Frank-Wolfe Optimization Algorithms

A Brief Tutorial

Martin Jaggi
ETH Zurich

Optimization and Big Data 2015, May 7 2015, Edinburgh

Pointers

- ICML 2014 Tutorial Material

  google "Frank-Wolfe tutorial"

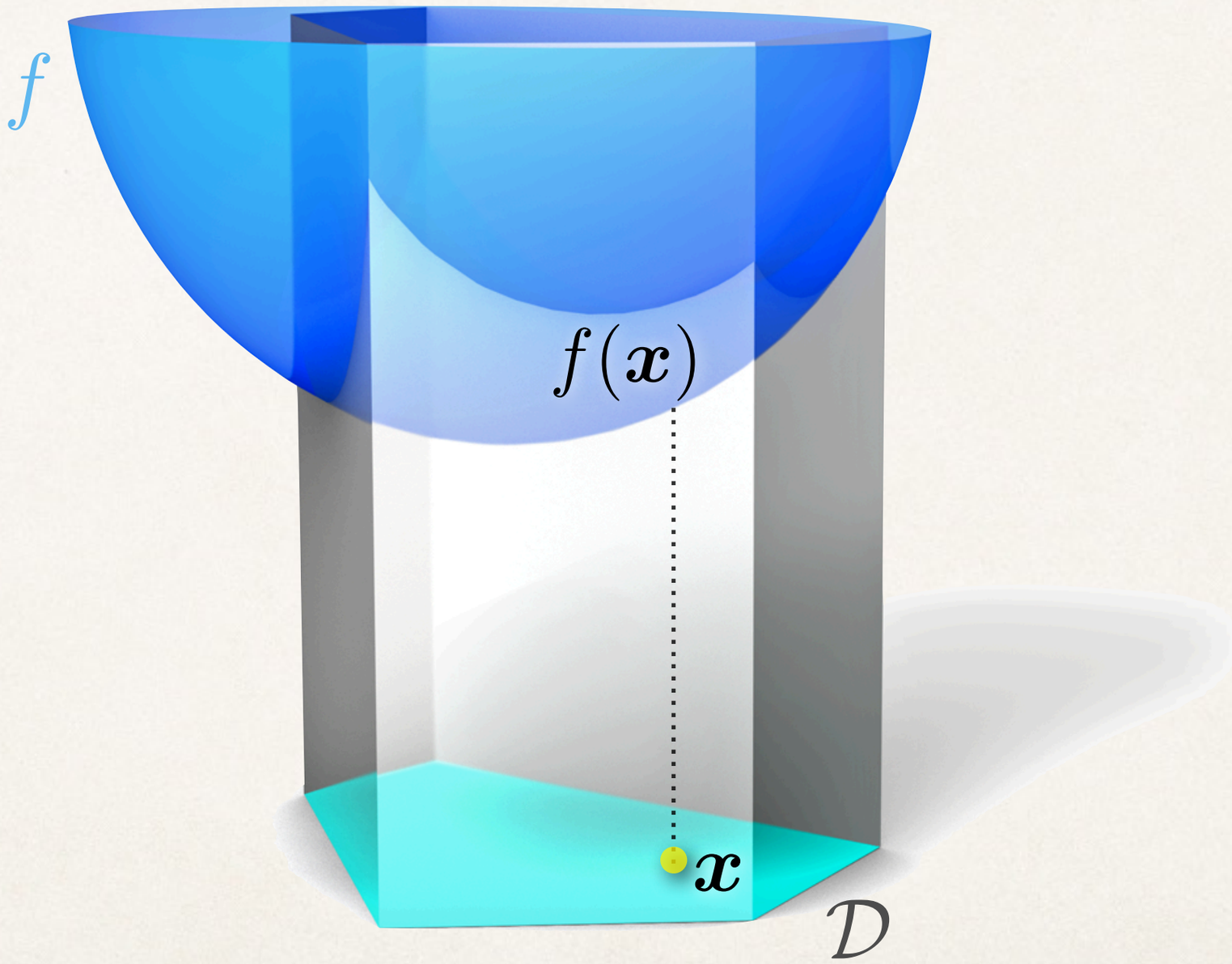
- distributed optimization talk:

