Boosting Variational Inference: An Optimization Perspective

Francesco Locatello[1,2], Rajiv Khanna[3], Joydeep Ghosh[3], Gunnar Rätsch[1]

NIPS

Variational Inference & Boosting

Variational Inference

$$\min_{q \in \mathcal{Q}} D^{KL}(q \| p_{\boldsymbol{x}})$$

- Efficient approximate inference
- Poor approximation of the posterior
- Can't trade time with accuracy

Boosting Variational Inference

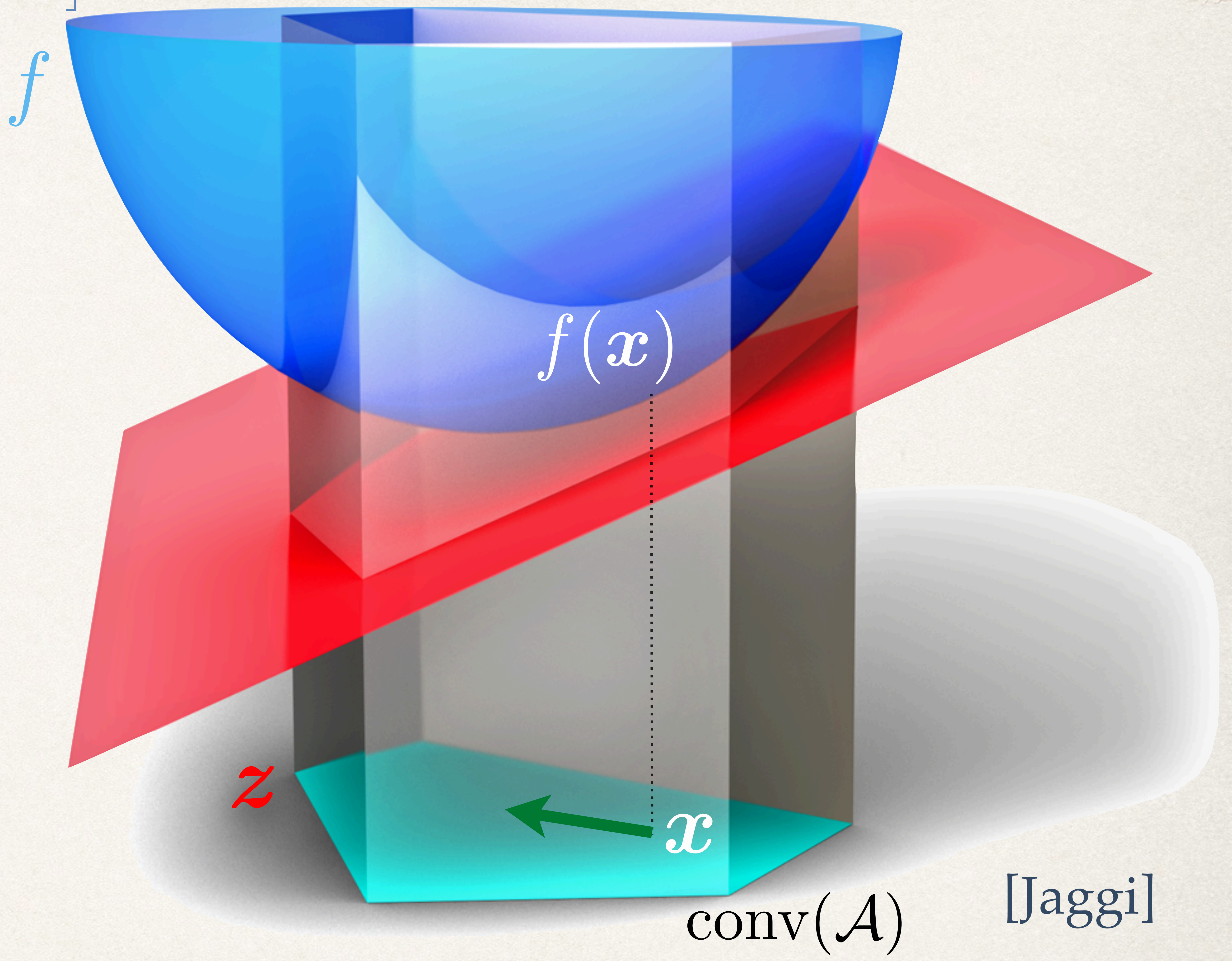
$$\min_{q \in \mathrm{conv}(\mathcal{Q})} D^{KL}(q \| p_{\boldsymbol{x}})$$

