Randomized Projection Methods for Convex Feasibility Problems

Peter Richtárik

joint work with Ion Necoara and Andrei Patrascu

arXiv:1801.04873

ISMP 2018 – Bordeaux – July 5, 2018

Convex Feasibility

Convex Feasibility Problem

Nonempty closed convex set

$$\text{Find } x \in C \subseteq \mathbb{R}^n$$

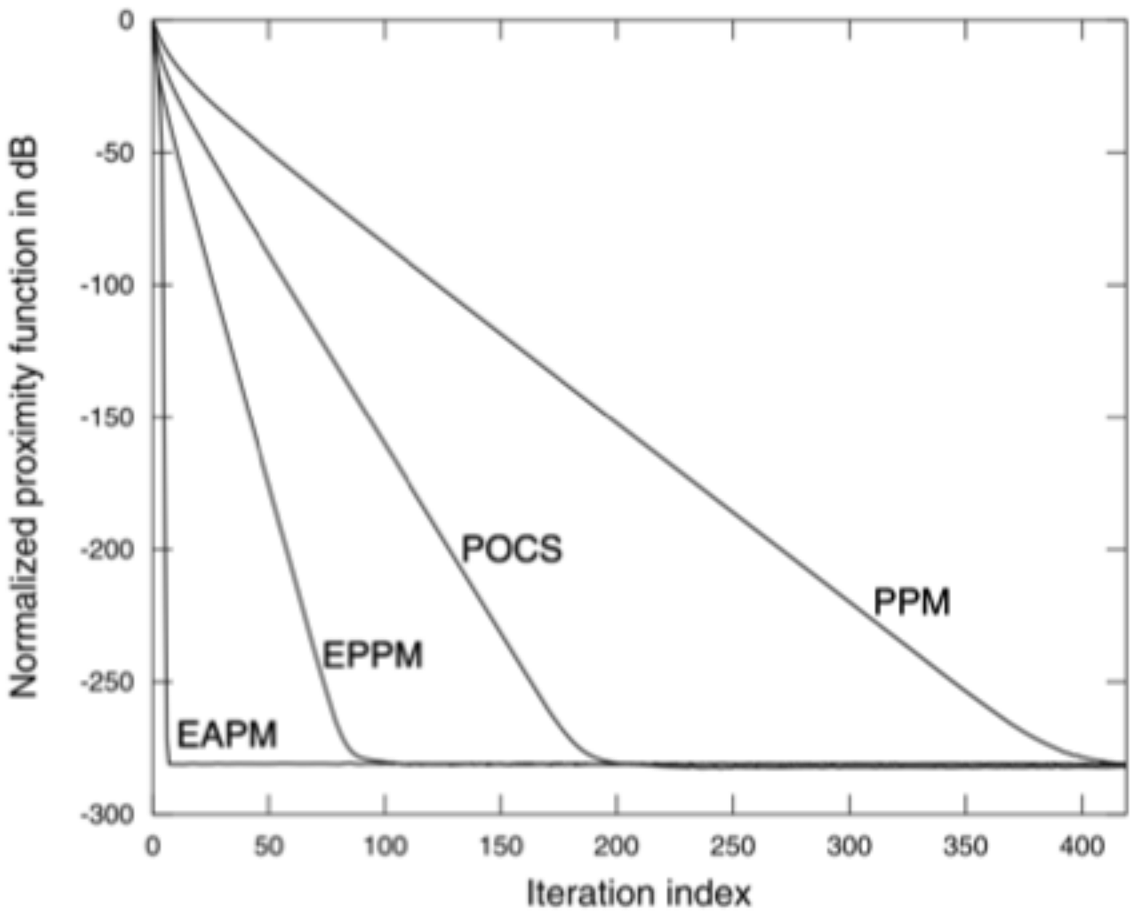
Applications

- communications, optics, neural networks, image processing (Stark-Yang '98)
- color imaging (Sharma '00)
- magnetic resonance imaging (Samsonov-Kholmovski-Parker-Johnson '04)
- wavelet-based denoising (Choi-Baraniuk '04)
- antenna design (Gu-Stark-Yang '04)
- data compression (Liew-Yan-Law '05)
- sensor networks problems (Blatt-Hero '06)
- intensity modulated radiation therapy (Herman-Chen '08)
- computerized tomography (Herman '09)
- demosaicking (Lu-Karzand-Vetterli '10)

