda}_1||^2 + \sum_{k \in S^*(\vec{\boldsymbol{\mu}})} \frac{w_k}{2} \min_{j \in [d]} (\boldsymbol{\mu}_k^j - \boldsymbol{\lambda}_1^j)_+^2$$

and that is

$$\min_{\phi : S^*(\vec{\boldsymbol{\mu}}) \to [d]} \underbrace{\min_{\boldsymbol{\lambda}_1} \frac{w_1}{2}||\boldsymbol{\mu}_1 - \boldsymbol{\lambda}_1||^2 + \sum_{k \in S^*(\vec{\boldsymbol{\mu}})} \frac{w_k}{2} (\boldsymbol{\mu}_k^{\phi(k)} - \boldsymbol{\lambda}_1^{\phi(k)})_+^2}_{\text{separable convex problem}}$$

The heart of the insight

The cost for moving point 1 onto the front is:

$$\min_{\boldsymbol{\lambda}_1} \; \frac{w_1}{2}||\boldsymbol{\mu}_1 - \boldsymbol{\lambda}_1||^2 + \sum_{k \in S^*(\vec{\boldsymbol{\mu}})} \frac{w_k}{2} \min_{j \in [d]} (\boldsymbol{\mu}_k^j - \boldsymbol{\lambda}_1^j)_+^2$$

and that is

$$\min_{\phi : S^*(\vec{\boldsymbol{\mu}}) \to [d]} \underbrace{\min_{\boldsymbol{\lambda}_1} \; \frac{w_1}{2}||\boldsymbol{\mu}_1 - \boldsymbol{\lambda}_1||^2 + \sum_{k \in S^*(\vec{\boldsymbol{\mu}})} \frac{w_k}{2} (\boldsymbol{\mu}_k^{\phi(k)} - \boldsymbol{\lambda}_1^{\phi(k)})_+^2}_{\text{separable convex problem}}$$

Not all $\phi : S^*(\vec{\mu}) \to [d]$ need to be attempted.
Only $\binom{K+d-1}{d-1}$ due to geometry of $\mathbb{R}^d$.

Outline

Conclusion

With that, everything slots in place and we obtain an algorithm for Pareto Front Identification with

- asymptotically optimal sample complexity
- **polynomial time** cost per round

Now interested in going beyond

- Gaussian
- $\epsilon = 0$
- independence

Thanks!

References i

c Degenne, R. and W. M. Koolen (Dec. 2019). **"Pure Exploration with Multiple Correct Answers".** In: *Advances in Neural Information Processing Systems (NeurIPS) 32*.

c Degenne, R., W. M. Koolen, and P. Ménard (Dec. 2019). **"Non-Asymptotic Pure Exploration by Solving Games".** In: *Advances in Neural Information Processing Systems (NeurIPS) 32*.

c Garivier, A. and E. Kaufmann (June 2016). **"Optimal Best Arm Identification with Fixed Confidence".** In: *29th Annual Conference on Learning Theory*. Vol. 49. Proceedings of Machine Learning Research.

c Kone, C., E. Kaufmann, and L. Richert (2023). **"Adaptive Algorithms for Relaxed Pareto Set Identification".** In: *arXiv preprint arXiv:2307.00424*.