- Efficient approximate inference
- Highly expressive family
- Greedily adding components: convex optimization problem

Frank-Wolfe [Frank&Wolfe 1956]

Linearization

$$\min_{z \in \mathcal{A}} f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}), \boldsymbol{z} - x \rangle$$



[Jaggi]

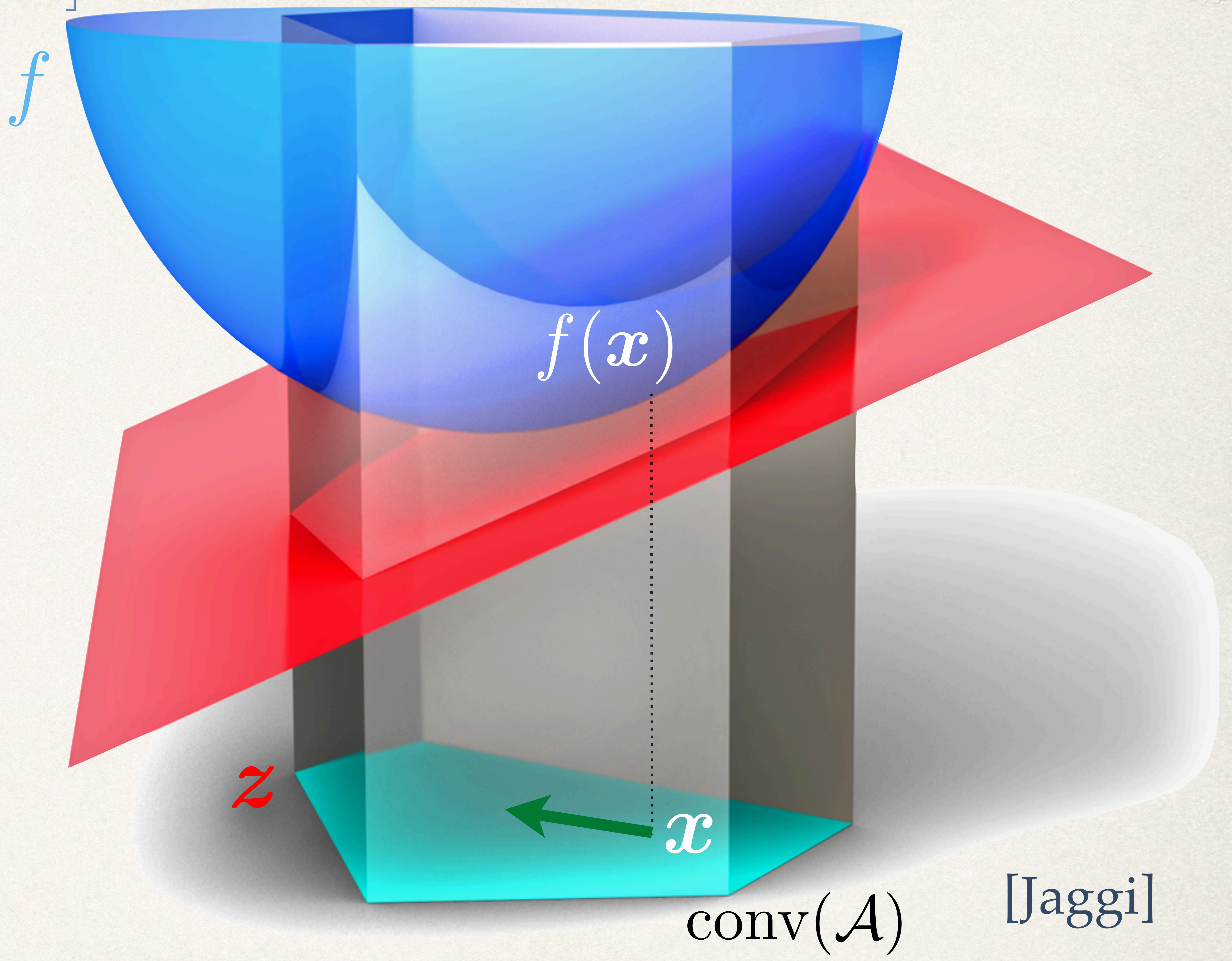
Frank-Wolfe [Frank&Wolfe 1956]

