A Conditional-Gradient-Based Augmented Lagrangian Framework

Alp Yurtsever

alp.yurtsever@epfl.ch

joint work with

Olivier Fercoq & Volkan Cevher

Télécom ParisTech EPFL

ICML2019 - Long Beach

Télécom ParisTech
Ecole Polytechnique Fédérale de Lausanne (EPFL)

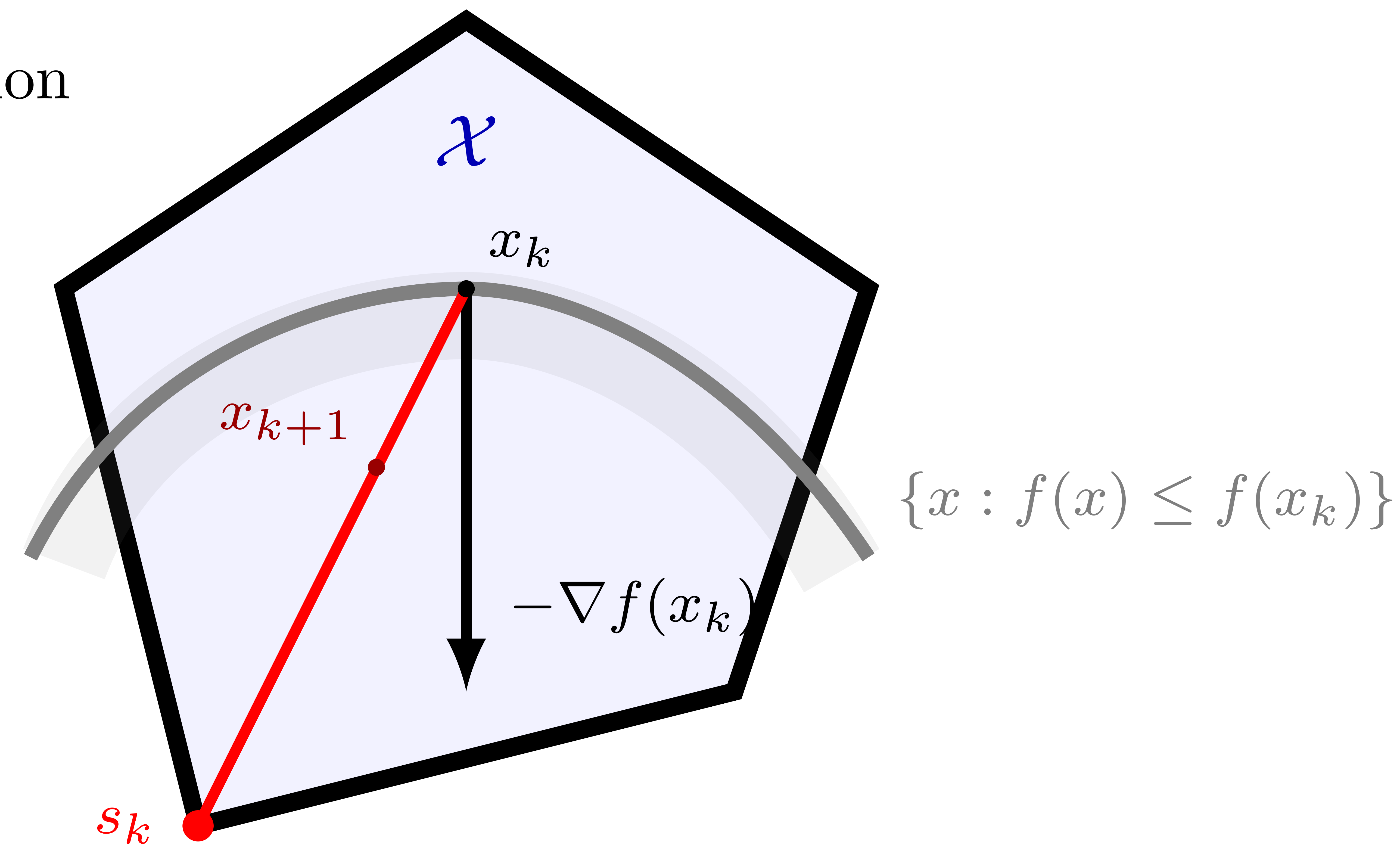
Conditional Gradient Method (CGM)

