Online optimization and machine learning: Applications to resource allocation problems in wireless networks

E. Veronica Belmega[1,2]

[1]Université Gustave Eiffel, CNRS, LIGM
[2]ETIS, CY Cergy-Paris University, ENSEA, CNRS

January 6th, 2024

Collaborators and students

P. Mertikopoulos (CNRS, LIG)
I. Fijalkow (ENSEA, ETIS)
A. Savard (IMT Nord Europe)
R. Negrel (ESIEE Paris, LIGM)
M. Debbah (Khalifa Univ. of Sci. and Technol.)

A. Marcastel (PhD)
H. El Hassani (PhD)
I. Chafaa (PhD)
O. Bilenne (PostDoc)

Supported by

Chair IoT Orange with Univ. Cergy Pontoise foundation, ANR ORACLESS, ANR ELIOT, GdR ISIS, ENSEA, . . .

Based on

E.V. Belmega, P. Mertikopoulos, and R. Negrel, "Online convex optimization in wireless networks and beyond: The feedback - performance trade-off", ***invited paper*** *at RAWNET intl. workshop in conjunction with WiOpt*, Sep. 2022.

I. Chafaa, R. Negrel, E.V. Belmega, and M. Debbah, "Self-supervised deep learning for mmWave beam steering exploiting sub-6 GHz channels", *IEEE Trans. on Wireless Commun.*, Mar. 2022.

I. Online resource optimization policies: feedback vs. performance

- Preliminaries
- First order (gradient) feedback
- Zeroth-order (scalar) feedback
- 1-bit feedback

II. Deep learning vs. online policies

- mmWave beamforming from sub-6GHz channels

I. Online resource optimization policies: feedback vs. performance

- Preliminaries
- First order (gradient) feedback
- Zeroth-order (scalar) feedback
- 1-bit feedback

II. Deep learning vs. online policies

- mmWave beamforming from sub-6GHz channels

Distributed and dynamic wireless networks

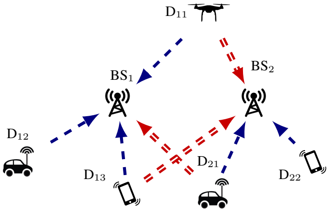
B5G, 6G, IoT [Saad'19, Tataria'21]

- **Characteristics and requirements:** dense, heterogeneous, autonomous, decentralized, energy-efficient, . . .
- **Challenges**

  Highly *mobile* radios & networks
  Unpredictable and *arbitrary* connectivity patterns
  Limited network/channel knowledge (*potentially outdated*)
  Limited processing capabilities

- Static or stochastic (optimal or Nash) solutions: not suitable
  $\Longrightarrow$ **dynamic, non-equilibrium** solutions

Obj: Design **energy-efficient** resource allocation policies coping with **arbitrary** network dynamics and **scarce** feedback



M transmitting devices, N receivers, S orthogonal bands

Shannon rate of an arbitrary device $R_t(\mathbf{p}(t)) = \frac{1}{S}\sum_{s=1}^{S} \log(1 + \overbrace{w_s(t)p_s(t)}^{SNR_s})$

$\mathbf{p}(t)$ - power allocation vector

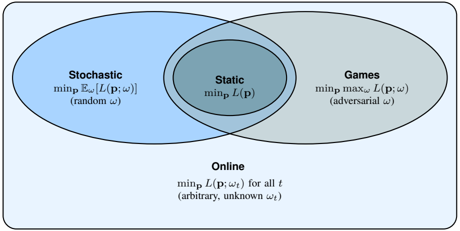
Effective channel gain $w_s(t) = \dfrac{|h_s(t)|^2}{\sigma^2 + \underbrace{\sum_j |h_s^j(t)|^2 p_s^j(t)}_{\text{interference}}}$

$$\text{minimize} \quad L_t(\mathbf{p}(t)) \triangleq \overbrace{\sum_{s=1}^{S} p_s(t)}^{\text{power consumption}} + \underbrace{\lambda\ [R_{\min} - R_t(\mathbf{p}(t))]^+}_{\text{minimum rate penalty}}$$

$$\text{s.t.} \quad p_s(t) \geq 0, \forall s,\ \sum_{s=1}^{S} p_s(t) \leq P_{\max}$$

Rk: **Arbitrary time-varying objective**, unknown at the transmission instant

- How to optimize an unknown $L(\mathbf{p}, \omega_t)$ objective at the decision time t?