Algorithms

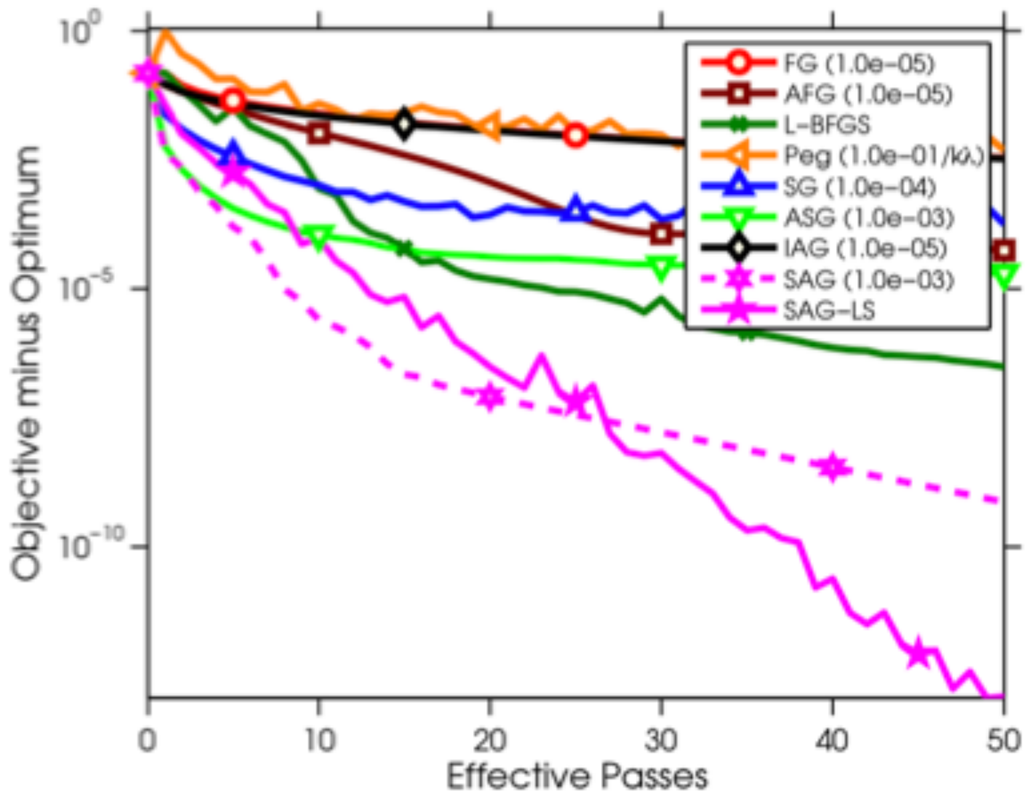
- linear equations (Kaczmarz ‘37)
- linear inequalities (Motzkin-Shoenberg ’54, Censor et al ‘11)
- convex feasibility (Polyak-Gubin-Raik ‘67), (Bauschke-Borwein ‘96), (Combettes ‘96)
- random methods (Nedic ’10, ‘11)
- monotone operators viewpoint (Bauschke-Combettes ‘11)
- conic feasibility (Henrion ‘11)
- review book (Escalante-Raydan ‘11)

Motivation 1



Extrapolated (alternating/parallel) projection methods are much better in practice than non-extrapolated variants, but **there is no theory that supports this empirical observation.** (Censor-Chen-Combettes-Davidi-Herman '11)

Motivation 2



IAG -> SAG by (LeRoux-Schmidt-Bach '12)

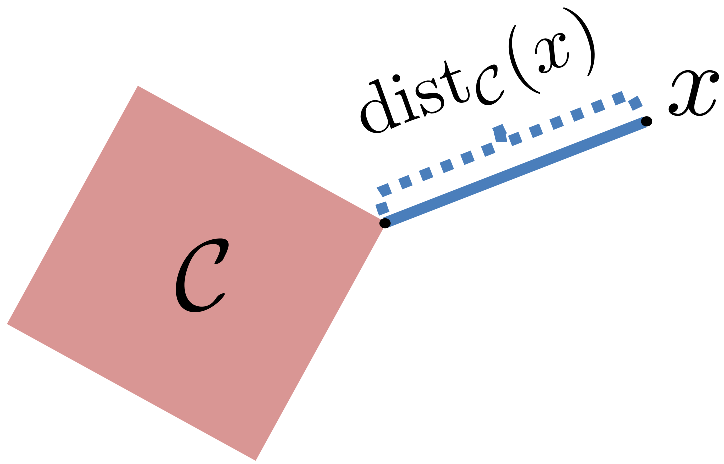
Randomized variants of deterministic methods are easier to analyze, and often lead to better complexity and practical behavior. **Develop and analyze randomized parallel projection methods.**

