KTH Royal Institute of Technology

Data-based certificates in stochastic Nash games

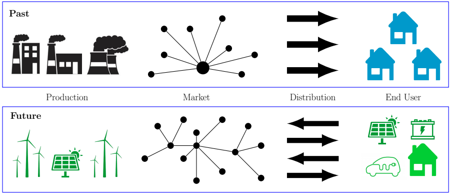
Filippo Fabiani
IMT School for Advanced Studies Lucca

Game On! Seminar series
May 26, 2026 – Stockholm, Sweden

Outline

Outline

Uncertainty and data in modern multi-agent systems



Source: B. Franci

Kroposki et al., "*"Achieving a 100% renewable grid: Operating electric power systems with extremely high levels of variable renewable energy*", IEEE P&E Magazine, 2017

Pandey et al., *"On the needs for MaaS platforms to handle competition in ridesharing mobility"*, Transportation Research 2019

Source: ChatGPT

- Statistics underlying the uncertainty is tipically unknown
- Information available through limited data samples

This talk

- Fabiani, Franci, *"Finite-sample guarantees for data-driven forward-backward operator methods"*, IEEE TAC 2026 (In press)

- Fabiani, Franci, *"Certifying ε-equilibria in stochastic games with limited data availability"*, EJC Special Issue of 2026 ECC (under review)

Prof. B. Franci
Politecnico di Torino

Stochastic Nash games

- N autonomous agents (or decision-makers)
- Each agent i has decision variable (or strategy) x_i
- Each agent i has its local cost $J_i(\cdot, \boldsymbol{x}_{-i}, \xi)$, constraints Ω_i
- $\xi \in \Xi$ is a random variable with **unknown** statistics

$$\forall i : \begin{cases} \min\limits_{x_i \in \Omega_i} & \mathbb{E}_{\mathbb{P}}[J_i(x_i, \boldsymbol{x}_{-i}, \xi)] \\ \text{s.t.} & g(x_i, \boldsymbol{x}_{-i}) \leq 0 \end{cases}$$

We call $\boldsymbol{x}^\star$ a stochastic generalized Nash equilibrium if, for all i,

$$\mathbb{J}_i(x_i^\star, \boldsymbol{x}_{-i}^\star) \leq \min_{y_i \in \mathcal{X}_i(\boldsymbol{x}_{-i}^\star)} \mathbb{J}_i(y_i, \boldsymbol{x}_{-i}^\star)$$

A modern approach to S(G)NE seeking

Under assumptions, one can look for $\boldsymbol{x}^\star$ s.t., for some $\lambda \geq 0$,

$$\forall i : 0 \in \mathbb{E}_{\mathbb{P}}\left[\nabla_{x_i} J_i\left(x_i^\star, \boldsymbol{x}_{-i}^\star, \xi\right)\right] + \mathrm{N}_{\Omega_i}(x_i^\star) + \nabla_{x_i} g\left(x_i^\star, \boldsymbol{x}_{-i}^\star\right)^\top \lambda$$

$$0 \in \mathrm{N}_{\mathbb{R}_{\geq 0}}(\lambda) - g(\boldsymbol{x}^\star)$$

This turns the S(G)NE seeking into a zero-finding problem:

$$0 \in \mathcal{T}(\boldsymbol{x}, \lambda) = \begin{bmatrix} \mathbb{F}(\boldsymbol{x}) + \mathrm{N}_\Omega(\boldsymbol{x}) + \nabla_{\boldsymbol{x}} g(\boldsymbol{x})^\top \lambda \\ \mathrm{N}_{\mathbb{R}_{\geq 0}}(\lambda) - g(\boldsymbol{x}) \end{bmatrix}$$

- Suitable splitting $\mathcal{T} = \mathcal{A} + \mathcal{B} \Rightarrow$ Different iterative schemes for $\boldsymbol{x}^\star$
- Approximation methods for $\mathbb{F}$ based on data samples $\xi^{(j)}$

Main features and potential limitations

	(Boț, 2021)	(Fabiani, 2022)	(Franci, 2021)	(Franci, 2021)	(Gower, 2020)	(Franci, 2020)	(Franci, 2025)	(Iusem, 2017)
Alg. type	FBF	Tik	RFB	RFB	VRG	FB	LEG	EG
Sample(s)	N_k	1	N_k	1	$1 + NB_k(p)$	N_k	$1 + NB_k(p)$	N_k
Convergence	$k \to \infty$	$k \to \infty$	$k \to \infty$	$k \to \infty$	$k \to \infty$	$k \to \infty$	$k \to \infty$	$k \to \infty$
Equilibrium	SNE	SGNE	SGNE	SNE	NE*	SGNE	SGNE*	SGNE