Linearization

$$\min_{z\in\mathcal{A}} f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}), \boldsymbol{z} - x\rangle$$

The Linear Minimization Oracle

$$\mathrm{LMO}_{\mathcal{A}}(\boldsymbol{d}) := \arg\min_{z\in\mathcal{A}} \langle \boldsymbol{d}, \boldsymbol{z}\rangle$$



[Jaggi]

Frank-Wolfe [Frank&Wolfe 1956]

Linearization

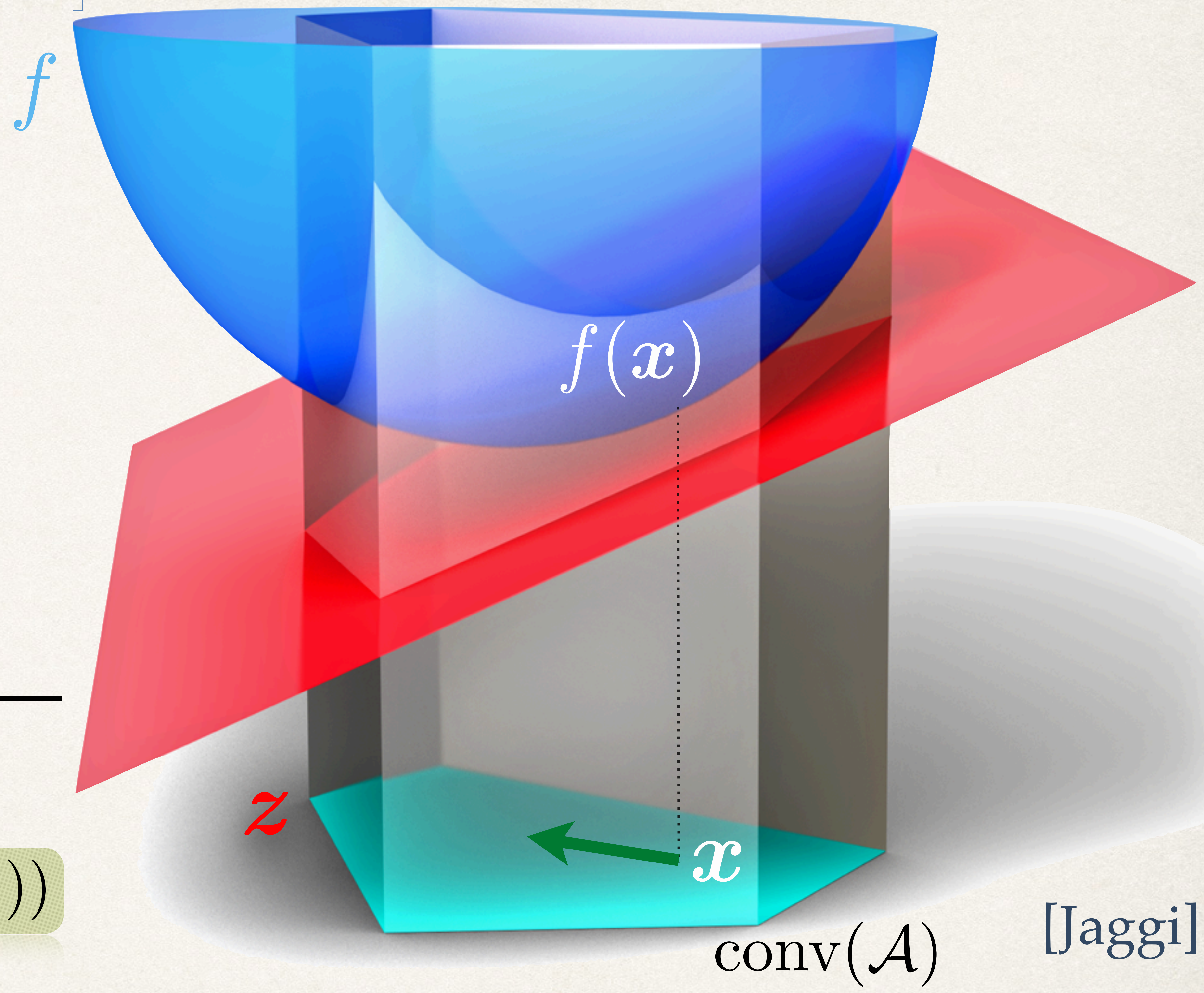
$$\min_{z \in \mathcal{A}} f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}), \boldsymbol{z} - x \rangle$$