  tomorrow 15:55 - Jakub Konecny

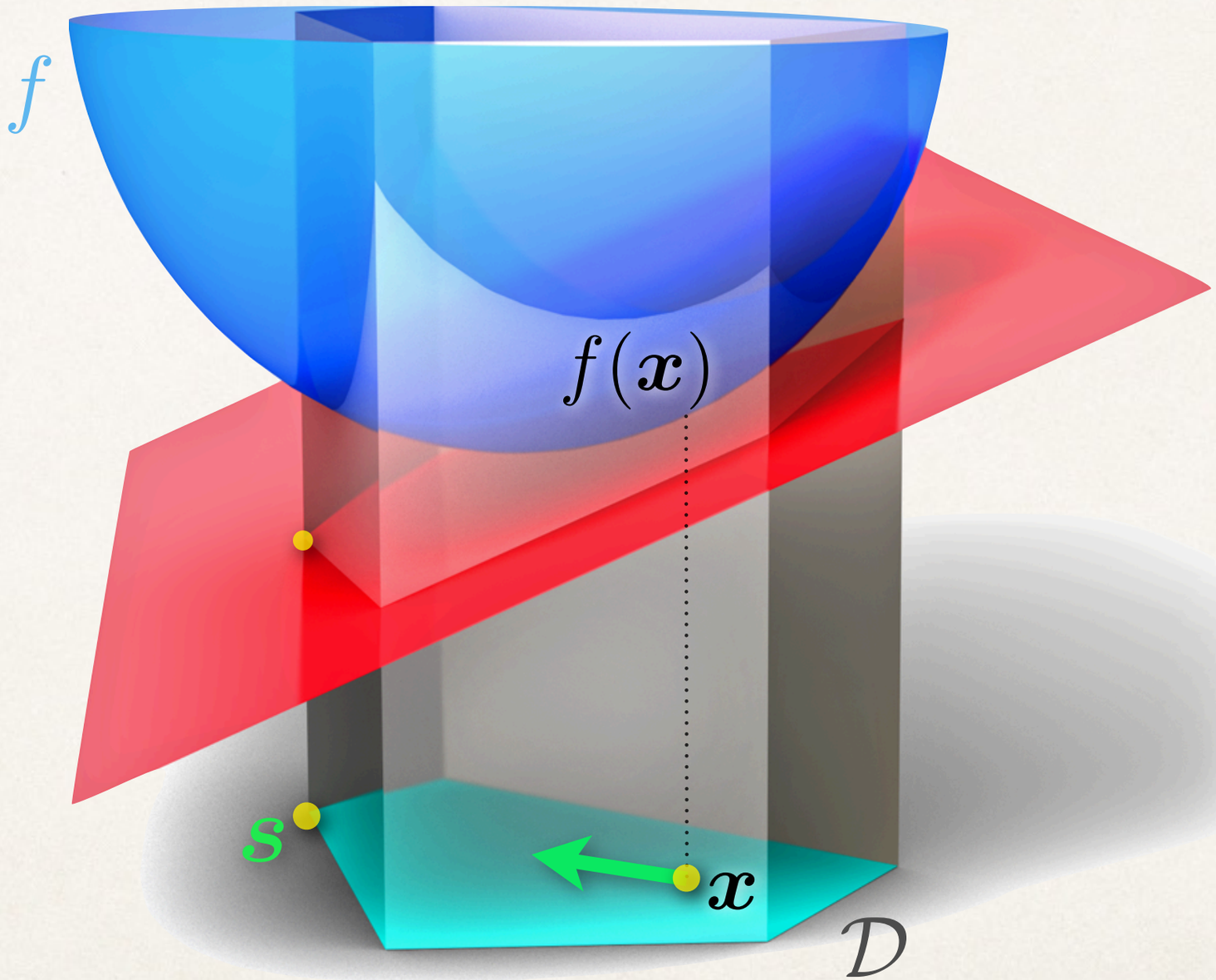
Outline

- Frank-Wolfe basics and history
- convergence analysis and geometry independence
- optimality & lower bounds
- applications for atomic domains
- faster rates under additional assumptions