- All algorithms enjoy convergence to an exact S(G)NE
- In all cases, an infinite number of samples is required

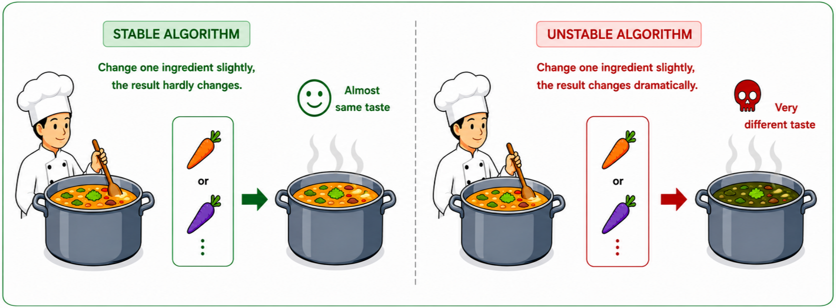
Practical problem considered

How far can we get from some $\boldsymbol{x}^\star$ by running a certain algorithm with an approximation of $\mathbb{F}$ exploiting a fixed and finite dataset?

Spoiler alert: We certify ε-convergence to $\boldsymbol{x}^\star$ and compute ε!

Outline

A cooking-inspired illustration



Source: ChatGPT

More formally

Given a collection of i.i.d. data samples $\mathcal{D}_s = \{\xi^{(j)}\}_{j=1}^s$, define set:

$$\mathcal{D}_s^j = \{\xi^{(1)}, \dots, \xi^{(j-1)}, \xi', \xi^{(j+1)}, \dots, \xi^{(s)}\}$$

- **Algorithm** A returns some hypothesis $H = A(\mathcal{D}_s)$
- Performance are measured by a **loss function** $\ell(H, \xi)$

Definition (Uniform stability)

A possesses β uniform stability w.r.t. ℓ if, for all $j \in \{1, \dots, s\}$,

$$\max_{\xi \in \Xi} |\ell(A(\mathcal{D}_s), \xi) - \ell(A(\mathcal{D}_s^j), \xi)| \leq \beta$$

$\Longrightarrow$ Changing one sample does not significantly alter the output of the algorithm w.r.t. to the metric ℓ!

Exponential bound for uniformly stable algorithms

Attached to a loss function ℓ, we have two performance measures:

Generalization error

$$r(A, s) = \mathbb{E}_{\mathbb{P}}[\ell(H, \xi)]$$

Empirical error

$$\hat{r}(A, s) = \tfrac{1}{s} \textstyle\sum_{j=1}^{s} \ell(H, \xi^{(j)})$$

Theorem (Kutin, 2002)

If A is β-uniformly stable w.r.t. some $\ell(H, \xi) \in [0, \bar{\ell}]$, then

$$r(A, s) \le \hat{r}(A, s) + \beta + (s\beta + \bar{\ell})\sqrt{2\ln(1/\delta)/s}$$

with probability at least $1 - \delta$, $\delta \in (0, 1)$

$\Longrightarrow$ Design of loss function ℓ is crucial!

1. "Meaningful", independent on unknown quantities (e.g., $\boldsymbol{x}^\star$)
2. Guarantee $\beta \propto 1/s$ uniform stability

Outline

The data-driven iterative scheme

$$x^\star \in \mathrm{zer}(\mathcal{T}) \iff x^\star \in \mathrm{zer}(\mathcal{A} + \mathcal{B})$$

- Max. monotone $\mathcal{A}$ is cheap, but $\mathcal{B}$ is unknown (e.g., $\mathbb{F}$)
- Noisy oracle gives measures of $\mathcal{B}$ through $\mathcal{D}_s$, i.e., $\mathcal{O}(\cdot, \xi^{(j)})$
- $\mathcal{O}$ is **unbiased** and inherits the properties possessed by $\mathcal{B}$

Data-driven FB

$$\begin{aligned} y^k &= x^k - \frac{\gamma}{s} \sum_{j=1}^{s} \mathcal{O}(x^k, \xi^{(j)}) \\ x^{k+1} &= \mathrm{J}_{\gamma \mathcal{A}}(y^k) \end{aligned}$$