(Frank & Wolfe, 1956)
(Hazan, 2008)
(Jaggi, 2013)

$$\min_{x \in \mathcal{X}} f(x)$$

- ▷ $\mathcal{X} \subset \mathbb{R}^n$ is a convex compact set
- ▷ $f : \mathcal{X} \to \mathbb{R}$ is a smooth convex function

Input: $x_1 \in \mathcal{X}$
for $k = 1, 2, \ldots,$ **do**
  $\eta_k = 2/(k+1)$
  $s_k = \arg\min_{x \in \mathcal{X}} \langle \nabla f(x_k), x \rangle$
  $x_{k+1} = x_k + \eta_k(s_k - x_k)$
end for



Motivation: Solving Large-Scale SDP

(Hazan, 2008)
(Yurtsever et al., 2017)

$$\min_{x \in \mathcal{X}} f(x)$$

- $\triangleright$ $\mathcal{X} \subset \mathbb{R}^n$ is a convex compact set
- $\triangleright$ $f : \mathcal{X} \to \mathbb{R}$ is a smooth convex function

When $\mathcal{X}$ is PSD-cone with bounded trace

$\rightarrow$ *lmo* is cheap (Arithmetic Scalability)
$\rightarrow$ updates are rank-1 (Storage Scalability)

Motivation: Solving Large-Scale SDP

(Hazan, 2008)
(Yurtsever et al., 2017)

$$\min_{x \in \mathcal{X}} f(x)$$

- $\triangleright$ $\mathcal{X} \subset \mathbb{R}^n$ is a convex compact set
- $\triangleright$ $f : \mathcal{X} \to \mathbb{R}$ is a smooth convex function

When $\mathcal{X}$ is PSD-cone with bounded trace

$\rightarrow$ *lmo* is cheap (Arithmetic Scalability)
$\rightarrow$ updates are rank-1 (Storage Scalability)

$$\min_{x \in \mathcal{X}} f(x) \quad \text{s.t.} \quad Ax \in \mathcal{K}$$

- $\triangleright$ $A : \mathcal{X} \to \mathbb{R}^d$ is a given linear map
- $\triangleright$ $\mathcal{K}$ is a simple convex set

This paper:
A new **CGM-type** method
based on **augmented Lagrangian**

CGM via Quadratic Penalty: HCGM

(Yurtsever et al., 2018)

$$\min_{x \in \mathcal{X}} f(x) + \frac{\lambda}{2}\mathrm{dist}^2(Ax, \mathcal{K})$$ (QP formulation)

(QP formulation) $\xrightarrow{\lambda \to +\infty}$ (Original problem)

Start with some λ

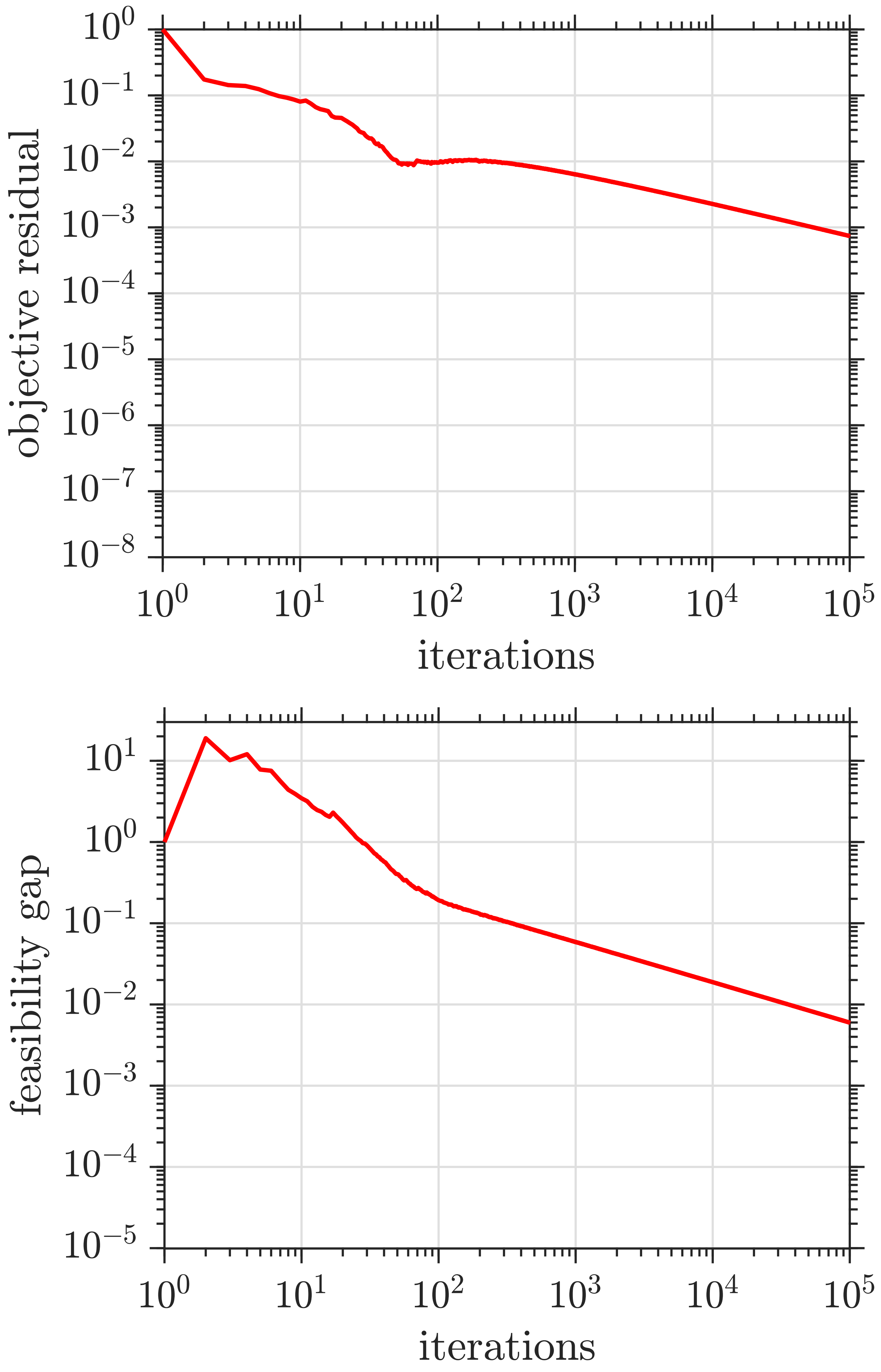
apply CGM step w.r.t. QP form.