Plan

- We design and analyze randomized versions of projection methods
- Our rates explain why extrapolation helps
- Existing extrapolation rules are interpretable by our theory as online numerical approximations of a certain long-stepsize rule
- Our approach will involve new ideas, such as:
  - Stochastic approximation of convex sets
  - Stochastic reformulations of convex feasibility
  - Our algorithm: SGD / stochastic fixed point method / stochastic projection method (+ minibatching)
  - Sublinear (always) and linear rates (sometimes)

Our Goal

$$\text{Find } x \in \mathcal{C} \subseteq \mathbb{R}^n$$



Deterministic Algorithm

$$\mathrm{dist}^2_{\mathcal{C}}(x) \leq \epsilon$$

Deterministic vector output by the algorithm

Randomized Algorithm

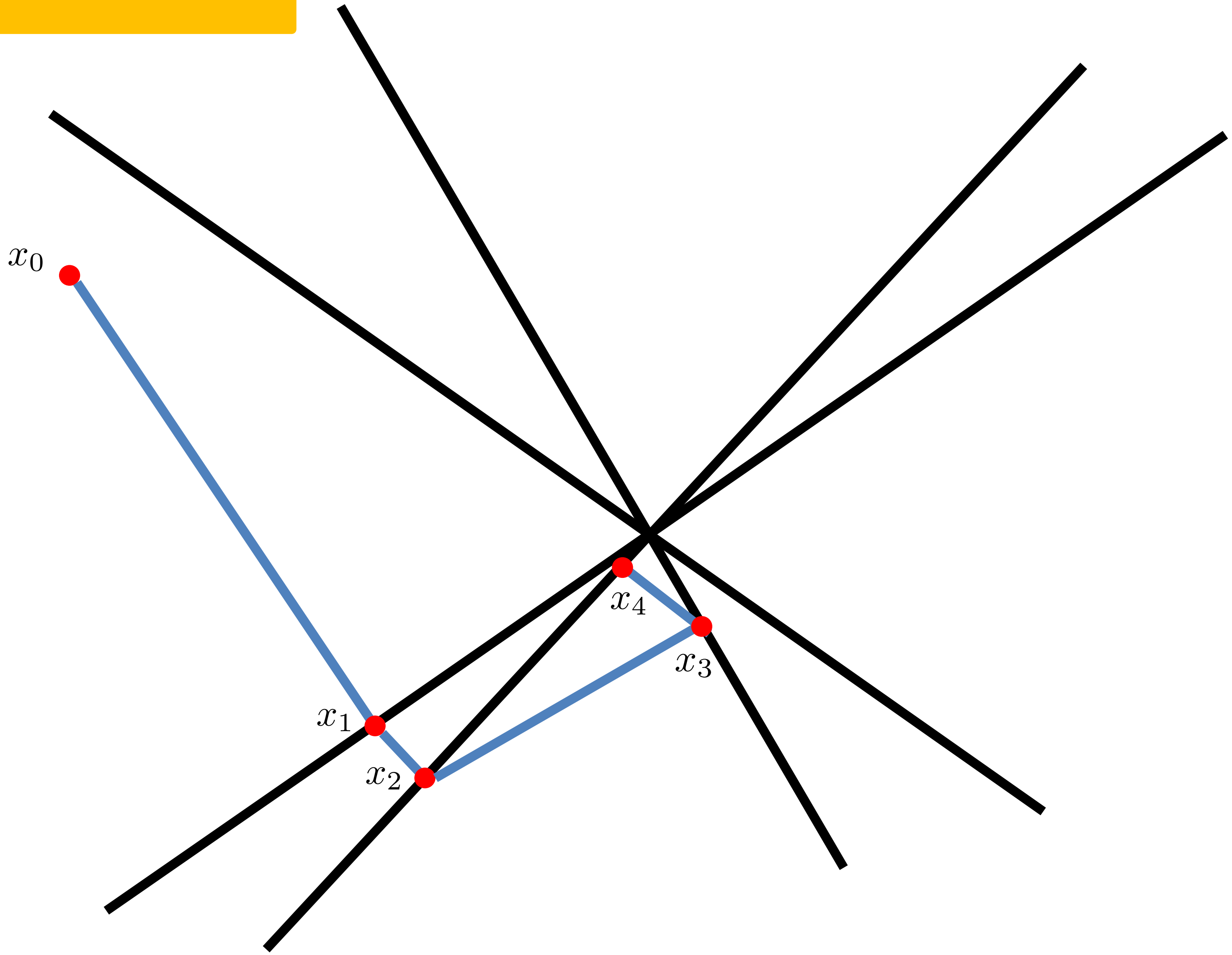
Expectation with respect to the randomness of the algorithm

$$\mathrm{E}\left[\mathrm{dist}^2_{\mathcal{C}}(x)\right] \leq \epsilon$$

Random vector output by the algorithm

Minibatch size: $\tau = 1$
Stepsize: $\alpha = 1$

$$x_{k+1} = (1-\alpha)x_k + \alpha\frac{1}{\tau}\sum_{i=1}^{\tau}\Pi_{\mathcal{S}_{ki}}(x_k)$$



Stochastic Approximation of Sets

Stochastic Approximation of Sets

Definition

Let $\mathcal{C} \subseteq \mathbb{R}^n$ be a closed and convex set. Let (Ω, Σ, P) be a probability space and consider a mapping $\mathcal{S} : \Omega \to 2^{\mathbb{R}^n}$.