Online iterative process

For $t = 1$ to T

- Choose policy $\mathbf{p}(t)$
- Incur loss $L(\mathbf{p}(t), \omega_t)$
- Receive **feedback**
- **Update**: $\mathbf{p}(t+1) \leftarrow \mathbf{p}(t)$ based on feedback

- **Goal:** develop *efficient* online processes based on strictly causal feedback
- Link with machine learning: **multi-armed bandits** (MABs) reinforcement learning
- **Assumptions:** on the objective function $L(\mathbf{p}, \omega_t)$ w.r.t. $\mathbf{p}$ but **not** on the underlying dynamics of ω_t

- Ideal benchmark

$$\forall t, \quad L_t(\mathbf{p}(t)) \quad - \quad \underbrace{\min_{\mathbf{q}\in\mathcal{P}} L_t(\mathbf{q})}_{\text{dynamic optimal loss}}$$

Too ambitious under the worse-case assumption of an arbitrarily time-varying network!

- Less ambitious benchmark: **regret** [Hannan'57]

$$\underbrace{Reg_T}_{\text{overall regret}} \triangleq \sum_{t=1}^{T} L_t(\mathbf{p}(t)) \quad - \quad \underbrace{\min_{\mathbf{q}\in\mathcal{P}} \sum_{t=1}^{T} L_t(\mathbf{q})}_{\text{optimal overall loss of \textbf{fixed} policies}}$$

- **Rk:** both benchmarks require knowledge of the future!

Objective: design online policies $\mathbf{p}(t)$ that lead to **no regret**:

$$\limsup_{T\to\infty} \frac{1}{T} Reg_T \leq 0.$$

No-regret policies

Static convex optimization $L_t(\mathbf{p}) = L(\mathbf{p}), \forall t$

$\mathbf{p}(t)$ no-regret policy $\Rightarrow \overline{\mathbf{p}}(T) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{p}(t)$ converges to the optimal solution

Stochastic convex optimization $L_t(\mathbf{p}) = L(\mathbf{p}, \omega_t), \forall t$

$\mathbf{p}(t)$ no-regret policy $\Rightarrow \overline{\mathbf{p}}(T) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{p}(t)$ converges to the optimal solution

Non-cooperative games [Viossat'13, Mertikopoulos'19]

A no-regret policy (cumulative) converges to the Nash equilibrium in
two-player (discrete or convex) **zero-sum** games, potential (discrete or convex) games, ...

Adversarial (non stochastic) MABs

- S arms of slot machines
- **Variable:** probability distribution $\mathbf{p}(t)$

$$\mathbf{p}(t) \in \mathcal{S} \triangleq \left\{ \mathbf{p} \in \mathbb{R}_{+}^{S} : \sum_{s=1}^{S} p_s = 1 \right\}$$

- **Objective:** maximize the linear expected gains $U_t(\mathbf{p}(t)) = \mathbf{r}(t)^T \mathbf{p}(t)$
  $r(t)$ - vector of arms' rewards at time t
- **Optimal policy**: exponential/multiplicative weights [Auer'02]

Our online power allocation problem

- S channels or frequency subcarriers
- **Variable:** power allocation vector $\mathbf{p}(t)$

$$\mathbf{p}(t) \in \mathcal{P} \triangleq \left\{ \mathbf{p} \in \mathbb{R}_{+}^{S} : \sum_{s=1}^{S} p_s \leq P_{\max} \right\}$$

- **Objective:** minimize the convex loss $L_t(\mathbf{p}(t))$ - tradeoff power consumption vs. rate
- **Hint:** adapted version of exponential weights

I. Online resource optimization policies: feedback vs. performance

- Preliminaries
- First order (gradient) feedback
- Zeroth-order (scalar) feedback
- 1-bit feedback