Frank-Wolfe Algorithm



Frank-Wolfe Algorithm

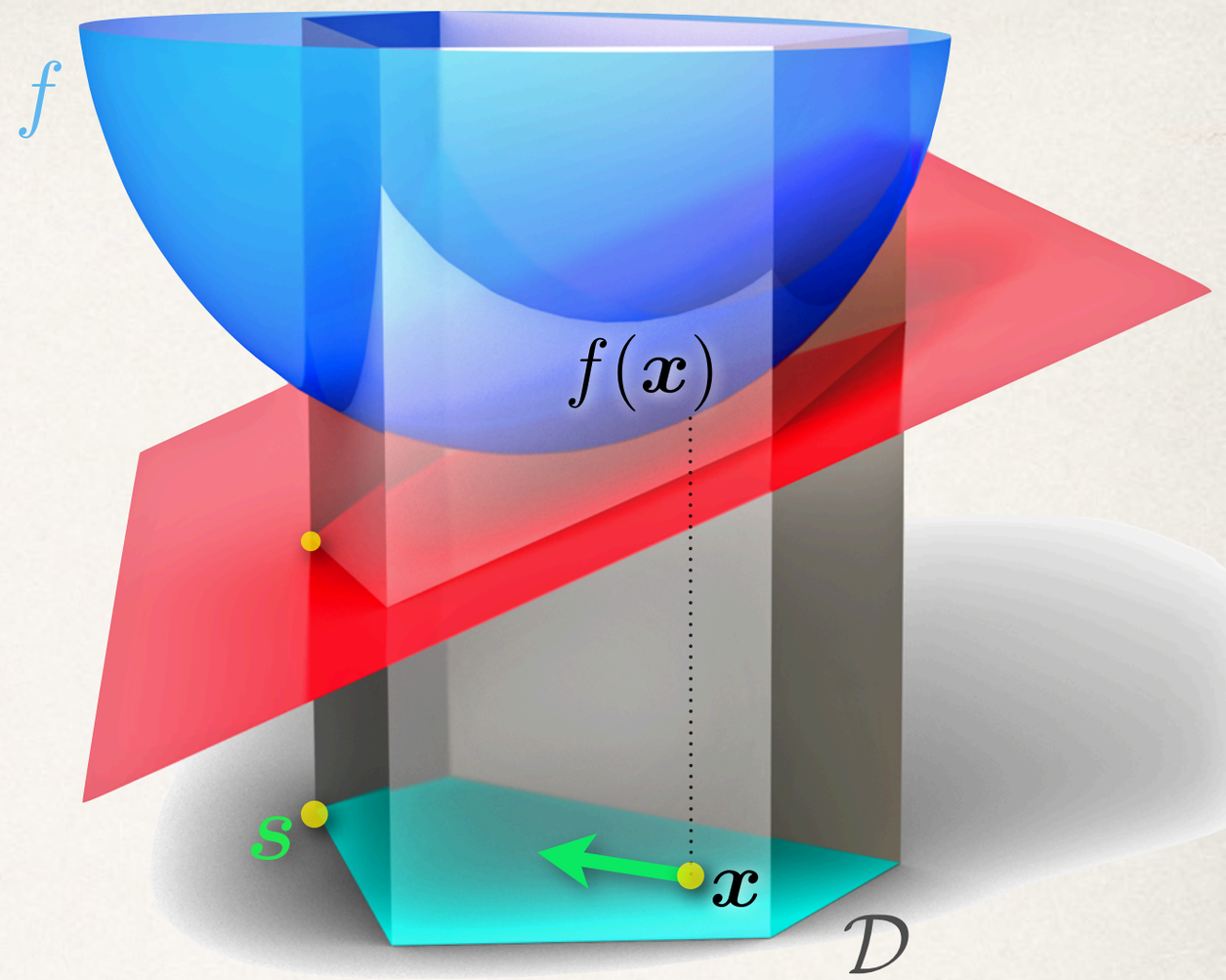


[ICML 2014 Tutorial]

$$\min_{\boldsymbol{x}\in\mathcal{D}} f(\boldsymbol{x})$$

The Linear Minimization Oracle