If

1. $\mathcal{S}(\omega)$ is closed and convex for all $\omega \in \Omega$,
2. $\mathcal{C} \subseteq \mathcal{S}(\omega)$ for all $\omega \in \Omega$,
3. $\omega \mapsto \mathrm{dist}^2_{\mathcal{S}(\omega)}(x)$ is measurable for all $x \in \mathbb{R}^n$,

then we say that $(\Omega, \Sigma, P, \mathcal{S})$ is a stochastic approximation of $\mathcal{C}$.

$$x \in \mathcal{C} \quad \Rightarrow \quad x \in \mathcal{S} \quad \Rightarrow \quad \mathrm{E}_{\mathcal{S}\sim P}[\mathrm{dist}^2_{\mathcal{S}}(x)] = 0$$

If, moreover,

$$\mathrm{E}_{\mathcal{S}\sim P}[\mathrm{dist}^2_{\mathcal{S}}(x)] = 0 \quad \Leftrightarrow \quad x \in \mathcal{C},$$

then we say that the approximation is exact.

Stochastic Approximation of Sets: Intersection

$$\mathcal{C} = \bigcap_{i=1}^{m} \mathcal{C}_i$$

Exact

1. Trivial — $\mathcal{S} = \mathcal{C}$ with probability 1
2. Natural — $\mathcal{S} = \mathcal{C}_i$ with probability $p_i \geq 0$
3. Composite — $\mathcal{S} = \bigcap_{i \in S} \mathcal{C}_i$ with probability $p_S \geq 0$

Stochastic Approximation of Sets: Linear Systems

$\mathcal{C}_i = \{x \in \mathbb{R}^n \;:\; \mathbf{A}_{i:}x = b_i\}$

$$\mathcal{C} = \{x \in \mathbb{R}^n \;:\; \mathbf{A}x = b\} = \bigcap_{i=1}^{m} \mathcal{C}_i$$

1. Trivial — $\mathcal{S} = \mathcal{C}$ with probability 1
2. Natural — $\mathcal{S} = \mathcal{C}_i$ with probability $p_i \geq 0$
3. Composite — $\mathcal{S} = \bigcap_{i \in S} \mathcal{C}_i$ with probability $p_S \geq 0$
4. Sketch — $\mathcal{S} = \{x \in \mathbb{R}^n \;:\; \mathbf{S}^\top \mathbf{A}x = \mathbf{S}^\top b\}$

Random matrix

Stochastic Approximation of Sets: System of Linear Inequalities

$\mathcal{C}_i = \{x \in \mathbb{R}^n \ : \ \mathbf{A}_{i:} x \leq b_i\}$

$$\mathcal{C} = \{x \in \mathbb{R}^n \ : \ \mathbf{A}x \leq b\} = \bigcap_{i=1}^{m} \mathcal{C}_i$$

1. Trivial — $\mathcal{S} = \mathcal{C}$ with probability 1
2. Natural — $\mathcal{S} = \mathcal{C}_i$ with probability $p_i \geq 0$
3. Composite — $\mathcal{S} = \bigcap_{i \in S} \mathcal{C}_i$ with probability $p_S \geq 0$
4. Sketch — $\mathcal{S} = \{x \in \mathbb{R}^n \ : \ \mathbf{S}^\top \mathbf{A}x \leq \mathbf{S}^\top b\}$

Random vector with nonnegative entries

Stochastic Reformulations of Convex Feasibility

Stochastic Reformulations of Convex Feasibility

$$f_{\mathcal{S}}(x) \overset{\text{def}}{=} \tfrac{1}{2}\mathrm{dist}^2_{\mathcal{S}}(x) = \tfrac{1}{2}\|x - \Pi_{\mathcal{S}}(x)\|^2$$

Stochastic Optimization Problem (SOP)

$$\text{Minimize } f(x) \overset{\text{def}}{=} \mathrm{E}_{\mathcal{S}\sim P}[f_{\mathcal{S}}(x)]$$

Stochastic Fixed Point Problem (SFPP)

$$\text{Solve } x = \mathrm{E}_{\mathcal{S}\sim P}\left[\Pi_{\mathcal{S}}(x)\right]$$