II. Deep learning vs. online policies

- mmWave beamforming from sub-6GHz channels

Gradient-based online policies

Low complexity, parallel computing

Online Gradient Descent (OGD) **[Zinkevich'03]**

- Gradient feedback: $\mathbf{v}(t) = -\nabla L_t(\mathbf{p}(t))$
- Update: Euclidean projection of $\mathbf{p}(t) + \gamma \mathbf{v}(t)$

$$\begin{aligned} \mathbf{p}(t+1) &\triangleq \arg\min_{\mathbf{q}\in\mathcal{P}} \ \tfrac{1}{2}\left\|\mathbf{p}(t) + \gamma\mathbf{v}(t) - \mathbf{q}\right\|^2 \\ &= \arg\min_{\mathbf{q}\in\mathcal{P}} \ \left\{\gamma\mathbf{v}(t)^T(\mathbf{p}(t) - \mathbf{q}) + \tfrac{1}{2}\left\|\mathbf{p}(t) - \mathbf{q}\right\|^2\right\} \end{aligned}$$

Based on the mirror descent [Nemirovski'83]

Online Mirror Descent (OMD) **[Shalev-Shwartz'07]**

- Gradient feedback: $\mathbf{v}(t) = -\nabla L_t(\mathbf{p}(t))$
- Update: Mirror-mapping on $\mathbf{p}(t) + \gamma \mathbf{v}(t)$

$$\mathbf{p}(t+1) \triangleq \arg\min_{\mathbf{q}\in\mathcal{P}} \left\{\gamma\mathbf{v}(t)^T(\mathbf{p}(t) - \mathbf{q}) + D_h(\mathbf{q}, \mathbf{p}(t))\right\}$$

$D_h(\mathbf{q}, \mathbf{p}) \triangleq h(\mathbf{q}) - h(\mathbf{p}) - \nabla h(\mathbf{p})^T(\mathbf{q} - \mathbf{p})$ the Bregman divergence of regularizer $h(\cdot)$

Online Mirror Descent (OMD)

- Particular cases
  - OGD: Euclidean regularizer $h(\mathbf{p}) = \frac{1}{2}\|\mathbf{p}\|^2, \;\; D_h(\mathbf{q}, \mathbf{p}) = \frac{1}{2}\|\mathbf{q} - \mathbf{p}\|^2$
  - Exponential weights for MAB:

    entropic regularizer $h(\mathbf{p}) = \sum_{s=1}^{S} p_s \log p_s$, Kullback-Leibler $D_h(\mathbf{q}, \mathbf{p}) = \sum_{s=1}^{S} q_s \log \frac{q_s}{p_s}$
- MAB (simplex) regret performance: **scalability**
  - OGD: $Reg_T = \mathcal{O}\left(\sqrt{T\ S}\right)$ [Zinkevich'03]
  - Exponential weights: $Reg_T = \mathcal{O}\left(\sqrt{T \log S}\right)$ [Auer'02]

OMD allows to design algorithms that are **tailored to the feasible set** compared with OGD.

OXL based on gradient feedback

Main idea: we exploit the similarity of $\mathcal{P}$ with the simplex and the entropic regularizer

Online exponential learning (OXL)

- Gradient feedback: $\mathbf{v}(t) = -\nabla L_t(\mathbf{p}(t))$
- Update:

$$\begin{aligned} \mathbf{y}(t+1) &= \mathbf{y}(t) + \gamma \mathbf{v}(t) && \text{cumulative gradient score} \\ p_s(t+1) &= P_{\max} \frac{\exp(y_s(t+1))}{1 + \sum_{i=1}^{S} \exp(y_i(t+1))}, \ \forall s && \text{exponential mapping} \end{aligned}$$

- Similar to the exponential weights for MAB
- Closed-form update as opposed to OGD, which requires an Euclidean projection
- We proved: **no-regret** property, $Reg_T \leq P_{\max} \sqrt{2V \log(1+S)T}$, $V = \max \|v(t)\|^2$

  Extends to the imperfect gradient feedback case: $\tilde{\mathbf{v}}(t) = \mathbf{v}(t) +$ **error**

A. Marcastel, E. V. Belmega, P. Mertikopoulos, and I. Fijalkow, "Online power optimization in feedback-limited, dynamic and unpredictable IoT networks", IEEE Trans. on Signal Processing, 2019.

I. Online resource optimization policies: feedback vs. performance

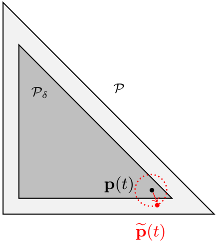
- Preliminaries
- First order (gradient) feedback
- Zeroth-order (scalar) feedback
- 1-bit feedback

II. Deep learning vs. online policies

- mmWave beamforming from sub-6GHz channels

Can the feedback be reduced to a scalar?