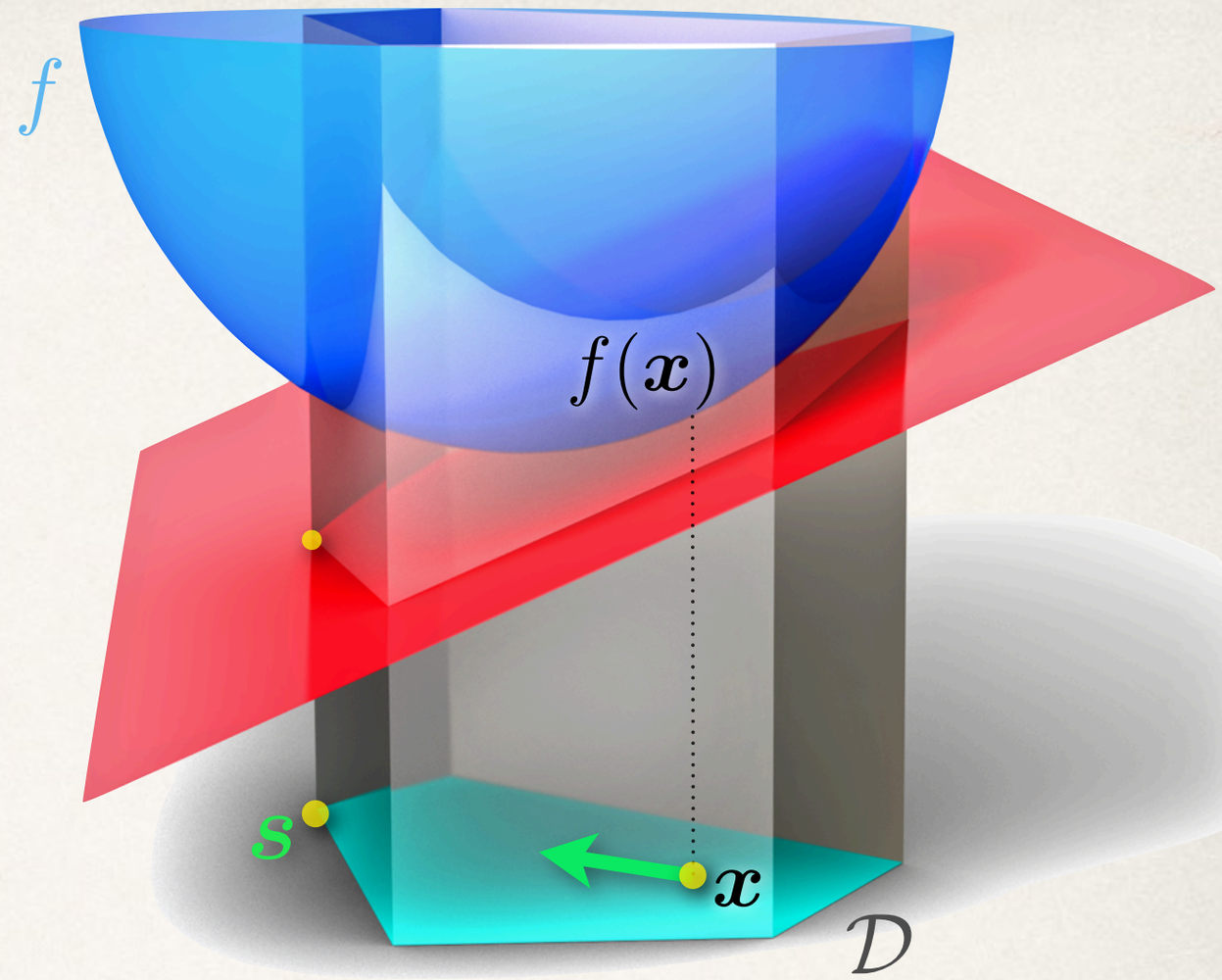
$$\mathrm{LMO}_{\mathcal{D}}(\boldsymbol{d}) := \arg\min_{\boldsymbol{s}\in\mathcal{D}} \langle \boldsymbol{d}, \boldsymbol{s} \rangle$$