Stochastic Feasibility Problem (SFP)

$$\text{Find } x \in \mathbb{R}^n \text{ such that } \mathrm{P}(x \in \mathcal{S}) = 1$$

In the case of linear feasibility, these reformulations were studied in (R-Takáč '17)

Equivalence & Exactness

Theorem (Equivalence)

The three stochastic reformulations of the convex feasibility problem have the same solution sets:

$\mathcal{C} \subseteq \mathcal{C}' \stackrel{\text{def}}{=}$ **minimizers of SOP = fixed points of SFPP = solutions of SFP**

Theorem (Exactness)

$\exists \mu > 0$ such that for all $x \in \mathbb{R}^n$:

$$\mu \cdot \|x - \Pi_{\mathcal{C}}(x)\|^2 \leq \mathrm{E}_{\mathcal{S} \sim P}\left[\|x - \Pi_{\mathcal{S}}(x)\|^2\right]$$

$$\mathcal{C} = \mathcal{C}'$$

Assumption

Assumption (stochastic linear regularity)

$\exists \mu > 0$ such that for all $x \in \mathbb{R}^n$:

$$\mu \cdot \|x - \Pi_{\mathcal{C}}(x)\|^2 \leq \mathrm{E}_{\mathcal{S}\sim P}\left[\|x - \Pi_{\mathcal{S}}(x)\|^2\right]$$

Jensen inequality: $L \leq 1$

Also define:

$$\|\mathrm{E}_{\mathcal{S}\sim P}\left[x - \Pi_{\mathcal{S}}(x)\right]\|^2 \leq L \cdot \mathrm{E}_{\mathcal{S}\sim P}\left[\|x - \Pi_{\mathcal{S}}(x)\|^2\right]$$

$$\mu \leq L \leq 1$$

Stochastic Algorithms

"Basic" Method

Minimize $f(x) \stackrel{\text{def}}{=} \mathrm{E}_{\mathcal{S}\sim P}[f_{\mathcal{S}}(x)]$

Stochastic Gradient Descent

$$x_{k+1} = x_k - \alpha \nabla f_{\mathcal{S}_k}(x_k)$$

Solve $x = \mathrm{E}_{\mathcal{S}\sim P}\left[\Pi_{\mathcal{S}}(x)\right]$

Stochastic Fixed Point Method

$$x_{k+1} = (1-\alpha)x_k + \alpha \Pi_{\mathcal{S}_k}(x_k)$$

Find $x \in \mathbb{R}^n$ such that $\mathrm{P}(x \in \mathcal{S}) = 1$

Stochastic Projection Method

$$x_{k+1} = (1-\alpha)x_k + \alpha \Pi_{\mathcal{S}_k}(x_k)$$

"Parallel" Method

Minibatch size

Stochastic Gradient Descent

$$x_{k+1} = x_k - \alpha \frac{1}{\tau} \sum_{i=1}^{\tau} \nabla f_{S_{ki}}(x_k)$$