- **Received feedback:** value of incurred loss $L_t(\mathbf{p})$ as in MABs with bandit feedback
- Stochastic gradient approximation: $\tilde{\mathbf{v}}(t) = \frac{S}{\delta} L_t(\mathbf{p} + \delta\mathbf{u})\,\mathbf{u}$, [Spall'97, Flaxman'05]
  $\mathbf{u}$ - uniformly drawn over the unit S-dimensional Sphere
- To estimate $\nabla L_t(\mathbf{p}(t))$, transmit at $\tilde{\mathbf{p}}(t) = \mathbf{p}(t) + \delta\mathbf{u} \longrightarrow \tilde{\mathbf{p}}(t)$ may fall outside $\mathcal{P}$!
- Our solution: **shrink the feasible set** s.t. $\forall\, \mathbf{p}(t) \in \mathcal{P}_\delta \subset \mathcal{P} \;\Rightarrow\; \tilde{\mathbf{p}}(t) \in \mathcal{P}$



$$\mathcal{P}_\delta \triangleq \left\{ \mathbf{p} \in \mathbb{R}_+^S : p_s \geq \delta, \sum_{s=1}^{S} p_s \leq P_{\max} - \sqrt{S}\delta \right\}$$

A. Marcastel, E. V. Belmega, P. Mertikopoulos, and I. Fijalkow, "Online power optimization in feedback-limited, dynamic and unpredictable IoT networks", IEEE Trans. on Signal Processing, 2019.

Our modified OXL_0

OXL_0 algorithm

- Transmit at $\tilde{\mathbf{p}}(t) = \mathbf{p}(t) + \delta \mathbf{u}$
- Receive **scalar feedback:** $L_t(\tilde{\mathbf{p}}(t))$
- Gradient estimation: $\tilde{\mathbf{v}} = -\frac{S}{\delta} L_t(\tilde{\mathbf{p}}(t))\, \mathbf{u}$
- **Update:**

$$\mathbf{y}(t+1) = \mathbf{y}(t) + \gamma \tilde{\mathbf{v}}(t)$$

cumulative score

$$p_s(t+1) = \delta + P_{\max}(1 - C_\delta) \frac{\exp(y_s(t+1))}{1 + \sum_{i=1}^{S} \exp(y_i(t+1))}, \quad \forall s$$

modified exponential mapping

adapted to the shrunk set $\mathcal{P}_\delta$

$C_\delta = \frac{\delta}{P_{\max}}(S + \sqrt{S})$

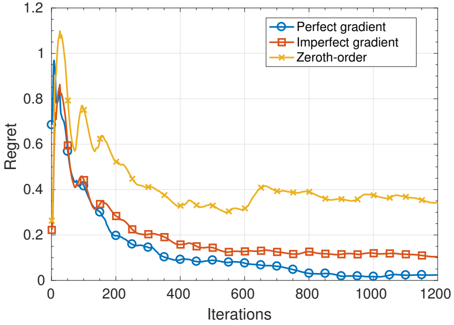
- We proved: **no-regret property**, $EReg_T = \mathcal{O}(T^{3/4})$
- Zeroth-order feedback greatly impacts the decay rate of the average regret

Tradeoff: regret decay rate vs. amount of feedback!

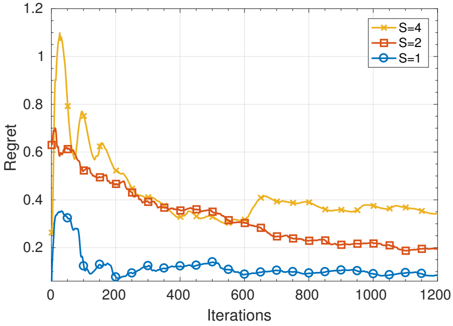
A. Marcastel, E. V. Belmega, P. Mertikopoulos, and I. Fijalkow, "Online power optimization in feedback-limited, dynamic and unpredictable IoT networks", IEEE Trans. on Signal Processing, 2019.

Impact of reducing the feedback

OXL vs. OXL_0



OXL_0 and S - problem dimension



Zeroth-order information reduces the decay rate of the average regret.
The decay rate becomes worse by increasing S - the problem dimensionality.