Linearization

$$\min_{\boldsymbol{s}\in\mathcal{D}} \; f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}), \boldsymbol{s} - \boldsymbol{x} \rangle$$

$$\min_{\boldsymbol{x} \in \mathcal{D}} f(\boldsymbol{x})$$



The Linear Minimization Oracle

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

Algorithm 1: Frank-Wolfe

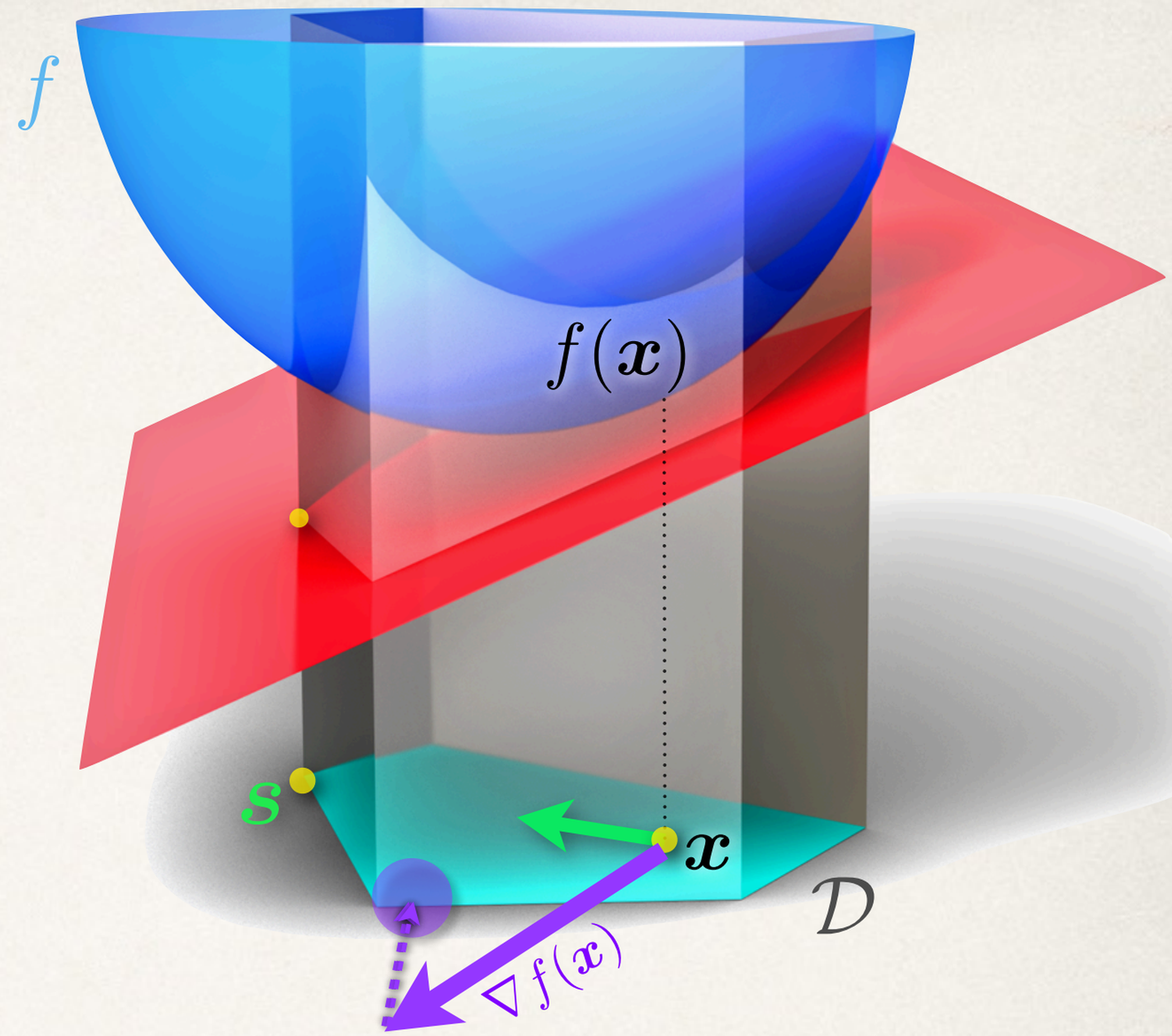
Let $\boldsymbol{x}^{(0)} \in \mathcal{D}$
for $t = 0 \dots T$ **do**
  Compute $\boldsymbol{s} := \mathrm{LMO}_{\mathcal{D}}\big(\nabla f(\boldsymbol{x}^{(t)})\big)$
  Let $\gamma := \frac{2}{t+2}$
  Update $\boldsymbol{x}^{(t+1)} := (1-\gamma)\boldsymbol{x}^{(t)} + \gamma \boldsymbol{s}$
end