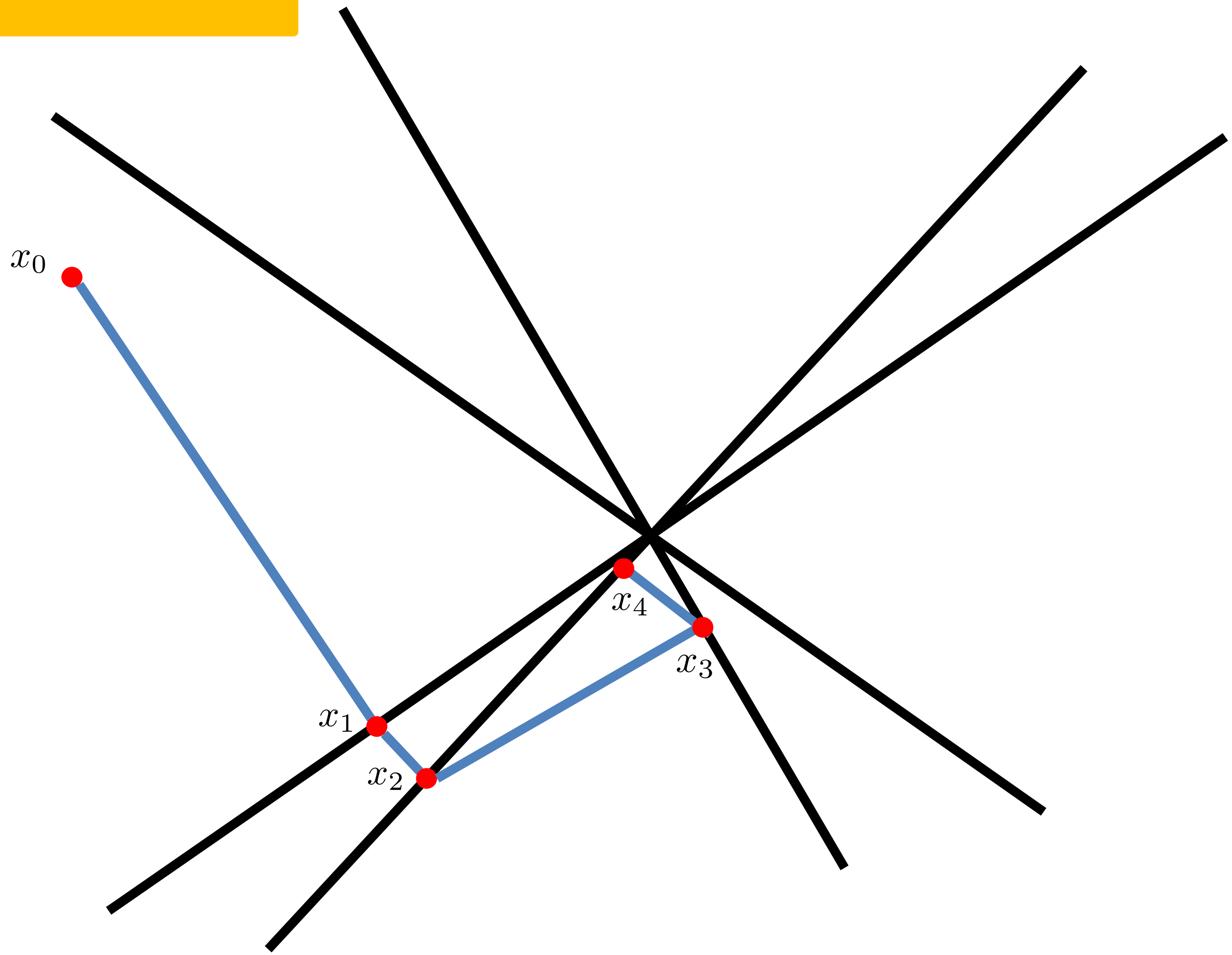
Minibatch size

Stochastic Fixed Point Method
Stochastic Projection Method

$$x_{k+1} = (1-\alpha)x_k + \alpha \frac{1}{\tau} \sum_{i=1}^{\tau} \Pi_{S_{ki}}(x_k)$$

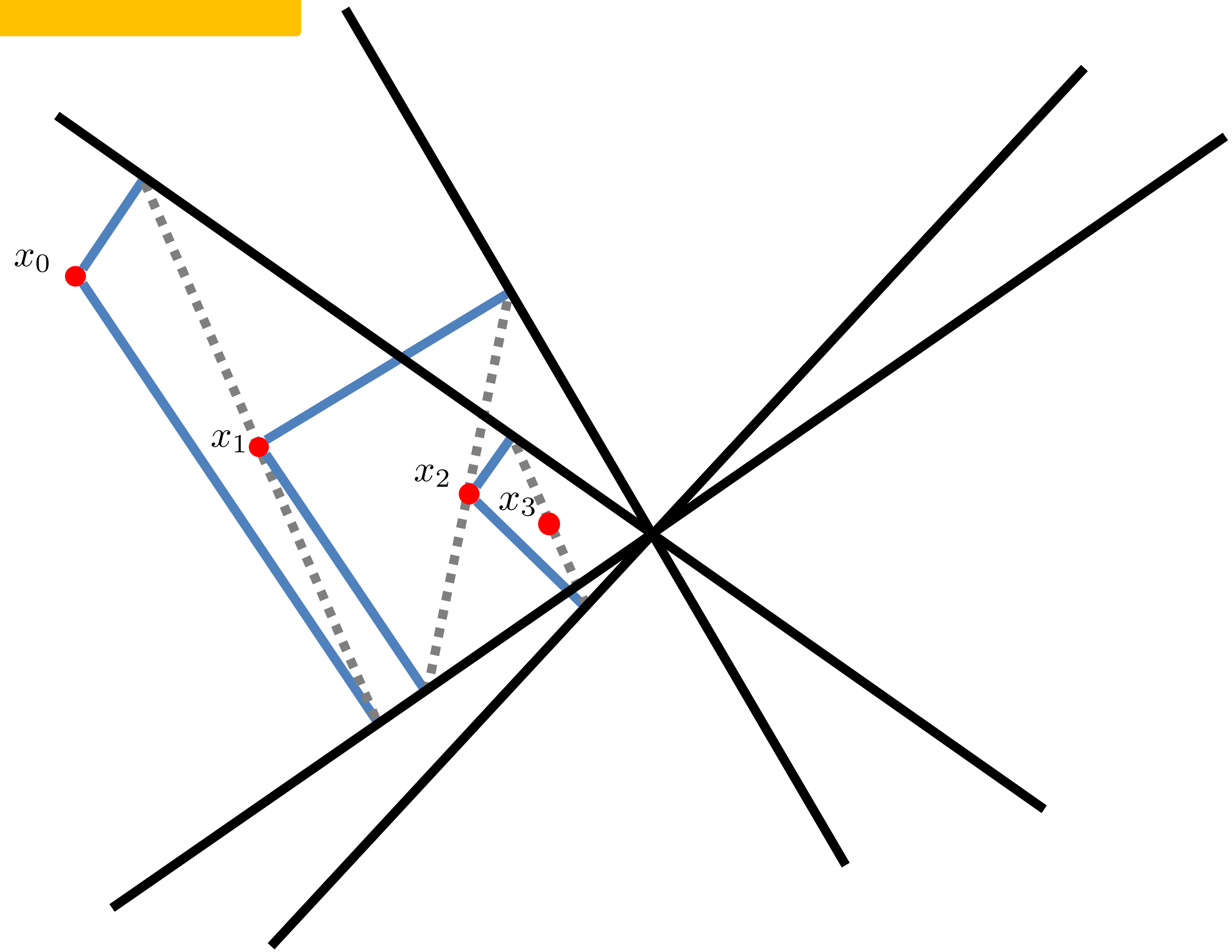
Minibatch size: $\tau = 1$
Stepsize: $\alpha = 1$