increase λ at each iteration

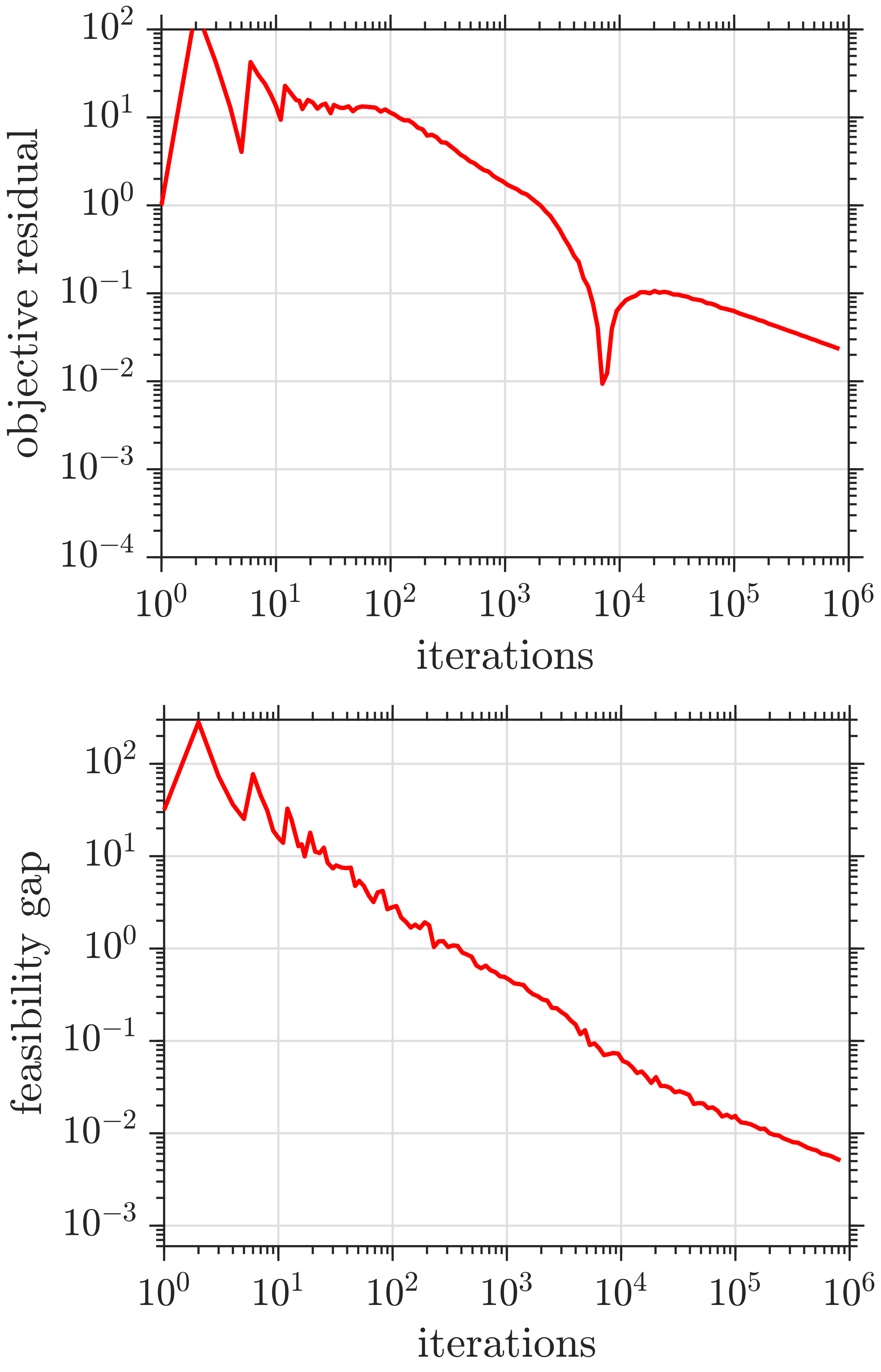
$$|f(x) - f^\star| = \mathcal{O}(1/\sqrt{k}) \quad \& \quad \mathrm{dist}(Ax, \mathcal{K}) = \mathcal{O}(1/\sqrt{k})$$