The Linear Minimization Oracle

$$\mathrm{LMO}_{\mathcal{A}}(\boldsymbol{d}) := \arg\min_{z \in \mathcal{A}} \langle \boldsymbol{d}, \boldsymbol{z} \rangle$$

---

init $\boldsymbol{x}_0 \in \mathrm{conv}(\mathcal{A})$
for $t = 0 \dots T$
    Find $\boldsymbol{z}_t := (\text{Approx-})\,\mathrm{LMO}_{\mathcal{A}}(\nabla f(\boldsymbol{x}_t))$
    Variant 0: $\gamma = \frac{2}{t+2}$
    Update $\boldsymbol{x}_{t+1} := (1-\gamma)\boldsymbol{x}_t + \gamma \boldsymbol{z}_t$
end for

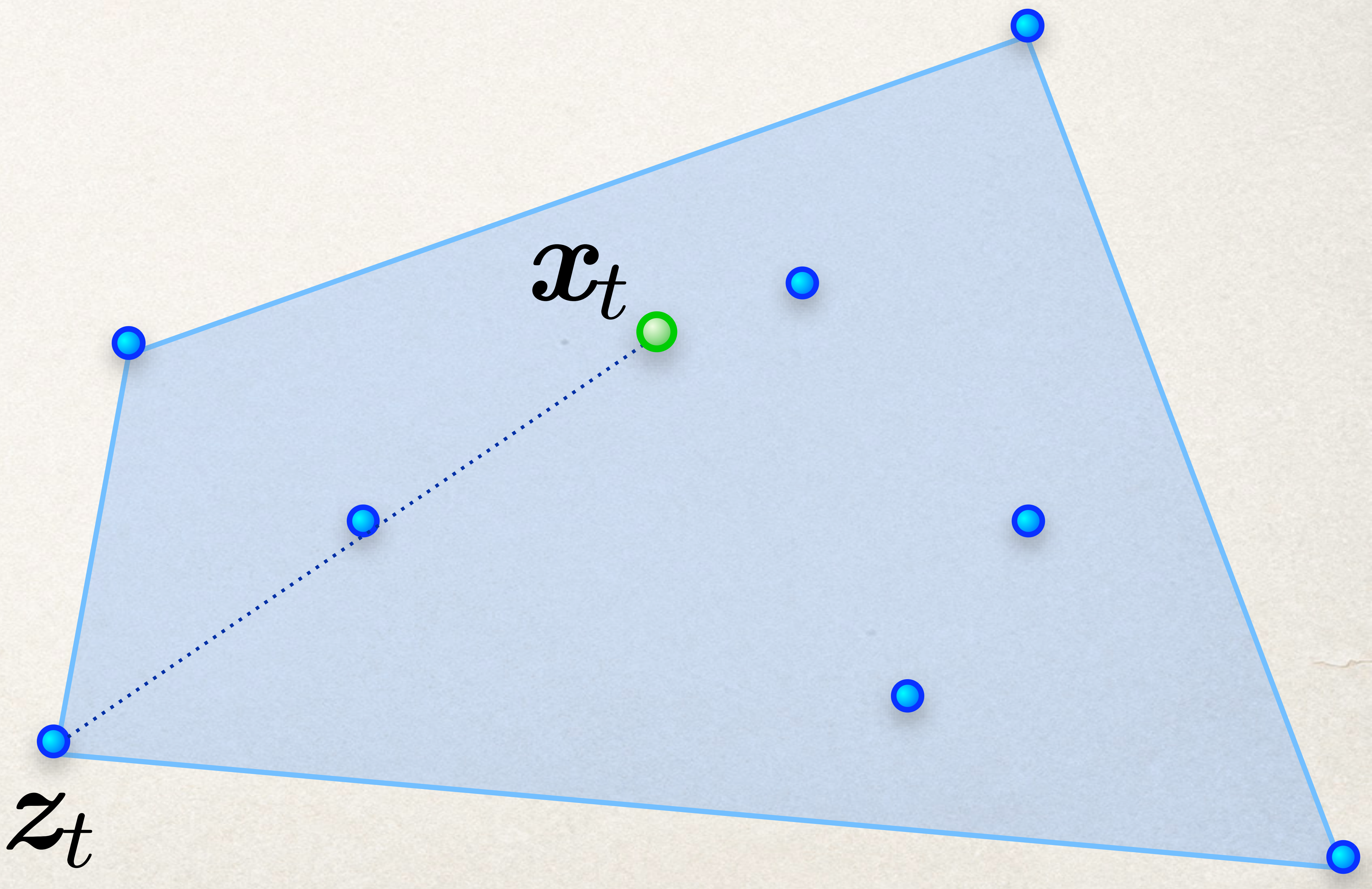


[Jaggi]

Frank-Wolfe [Frank&Wolfe 1956]

$$\min_{\boldsymbol{x} \in \operatorname{conv}(\mathcal{A})} \quad f(\boldsymbol{x})$$

init $\boldsymbol{x}_0 \in \operatorname{conv}(\mathcal{A})$
for $t = 0 \dots T$
    Find $\boldsymbol{z}_t :=$ (Approx-) $\mathrm{LMO}_{\mathcal{A}}(\nabla f(\boldsymbol{x}_t))$
    Variant 0: $\gamma = \frac{2}{t+2}$
    Update $\boldsymbol{x}_{t+1} := (1-\gamma)\boldsymbol{x}_t + \gamma \boldsymbol{z}_t$
end for



Frank-Wolfe [Frank&Wolfe 1956]

$$\min_{\boldsymbol{x} \in \operatorname{conv}(\mathcal{A})} \quad f(\boldsymbol{x})$$