Convergence

$$f(\boldsymbol{x}^{(t)}) - f^* \leq O(1/t)$$

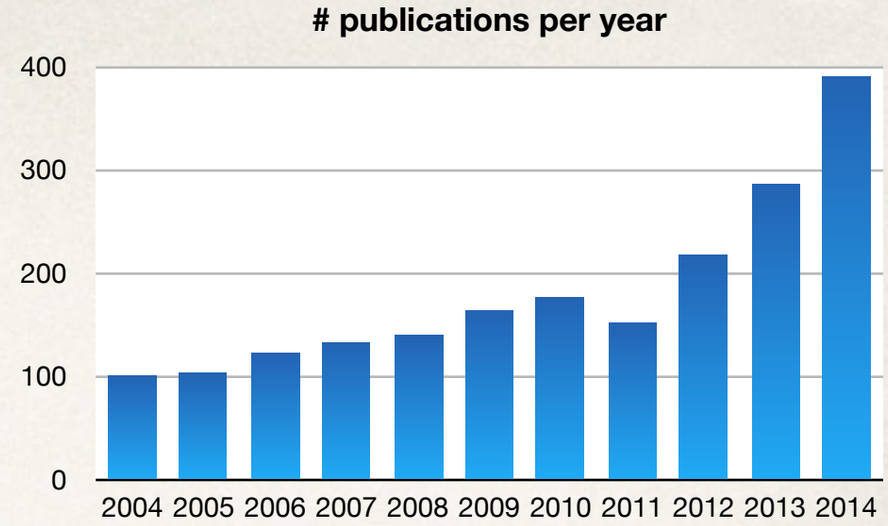
convexity, smoothness of f, boundedness of $\mathcal{D}$

Two kinds of first-order methods



	Frank-Wolfe	Gradient Descent and Proximal Methods
Iteration cost	*(approx.)* solve ***linear problem*** on D	***projection*** back to D (or prox operator)
Iterates	***sparse*** ✓ (in terms of used vertices)	***dense*** ✗
∞-dim domain	***yes*** ✓ (if oracle available)	?