Performance of HCGM

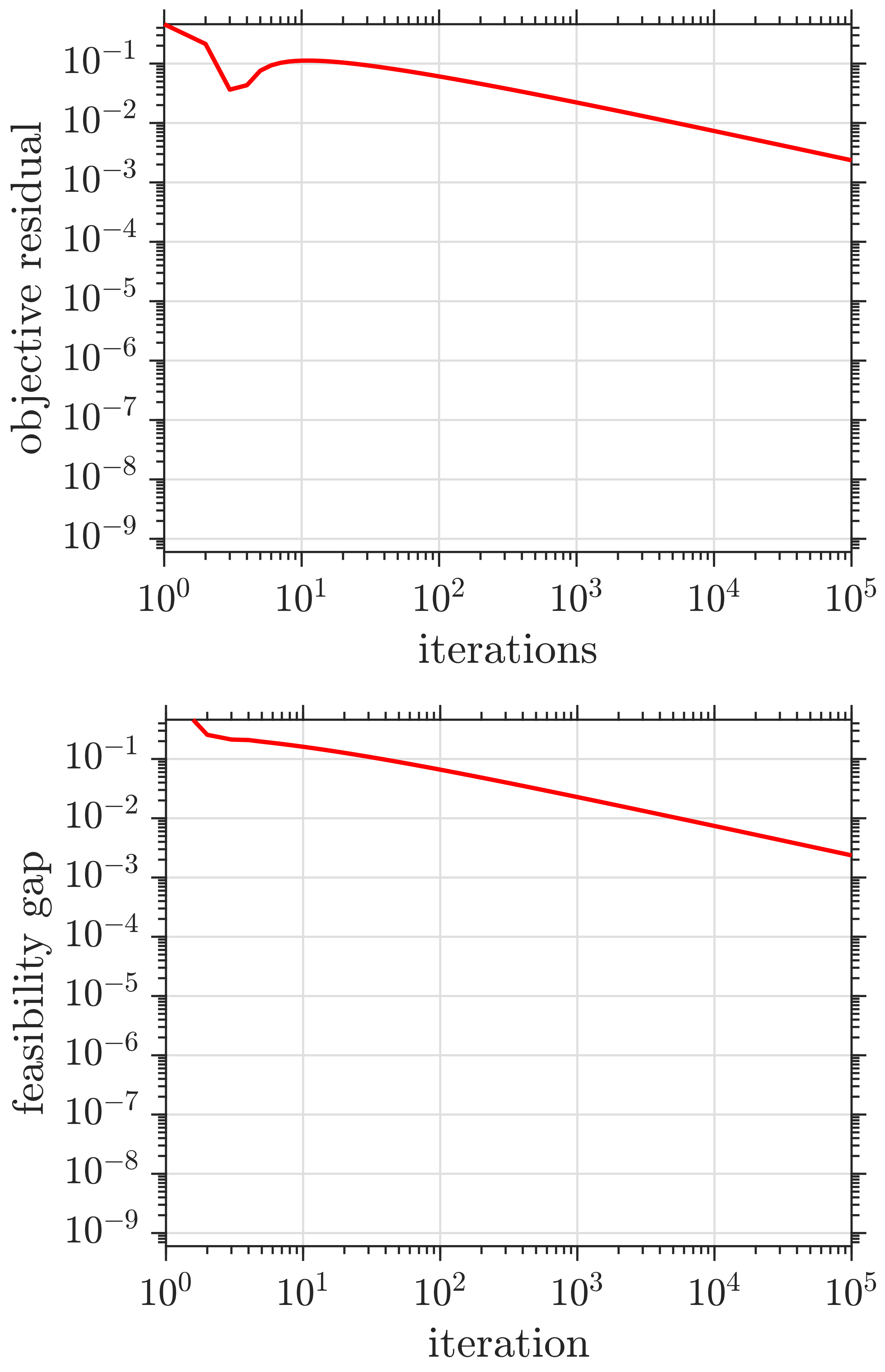
MaxCut



Clustering



Generalized Eig.Vec.



CGM via augmented Lagrangian: CGAL

$$\max_{y \in \mathbb{R}^d} \min_{x \in \mathcal{X}} \ \mathcal{L}_\lambda(x, y)$$ (AL formulation)

where $\mathcal{L}_\lambda(x, y) := f(x) + \min_{r \in \mathcal{K}} \left\{ \langle y, Ax - r \rangle + \frac{\lambda}{2} \|Ax - r\|^2 \right\}$