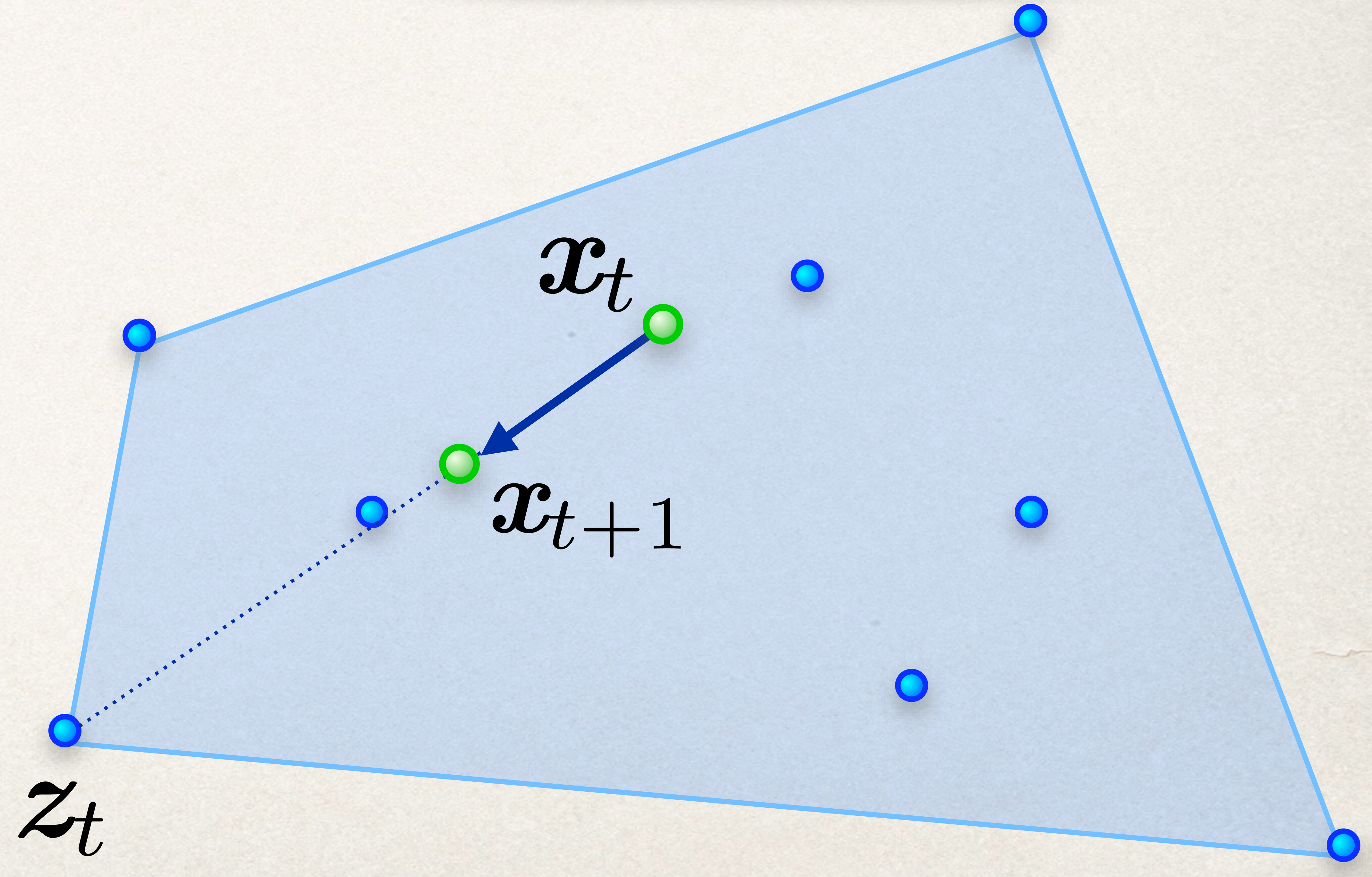
Sparse Iterate

$$\boldsymbol{x}_t = \sum_{i=0}^{t} \gamma_i \boldsymbol{z}_i$$

$$\sum_{i=0}^{t} \gamma_i = 1$$

---

init $\boldsymbol{x}_0 \in \operatorname{conv}(\mathcal{A})$
for $t = 0 \dots T$
  Find $\boldsymbol{z}_t :=$ (Approx-) $\mathrm{LMO}_{\mathcal{A}}(\nabla f(\boldsymbol{x}_t))$
  Variant 0: $\gamma = \frac{2}{t+2}$
  Update $\boldsymbol{x}_{t+1} := (1-\gamma)\boldsymbol{x}_t + \gamma \boldsymbol{z}_t$
end for



Outlook

- How to adapt the FW framework to boost variational inference
- Converge analysis in the Variational Inference case

Thank you

- Poster #31
- P.S. We are hiring PostDocs!