COST-HATA wireless channels, $M = 10$ transmitting devices, $S \in \{1, 2, 4\}$ bands, $N = 1$ receiver, $P_{\max} \in [0.5, 2]$, $R_{\min} \in [0.5, 3]$ bps/Hz, $\lambda \in [0.5, 10]$, average regret over 200 random draws of $\mathbf{u}$

Zeroth-order feedback with callbacks

- **Issues:** slow decay rate $\mathcal{O}(T^{3/4})$, poor scaling with problem dimensionality

  $\longrightarrow$ MXL0 performs poorly in MIMO networks

- Motivation: exponential weights yield $\mathcal{O}(\sqrt{T})$ regret in bandit feedback MABs

- Our idea: exploit the current feedback R_t jointly with the past R_{t-1}

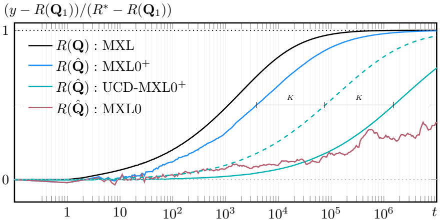
  $\longrightarrow$ improved two-point gradient estimator

- For **static convex objectives**, we can recover the $\mathcal{O}(\sqrt{T})$ regret as with perfect gradient feedback

K-user MIMO multiple access channel

$$R(\mathbf{Q}) = \log\det\left(\mathbf{I} + \sum_{k=1}^{K} \mathbf{H}_k \mathbf{Q}_k \mathbf{H}_k^{\dagger}\right)$$

$$\mathbf{Q}_k \succeq 0, \quad \mathrm{Tr}(\mathbf{Q}_k) \leq P_{\max}$$



Arbitrary static channels, $K = 20$ users, 3 transmit and 16 receive antennas

O. Bilenne, P. Mertikopoulos, and E.V. Belmega, "Fast Gradient-Free Optimization in Distributed Multi-User MIMO Systems", IEEE Trans. on Signal Processing, 2020.

I. Online resource optimization policies: feedback vs. performance

- Preliminaries
- First order (gradient) feedback
- Zeroth-order (scalar) feedback
- 1-bit feedback

II. Deep learning vs. online policies

- mmWave beamforming from sub-6GHz channels

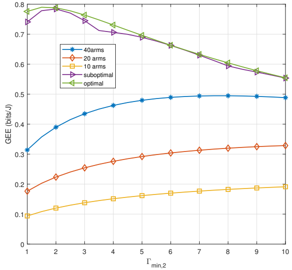
Can the feedback be reduced to one bit?

- **Our idea:** exploit MABs with outage (QoS) based rewards and ACK/NACK-type feedback
- Problems: beam alignment in mmWave networks, NOMA with no CSIT/CDIT
- **Issue:** feasible set needs to be quantized

  $\longrightarrow$ performance loss

2-user downlink NOMA

$$\mathrm{GEE}(i, \mathbf{p}) \triangleq \frac{(R_{\min,i}+R_{\min,j})(1-\mathbb{P}_{\text{out}}(i,\mathbf{p}))}{p_i+p_j+P_c}$$



$\Gamma_{\min,2} = 2^{R_{\min,2}} - 1$, $\sigma_k^2 = 0.1$, $R_{\min,1} = 1$ bpcu, $P_c = 1$ W, $P_{\max} = 100$ W, $T = 5000$, 10^3 channel (Rayleigh) realizations

H. El Hassani, A. Savard, and E.V. Belmega, "Adaptive NOMA in time-varying wireless networks with no CSIT/CDIT relying on a 1-bit feedback", IEEE Wireless Commun. Lett., 2021.

- **Online optimization and no-regret learning:** a suitable framework for resource allocation problems in dynamic and unpredictable networks

- **Assumptions:** Convex (extends to fractional programs) and Lipschitz objectives $L(\mathbf{x}, \omega_t)$ w.r.t. $\mathbf{x}$, convex feasible sets

  $\Rightarrow$ **Adaptive online algorithms**

  Decentralized and reinforcing
  Cope with **arbitrary network dynamics**
  Rely on **strictly causal** information
  Imperfect and **reduced feedback**
  Regret-based theoretical **guarantees**

- Tradeoff: performance vs. available feedback

- Applies to semi-definite programming: in **massive MIMO** networks reducing the (matrix) gradient feedback is crucial