$$x_{k+1} = (1-\alpha)x_k + \alpha \frac{1}{\tau}\sum_{i=1}^{\tau} \Pi_{\mathcal{S}_{ki}}(x_k)$$



Minibatch size: $\tau = 2$
Stepsize: $\alpha = 1$

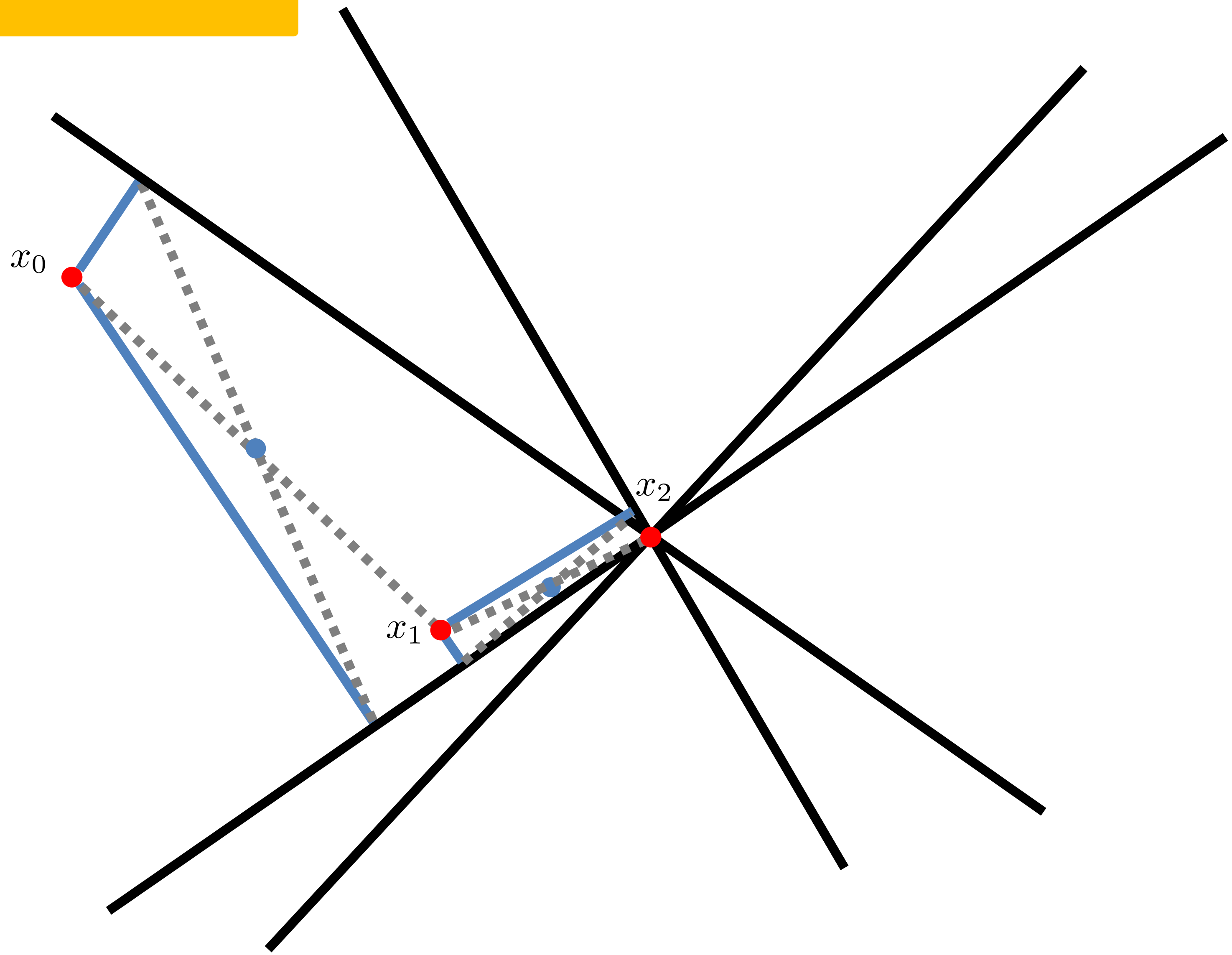
$$x_{k+1} = (1-\alpha)x_k + \alpha \frac{1}{\tau}\sum_{i=1}^{\tau} \Pi_{\mathcal{S}_{ki}}(x_k)$$