Start with some λ

apply CGM step on x

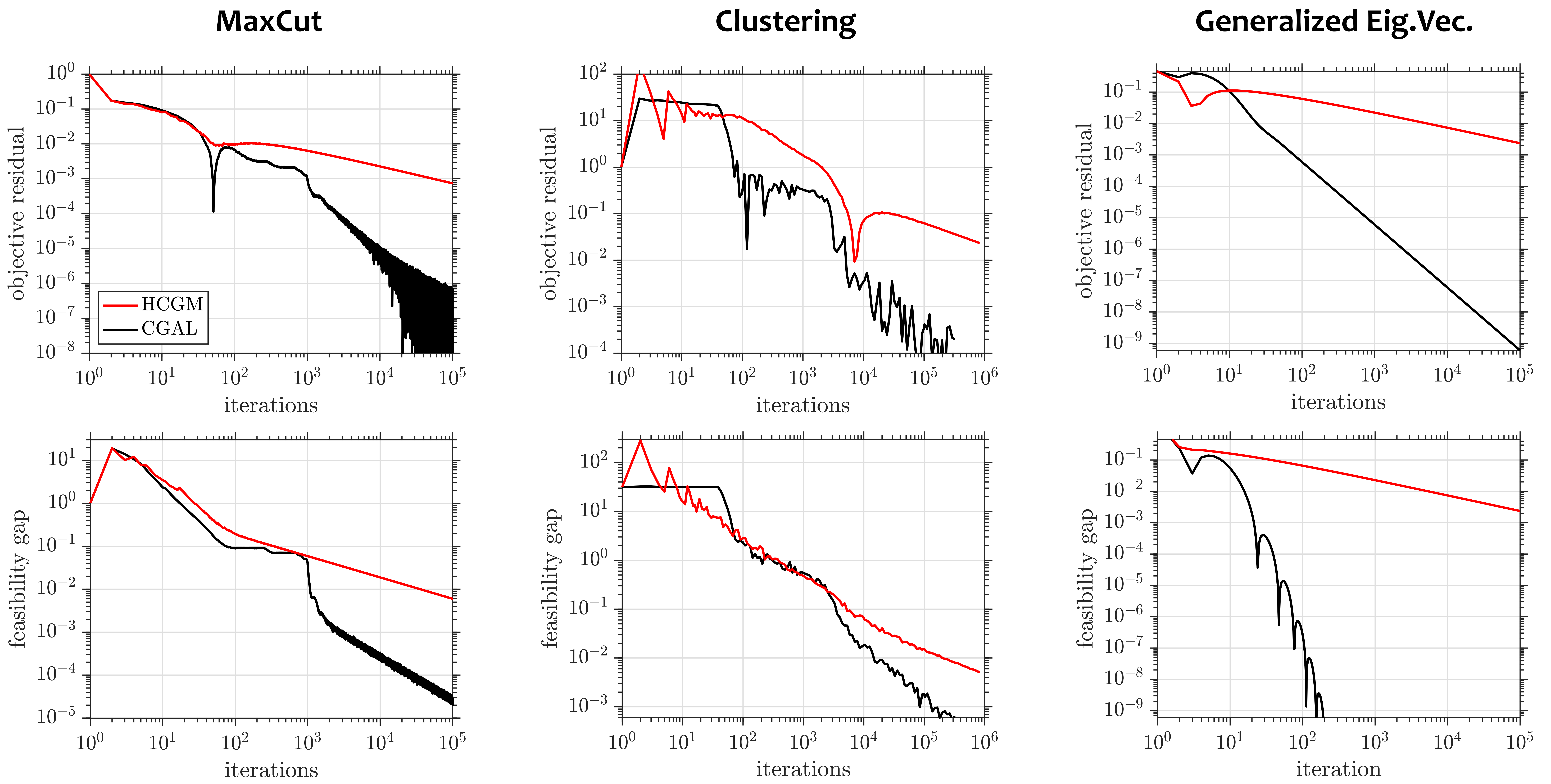
apply dual ascent step on y

increase λ at each iteration

$$|f(x) - f^\star| = \mathcal{O}(1/\sqrt{k}) \quad \& \quad \mathrm{dist}(Ax, \mathcal{K}) = \mathcal{O}(1/\sqrt{k})$$

CGAL vs HCGM

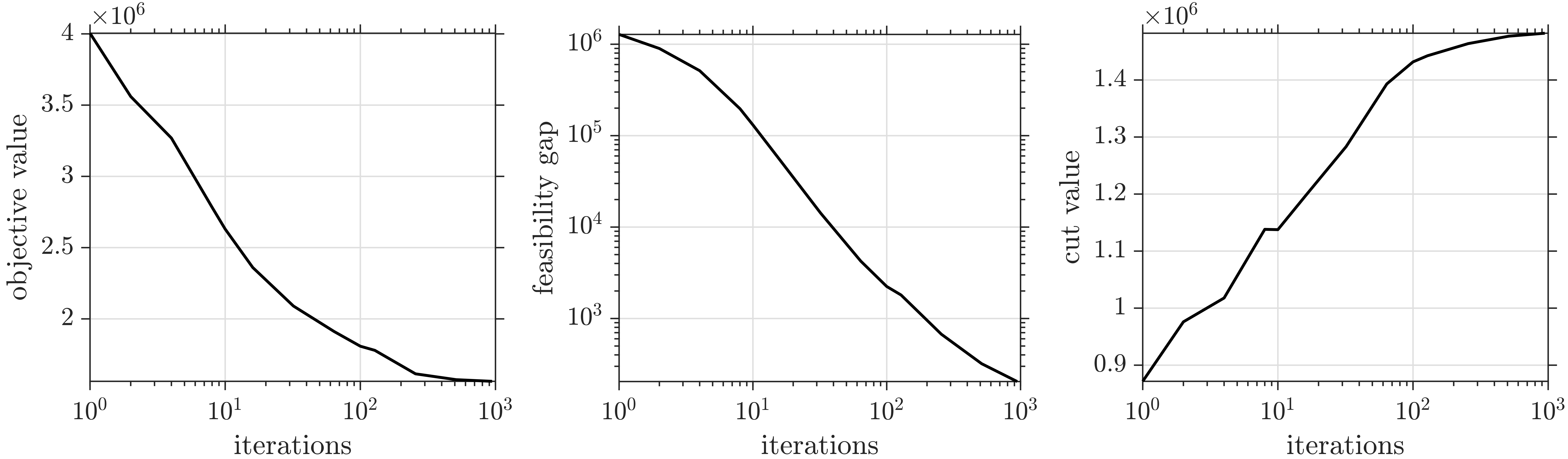
Poster today: **Pacific Ballroom #194**



Coming soon: Sketchy-CGAL

(Cevher-Tropp-Yurtsever, 2019)

$$\underset{x}{\text{minimize}} \quad -\frac{1}{4}\langle L, X\rangle \quad \text{subject to} \quad \text{diag}(X) = 1, \quad \text{trace}(X) = n, \quad X \text{ is PSD}$$



MaxCut with **(1`441`295 x 1`441`295)** dimensional **belgium-osm** Street Network from DIMACS10 Implementation Challenge Library