- Zeroth-order feedback with callbacks for **online** problems
- One-bit feedback policies: improve the performance, NOMA cope with continuous and discrete variables
- Beyond convex, Lipschitz objectives: more realistic (**non convex**) energy-efficiency measures
- Compare online algorithms with **deep learning** ones

I. Online resource optimization policies: feedback vs. performance

- Preliminaries
- First order (gradient) feedback
- Zeroth-order (scalar) feedback
- 1-bit feedback

II. Deep learning vs. online policies

- mmWave beamforming from sub-6GHz channels

Model-based approaches tradeoff

simple (unrealistic) but tractable problems **vs.** *practical but not tractable problems*

$\Longrightarrow$ machine (deep) learning has the potential to bridge this gap

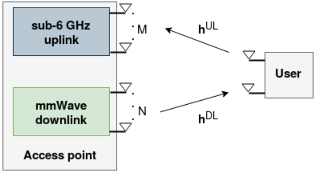
Deep learning based on neural networks [Zappone'19, Tataria'21]

+ Powerful toolbox: *universal approximators* of complex relationships
+ Pervasive in 6G network design, operation, ...
− Large and relevant training datasets
− High computing/processing capabilities

mmWave communications

- Exploit high frequency spectrum
- MIMO systems and **beam alignment** against severe pathloss
- **Challenges:** channel estimation, device mobility and network dynamics

  $\longrightarrow$ online optimization (MABs) and deep learning [Hashemi'18, Alrabeiah'20]
- **Idea:** Mapping sub-6GHz channels into mmWave beamforming vectors [Alrabeiah'20]

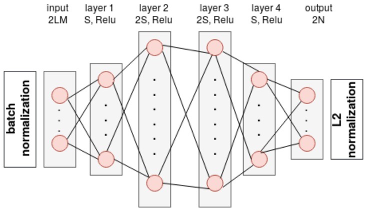


- Sub-6 GHz channels provide multipath signature, easier to estimate
- $\mathbf{h}^{\mathrm{UL}} \to \mathbf{f}^{\mathrm{DL}}$ highly complex and non linear mapping $\longrightarrow$ deep learning

Network architecture and DeepMIMO

Neural network architecture

- Input: uplink sub-6 GHz channels
- Output: mmWave beamforming vectors
- **Fully-connected**: structure agnostic

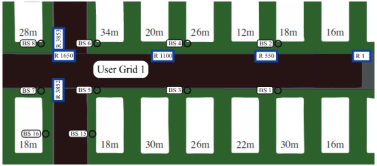


Training

- DeepMIMO dataset: $\left(\{\mathbf{h}_i[\ell]^{\mathrm{UL}}\}_{\ell=1}^{L}, \{\mathbf{h}_i[\ell]^{\mathrm{DL}}\}_{\ell=1}^{L}\right)_i$
- Custom loss function: $\mathcal{L} = -1/\mathcal{B}\sum_{i=1}^{\mathcal{B}} \mathcal{R}_i$

$$\mathcal{R}_i = \frac{1}{L}\sum_{\ell=1}^{L} \log_2\left(1 + \frac{P^{\mathrm{DL}}}{L\,(\sigma^{\mathrm{DL}})^2} \mid \mathbf{h}_i^{\mathrm{DL}\dagger}[\ell]\,\mathbf{f}_i\mid^2\right)$$

$\mathbf{f}_i$: predicted beamformer, P^{DL} transmit power, $\mathcal{B}$ size of mini-batch, L subcarriers, $(\sigma^{\mathrm{DL}})^2$ noise variance



DeepMIMO setup [Alkhateeb'19]

I. Chafaa, R. Negrel, E.V. Belmega, and M. Debbah, "Unsupervised deep learning for mmWave beam steering exploiting sub-6 GHz channels", IEEE Trans. on Wireless Commun., 2022.

Deep learning or MABs?