Minibatch size: $\tau = 2$

Stepsize: $\alpha = 2$

$$x_{k+1} = (1-\alpha)x_k + \alpha \frac{1}{\tau}\sum_{i=1}^{\tau} \Pi_{\mathcal{S}_{ki}}(x_k)$$



Convergence Results

Sublinear Convergence

(no need to assume linear regularity)

Theorem With stepsize $\alpha = 1/L_\tau$, we get:

$\hat{x}_k = \frac{1}{k}\sum_{t=0}^{k-1} x_t$

$L_\tau = \frac{1}{\tau} + (1 - \frac{1}{\tau})L$

$$\mathrm{E}\left[f(\hat{x}_k)\right] = \frac{1}{2}\mathrm{E}\left[\mathrm{dist}^2_{\mathcal{S}}(\hat{x}_k)\right] \leq \frac{L_\tau \mathrm{dist}^2_{\mathcal{C}}(x_0)}{2k}$$

Let $(\Omega, \Sigma, P, \mathcal{S})$ is a stochastic approximation of $\mathcal{C}$

Linear Convergence

(assuming linear regularity)

Theorem With stepsize $\alpha = 1/L_\tau$, we get:

Linear regularity parameter

$L_\tau = \frac{1}{\tau} + (1 - \frac{1}{\tau})L$

$$\mathrm{E}\left[\mathrm{dist}^2_{\mathcal{C}}(x_k)\right] \leq (1 - \mu/L_\tau)^k \, \mathrm{dist}^2_{\mathcal{C}}(x_0)$$

$$\mathrm{E}\left[f(x_k)\right] \leq \frac{L}{2}(1 - \mu/L_\tau)^k \, \mathrm{dist}^2_{\mathcal{C}}(x_0)$$

Best current rate of "parallel" projection method for convex feasibility

- Sketch approximations for linear systems (R.-Takáč '17) obtained as a special case
- Natural approximations for convex sets & $\tau = 1$ done in (Nedic '11)

Extrapolation Rules = Approximation of *L*

Optimal stepsize: $\alpha = 1/L_\tau$

$L_\tau = \frac{1}{\tau} + (1 - \frac{1}{\tau})L$

$$L = \sup_{x \notin \mathcal{C}} \frac{\left\| \mathrm{E}\left[x - \Pi_{\mathcal{S}}(x)\right]\right\|^2}{\mathrm{E}\left[\left\|x - \Pi_{\mathcal{S}}(x)\right\|^2\right]}$$

Online approximation of *L*:

$$L \approx \frac{\left\| \frac{1}{\tau}\sum_{i=1}^{\tau}\left[x_k - \Pi_{\mathcal{S}_{ki}}(x_k)\right]\right\|^2}{\frac{1}{\tau}\sum_{i=1}^{\tau}\left[\left\|x_k - \Pi_{\mathcal{S}_{ki}}(x_k)\right\|^2\right]}$$

Summary

- **New approach to convex feasibility** via:
  - Stochastic approximation of convex sets
  - Stochastic reformulations
    - Stochastic optimization
    - Stochastic fixed point
    - Stochastic feasibility
  - Natural algorithms for the stochastic reformulations
- First rate of a parallel projection method which is better than the rate of the non-parallel version
- Sheds light on the empirical success of extrapolated parallel projection methods (Censor-Chen-Combettes-Davidi-Herman '11)

The End