Origins of Frank-Wolfe / Conditional Gradient

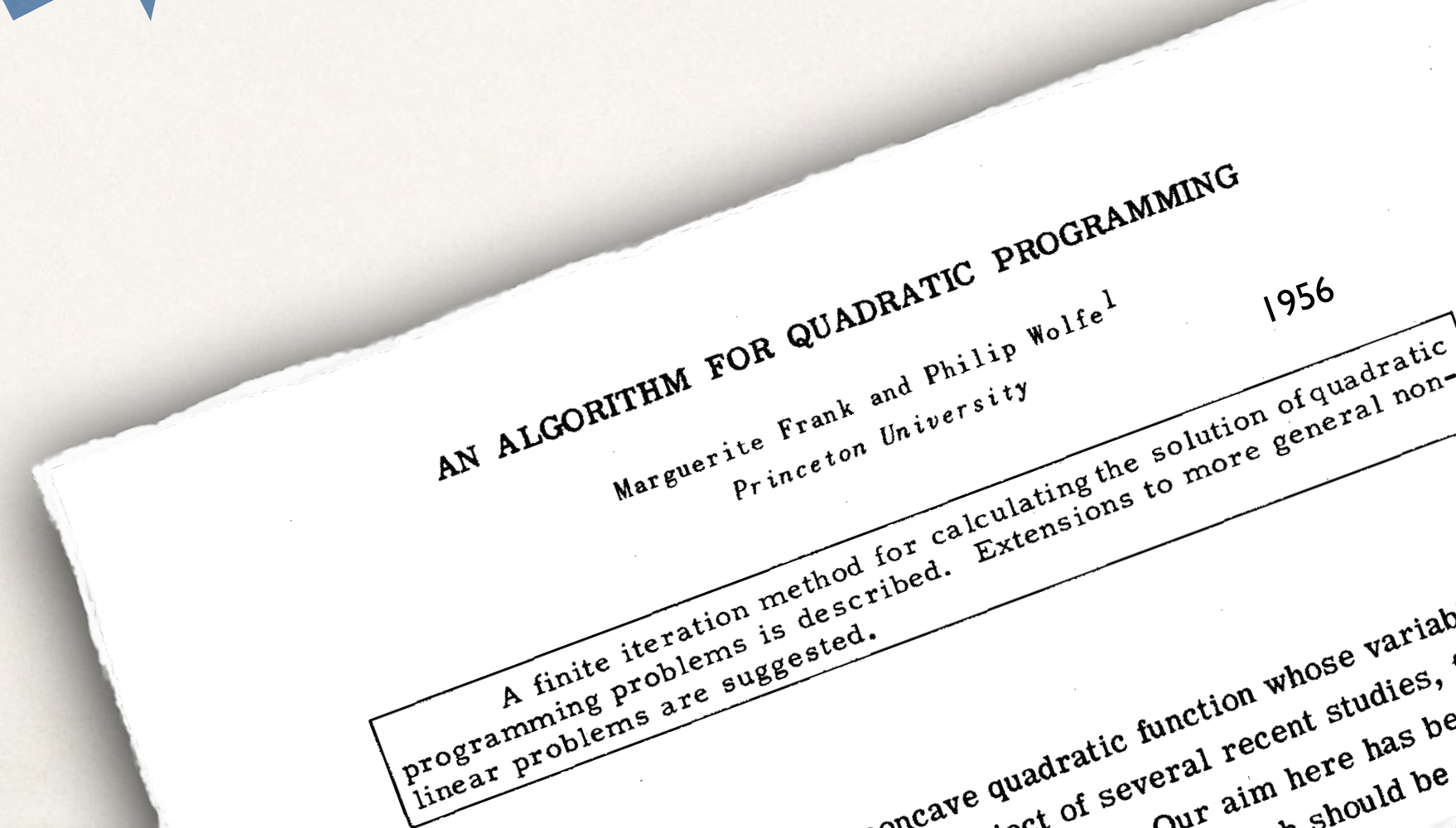


Linear Programming → Convex Programming

YouTube

"Marguerite Frank"