- Run for K iterations, return hypotesis $\omega^{K+1} = \mathrm{col}(y^K, x^{K+1})$
- Interested in: $\mathrm{zer}_\varepsilon(\mathcal{A} + \mathcal{B}) = \{x : \exists z \in \mathcal{A}(x) \text{ s.t.} \|z + \mathcal{B}(x)\| \leq \varepsilon\}$

A certificate independent on K

- Loss function $\ell(H, \xi) = \|x - \gamma \mathcal{O}(x, \xi) - y\|$
- $\mathcal{B}$ is μ-strongly monotone and κ-Lipschitz continuous

Lemma

With $\tau = \sqrt{1 - \gamma(2\mu - \gamma\kappa^2)} < 1$, *for all* $\xi \in \Xi$, $j \in \{1, \ldots, s\}$,

$$|\ell(A(\mathcal{D}_s), \xi) - \ell(A(\mathcal{D}_s^j), \xi)| \leq \frac{2\gamma M(1 + \tau)}{s(1 - \tau)}$$

Theorem

$x^{K+1} \in \mathrm{zer}_\varepsilon(\mathcal{A} + \mathcal{B})$ *with probability at least* $1 - \delta$, *where*

$$\varepsilon = \frac{1}{\gamma}\hat{r}(x^{K+1}, s) + \frac{2M(1 + \tau)}{s(1 - \tau)} + \left(\frac{2\gamma M(1 + \tau) + \bar{\ell}}{\gamma(1 - \tau)}\right)\sqrt{\frac{2\ln(1/\delta)}{s}}$$

Intuition: $0 = \|\mathrm{J}_{\gamma\mathcal{A}}(x^\star - \gamma\mathcal{B}(x^\star)) - x^\star\| \leq \|\mathrm{J}_{\gamma\mathcal{A}}(x^{K+1} - \gamma\mathcal{B}(x^{K+1})) - x^{K+1}\|$

$$\ldots \leq \mathbb{E}_{\mathbb{P}}[\ell(H, \xi)] = r(A, s)$$

Stochastic Nash game: PEVs charging coordination

- Minimize day-ahead charging schedule:

$$J_i(x_i, \sigma(\boldsymbol{x})) = \|x_i\|_{Q_i}^2 + c_i^\top x_i + \xi^\top x_i + \|\sigma(\boldsymbol{x}) - \bar{\sigma}\|_P^2$$

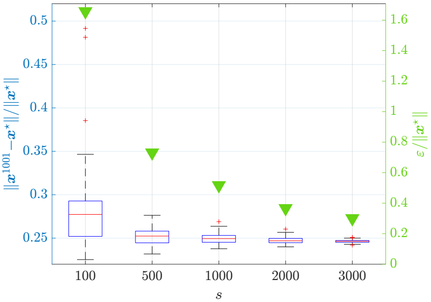
$$\Omega_i = \{x_i \in \mathbb{R}^T : x_i \in [0, \bar{x}_i]^T, \mathbf{1}_T^\top x_i \geq \zeta_i\}$$

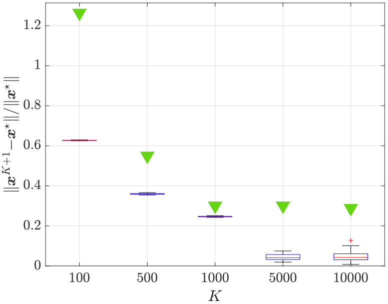
- Stochastic cost of electricity ξ with unknown statistics
- Historical samples $\xi^{(j)}$ taken from entsoe.eu

Iteration k: Agent i receives $\boldsymbol{x}_{-i}^k$, then updates

$$x_i^{k+1} = \mathrm{proj}_{\Omega_i}\left(x_i^k - \frac{\gamma}{s}\sum_{j=1}^{s} \nabla_{x_i} J_i(x_i^k, \boldsymbol{x}_{-i}^k, \xi^{(j)})\right)$$

Stochastic Nash game: PEVs charging coordination (+)





A weaker assumption yields a K-dependent certificate

- Loss function $\ell(H, \xi) = \|x - \gamma \mathcal{O}(x, \xi) - y\|$
- $\mathcal{B}$ is θ-cocoercive, $\theta > 0$, and $\gamma \in (0, 2\theta)$

Lemma

For all $\xi \in \Xi$, $j \in \{1, \ldots, s\}$,

$$|\ell(A(\mathcal{D}_s), \xi) - \ell(A(\mathcal{D}_s^j), \xi)| \leq \frac{4\gamma MK}{s}$$