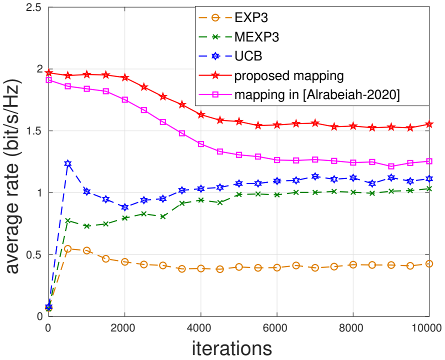
MABs

- \+ No *offline* training, online learning
- \+ Low complexity $\sim 10^2$ ops/iteration
- \+ 1-bit feedback (ACK/NACK)
- − Quantized beams $\longrightarrow$ performance loss
- − Trial and error $\longrightarrow$ transitory phase

Deep learning

- − Offline training on large and relevant datasets, *DeepMIMO* [Alkhateeb'19]
- − Higher complexity $\sim 10^6 - 10^7$ ops/iteration
- − sub-6 GHz channel information
- \+ Regression (no quantization)
- \+ Better running performance (no transitory phase)

Depends on the target application, its characteristics and requirements



MABs: EXP3, MEXP3, UCB

Deep learning [Alrabeiah'20]: classification

I. Chafaa, R. Negrel, E.V. Belmega, and M. Debbah, "Self-supervised deep learning for mmWave beam steering exploiting sub-6 GHz channels", IEEE Trans. on Wireless Commun., 2022.
I. Chafaa, E.V. Belmega, and M. Debbah, "One-bit Feedback exponential learning for beam alignment in mobile mmWave", IEEE Access, Oct. 2020.

- Deep learning approaches capture **complex relationships**:
  sub-6GHz uplink channels – downlink mmWave beams
- Extension to multi-cell multi-user networks
- **Robustness** to data imperfections
- **Generalization** capability: other wireless settings, problems
- Recurrent networks, reinforcement and online deep learning to capture the temporal network dynamics

AI-enhanced highly mobile and unpredictable IoT networks

OBJ-1 Design efficient **online optimization** algorithms
requiring **low-cost** and energy-efficient communication feedback

OBJ-2 Design online algorithms maximizing **non convex** energy efficiency
exploiting non convex online optimization and deep learning

Sustainable wireless communications: low-energy, low-cost and zero added electromagnetic waves

OBJ-1 Derive the fundamental **achievable Shannon rates**
in multi-user, multi-backscatter/RIS networks

OBJ-2 Develop **efficient algorithms** that jointly tune the transmit strategy
and the backscattering/RIS strategy

More intel on my webpage:
https://sites.google.com/site/evbelmega

E.V. Belmega, P. Mertikopoulos, and R. Negrel, “Online convex optimization in wireless networks and beyond: The feedback - performance trade-off”, ***invited paper*** *at RAWNET intl. workshop in conjunction with WiOpt*, Sep. 2022.

O. Bilenne, P. Mertikopoulos, and E.V. Belmega, “Fast Gradient-Free Optimization in Distributed Multi-User MIMO Systems”, *IEEE Trans. on Signal Processing*, Oct. 2020.

A. Marcastel, E. V. Belmega, P. Mertikopoulos, and I. Fijalkow, “Gradient-free online resource allocation algorithms for dynamic wireless networks”, ***invited paper*** *to IEEE SPAWC*, 2019.

A. Marcastel, E.V. Belmega, P. Mertikopoulos, and I. Fijalkow, "Online power optimization in dynamic IoT networks: The impact of feedback scarcity", *IEEE Trans. on Signal Processing*, Mar. 2019.

P. Mertikopoulos, and E.V. Belmega, “Learning to be green: robust energy efficiency maximization in dynamic MIMO-OFDM systems”, *IEEE Journal on Selected Areas in Communication*, Apr. 2016.

H. El Hassani, A. Savard, and E.V. Belmega, “Energy-efficient 1-bit feedback NOMA in wireless networks with no CSIT/CDIT”, *IEEE SSP Workshop*, Jun. 2021.

H. El Hassani, A. Savard, and E.V. Belmega, “Adaptive NOMA in time-varying wireless networks with no CSIT/CDIT relying on a 1-bit feedback”, *IEEE Wireless Commun. Lett.*, Apr. 2021.

I. Chafaa, R. Negrel, E.V. Belmega, and M. Debbah, “Self-supervised deep learning for mmWave beam steering exploiting sub-6 GHz channels”, *IEEE Trans. on Wireless Commun.*, 2022.

I. Chafaa, E.V. Belmega, and M. Debbah, “One-bit Feedback exponential learning for beam alignment in mobile mmWave”, *IEEE Access*, Oct. 2020.