Theorem

$x^{K+1} \in \mathrm{zer}_{\varepsilon}(\mathcal{A} + \mathcal{B})$ *with probability at least* $1 - \delta$, *where*

$$\varepsilon = \frac{1}{\gamma}\hat{r}(x^{K+1}, s) + \frac{4MK}{s} + \left(\frac{4\gamma MK + \bar{\ell}}{\gamma}\right)\sqrt{\frac{2\ln(1/\delta)}{s}}$$

An example beyond stochastic games

- Solve minimization problem with $P \succcurlyeq 0$ and convex $\mathcal{X}$:

$$\min_{x \in \mathcal{X}} q^\top x + \frac{1}{2} x^\top P x \iff 0 \in \underbrace{q + \bar{P} x}_{=\mathcal{B}(x)} + \underbrace{\mathrm{N}_{\mathcal{X}}(x)}_{=\mathcal{A}(x)}$$

- $\bar{P} = \frac{1}{2}(P + P^\top)$ is the symmetric part of P
- From (Bauschke, 2016), $\mathcal{B}$ is $(1/\lambda_{\max}(\bar{P}))$-cocoercive
- Samples are perturbed matrices $\bar{P} + \mathrm{diag}(\xi^{(i)})$

K	100	500	1000	5000	10000
$\mathrm{avg}(\lVert x^{K+1} - x^\star \rVert)$	0.44	0.02	0.001	0.001	0.001
$\mathrm{avg}(\varepsilon)$	0.49	0.85	1.29	2.37	3.67

Main considerations

Advantages

- Bounds require i.i.d. data only, independent from $\mathcal{D}_s$ and $\mathbb{P}$
- *A-posteriori*, easy-to-compute (for tailored losses)
- Certificate ε vanishes as $O(s^{-1/2})$
- Generalizaton of SGD in optimization (Hardt, 2016)

Limitations

- Carefull choice of the loss function ℓ
- Bounds are not tight in practice (and not expected to be...)

Outline

Including penality terms for SGNE seeking

$$\forall i : \min_{x_i \in \mathbb{R}^{n_i}} \mathbb{J}_i\left(x_i, \boldsymbol{x}_{-i}\right) + \phi(\boldsymbol{x})$$

with barrier function $\phi(\boldsymbol{x}) = \begin{cases} -\varrho \sum_{r=1}^{m} \log\left(-g_r(\boldsymbol{x})\right), & g(\boldsymbol{x}) < 0 \\ +\infty, & \text{otherwise} \end{cases}$

- $\boldsymbol{x} \mapsto \phi(\boldsymbol{x})$ is convex, $\mathcal{C}^2 \Rightarrow \nabla\phi_{\boldsymbol{x}}(\boldsymbol{x})$ monotone, ℓ_ϕ-Lipschitz
- $\varrho > 0$ small to obtain SGNE $\boldsymbol{x}^\star \in \operatorname{zer}(\underbrace{\mathbb{F}(\boldsymbol{x}) + \nabla_{\boldsymbol{x}}\phi(\boldsymbol{x})}_{=\mathbb{B}(\boldsymbol{x})})$
- Noisy oracle gives measures of $\mathbb{B}$ through $\mathcal{D}_s$, i.e., $\mathcal{O}(\cdot, \xi^{(j)})$
- $\mathcal{O}$ is **unbiased** and inherits the properties possessed by $\mathbb{F}$

Data-driven extragradient method

Iteration k**:** Agent i

1. Receives $\boldsymbol{x}_{-i}^k$, then updates

$$y_i^k = x_i^k - \frac{\gamma}{s} \sum_{j=1}^{s} \mathcal{O}(x_i^k, \boldsymbol{x}_{-i}^k, \xi^{(j)})$$

2. Receives $\boldsymbol{y}_{-i}^k$, then updates

$$x_i^{k+1} = x_i^k - \frac{\gamma}{s} \sum_{j=1}^{s} \mathcal{O}(y_i^k, \boldsymbol{y}_{-i}^k, \xi^{(j)})$$

- Run for K iterations, return hypotesis $\mathrm{col}(\boldsymbol{y}^K, \boldsymbol{x}^{K+1})$
- Set of interest: $\mathrm{zer}_\varepsilon(\mathbb{B}) = \{\boldsymbol{x} \in \mathcal{X} : \|\mathbb{B}(\boldsymbol{x})\| \leq \varepsilon\}$

Uniform stability of the EG scheme

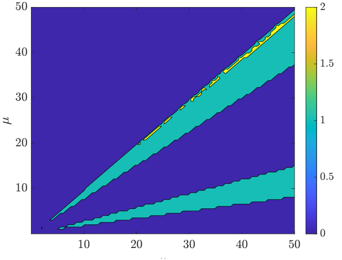
- Loss function $\ell(H, \xi) = \|x - \gamma \mathcal{O}(x, \xi)\|$
- $\mathbb{B}$ is μ-strongly monotone and κ-Lipschitz continuous