T. Chen, S. Barbarossa, X. Wang, G. B. Giannakis, and Z.-L. Zhang, “Learning and management for Internet-of-Things: Accounting for adaptivity and scalability”, Arxiv preprint arxiv:1810.11613, 2018.

J. Hannan, “Approximation to Bayes risk in repeated play”, *Contributions to the Theory of Games, Vol. III, fPrinceton University Press*, vol. 39, pp. 97–139, 1957.

Y. Viossat and A. Zapechelnyuk, “No-regret dynamics and fictitious play”, *Journal of Economic Theory*, vol. 148, no 2, pp. 825–842, 2013.

P. Mertikopoulos, and Z. Zhou, “Learning in games with continuous action sets and unknown payoff functions”, *Mathematical Programming*, vol. 173, no 1–2, pp. 465–507, 2019.

M. Zinkevich, "Online convex programming and generalized infinitesimal gradient ascent, *ICML’03: Proceedings of the 20th International Conference on Machine Learning*, pp. 928–936, 2003.

S. Shalev-Shwartz, "Online learning: Theory, algorithms, and applications," *Ph.D. dissertation, Hebrew University of Jerusalem*, 2007.

A. S. Nemirovski and D. B. Yudin, "Problem Complexity and Method Efficiency in Optimization", *Wiley, New York, NY*, 1983.

P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire, “The non stochastic multi armed bandit problem”, *SIAM Journal on Computing*, vol. 32, pp. 48–77, 2002.

J. C. Spall, “A one-measurement form of simultaneous perturbation stochastic approximation”, *Automatica*, vol. 33, no. 1, pp. 109–112, 1997.

A. D. Flaxman, A. T. Kalai, and H. B. McMahan, "Online convex optimization in the bandit setting: gradient descent without a gradient", *SODA’05: Proceedings of the 16th annual ACM-SIAM symposium on discrete algorithms*, 2005, pp. 385–394.

- C. Zhang, P. Patras, and H. Haddadi, “Deep learning in mobile and wireless networking: A survey”, *Arxiv preprint arxiv:1803.04311*, 2018.
- A. Zappone, M. Di Renzo, and M. Debbah, “Wireless networks design in the era of deep learning: Model-based, AI-based, or both?”, *IEEE Transactions on Communications*, vol. 67, no. 10, pp. 7331–7376, 2019.
- E. C. Strinati, S. Barbarossa, J. L. Gonzalez-Jimenez, D. Ktenas, N. Cassiau, and C. Dehos, “6G: The next frontier: From holographic messaging to artificial intelligence using subterahertz and visible light communication”, *IEEE Vehicular Technology Magazine*, vol. 14, no.3, pp.42–50, 2019.
- W. Saad, M. Bennis, and M. Chen, “A vision of 6G wireless systems: Applications, trends, technologies, and open research problems”, *IEEE Network*, vol. 34, no. 3, pp. 134–142, 2019.
- H. Tataria, M. Shafi, A.F. Molisch, M. Dohler, H. Sj oland, and F. Tufvesson, “6G wireless systems: Vision, requirements, challenges, insights, and opportunities”, *Proceedings of the IEEE*, vol. 109, no. 7, pp. 1166–1199, 2021.
- M. Alrabeiah and A. Alkhateeb, “Deep learning for mmwave beam and blockage prediction using sub-6 GHz channels”, *IEEE Trans. Commun.*, 2020.
- A. Alkhateeb, “DeepMIMO: A generic deep learning dataset for millimeter wave and massive MIMO applications”, arXivpreprint arXiv:1902.06435, 2019.
- X. Song, S. Haghighatshoar, and G. Caire, “A scalable and statistically robust beam alignment technique for mm-Wave systems”, IEEE Trans. Wireless Commun., vol. 17, no. 7, pp. 4792–4805, 2018.
- J.B. Wang, M. Cheng, J. Y. Wang, M. Lin, Y. Wu, H. Zhu, and J. Wang, “Bandit inspired beam searching scheme for mmWave high-speed train communications”, *arXiv preprint arXiv:1810.06150*, 2018.
- M. Hashemi, A. Sabharwal, C.E. Koksal, and N.B. Shroff, “Efficient beam alignment in millimeter wave systems using contextual bandits”, *IEEE Conference on Computer Communications (INFOCOM), 2018.*