Lemma

Choose $\gamma \in (0, 2\mu/\kappa^2)$ *s.t.* $\tau = (1 + \gamma\kappa)\rho + \gamma\kappa < 1$, *with* $\rho = \sqrt{1 - 2\gamma\mu + \gamma^2\kappa^2} < 1$. *Then, for all* $\xi \in \Xi$, $j \in \{1, \ldots, s\}$,

$$|\ell(A(\mathcal{D}_s), \xi) - \ell(A(\mathcal{D}_s)^j, \xi)| \leq \frac{2\gamma\rho M}{s(1 - \tau)}$$

$$\begin{aligned} \min_{\gamma \in (0, 2\mu/\kappa^2)} \quad & -\gamma \\ \text{s.t.} \quad & (1 + \gamma\kappa)\rho + \gamma\kappa < 1 \\ & \rho = \sqrt{1 - 2\gamma\mu + \gamma^2\kappa^2} \end{aligned}$$



Uniform stability of the EG scheme

- Loss function $\ell(H, \xi) = \|x - \gamma \mathcal{O}(x, \xi)\|$
- $\mathbb{B}$ is μ-strongly monotone and κ-Lipschitz continuous

Lemma

Choose $\gamma \in (0, 2\mu/\kappa^2)$ s.t. $\tau = (1 + \gamma\kappa)\rho + \gamma\kappa < 1$, with $\rho = \sqrt{1 - 2\gamma\mu + \gamma^2\kappa^2} < 1$. Then, for all $\xi \in \Xi$, $j \in \{1, \ldots, s\}$,

$$|\ell(A(\mathcal{D}_s), \xi) - \ell(A(\mathcal{D}_s)^j, \xi)| \leq \frac{2\gamma\rho M}{s(1 - \tau)}$$

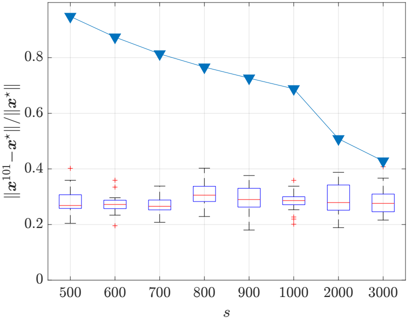
Theorem

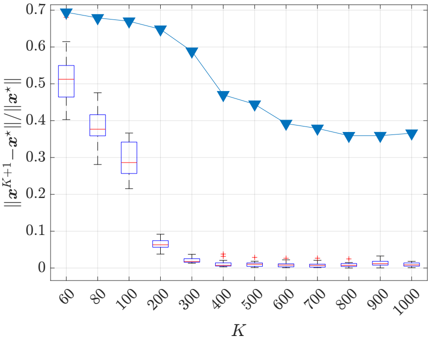
$\boldsymbol{x}^{K+1}$ is an ε-SGNE with probability at least $1 - \delta$, where

$$\varepsilon = \hat{r}(\boldsymbol{x}^{K+1}, s) + \frac{2\gamma\rho M}{s(1 - \tau)} + \frac{M(2\gamma\rho M + 1 - \tau)}{1 - \tau}\sqrt{\frac{2\ln(1/\delta)}{s}}$$

A toy example

- 20 agents, scalar decision variable x_i
- Cost $J_i(x_i, \boldsymbol{x}_{-i}, \xi) = a_i x_i^2 + (b_i + \xi)x_i + (\sum_{j \neq i} c_{i,j} x_j)x_i$
- Local and coupling linear constraints $\mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^N : G\boldsymbol{x} \leq p\}$
- $\xi \sim \mathcal{N}(-1, 0.2)$, other parameters are randomly generated





Outline

Conclusion & Outlook

Take-home messages

- Algorithmic stability offers general purpose tools
  - Minimum requirements on the available dataset
  - Suitable for both static and iterative algorithms
- Stronger assumptions on the operator $\implies$ better bounds
- If $s \to \infty$, $\varepsilon \to 0 \Rightarrow$ recovering exact convergence to S(G)NE

Possible research directions

- Randomized versions (to avoid averaging over s samples)
- Other operator splitting methods, e.g., FBF, DR, ...
- ...

Thanks for your attention!
Questions?

filippo.fabiani@imtlucca.it