Optima Article (2014)



AN ALGORITHM FOR QUADRATIC PROGRAMMING

Marguerite Frank and Philip Wolfe[1]
Princeton University

1956

A finite iteration method for calculating the solution of quadratic programming problems is described. Extensions to more general non-linear problems are suggested.

concave quadratic function whose variab
ject of several recent studies, Our aim here has be
h should be

Variants of Frank-Wolfe

Simple Step-Size

Algorithm 1: Frank-Wolfe

Let $\boldsymbol{x}^{(0)} \in \mathcal{D}$

for $t = 0 \dots T$ **do**

Compute $\boldsymbol{s} := \mathrm{LMO}_{\mathcal{D}}\big(\nabla f(\boldsymbol{x}^{(t)})\big)$

Let $\gamma := \frac{2}{t+2}$

Update $\boldsymbol{x}^{(t+1)} := (1-\gamma)\boldsymbol{x}^{(t)} + \gamma \boldsymbol{s}$

end

Line-Search

Optimize γ by line-search

$$\gamma := \underset{\gamma \in [0,1]}{\arg\min} \; f\Big((1-\gamma)\boldsymbol{x}^{(t)} + \gamma \boldsymbol{s}\Big)$$

Fully Corrective /

Restricted Simplicial Decomposition

Update $\boldsymbol{x}^{(t+1)} := \underset{\boldsymbol{x} \in \mathrm{conv}(\boldsymbol{s}^{(0)}, \dots, \boldsymbol{s}^{(t+1)})}{\arg\min} f(\boldsymbol{x})$

[Holloway:1974]

Approximate Oracles

and Inexact Gradients

$$\varepsilon_t := \frac{C_f}{2+t}$$

Algorithm 1: Frank-Wolfe

Let $\boldsymbol{x}^{(0)} \in \mathcal{D}$

for $t = 0 \dots T$ **do**

Compute $\boldsymbol{s} := \mathrm{LMO}_{\mathcal{D}}\big(\nabla f(\boldsymbol{x}^{(t)})\big)$

Let $\gamma := \frac{2}{t+2}$

Update $\boldsymbol{x}^{(t+1)} := (1-\gamma)\boldsymbol{x}^{(t)} + \gamma \boldsymbol{s}$

end

- Approximate Oracle

$\mathrm{LMO}_{\mathcal{D},\varepsilon_t}(\boldsymbol{d})$: returns $\boldsymbol{s}$ s.t.

$$\langle \boldsymbol{d}, \boldsymbol{s} \rangle \leq \min_{\boldsymbol{s} \in \mathcal{D}} \langle \boldsymbol{d}, \boldsymbol{s} \rangle + \varepsilon_t$$

- Inexact Gradients

$\mathrm{LMO}_{\mathcal{D}}(\hat{\boldsymbol{d}}_x)$

Same Convergence Bounds!

Convergence Analysis

Primal Rate

$$f(\boldsymbol{x}^{(t)}) - f^* \leq \frac{2C_f}{t+2}$$

[Frank & Wolfe 1956,
Demyanov & Rubinov 1967
Dunn et al. 1978, 1980]

Primal-Dual

$$g(\boldsymbol{x}^{(\hat{t})}) \leq \frac{7C_f}{t+2}$$

efficient ***certificates*** for approximation quality

[asymptotically: Demyanov & R. 1967
simplex domain: Clarkson 2008
general: Jaggi 2011,2013]

Theorem 1 (Primal Convergence Rate). *For each $t \geq 1$, the iterates $\boldsymbol{x}^{(t)}$ of the Frank-Wolfe algorithm satisfy*

$$f(\boldsymbol{x}^{(t)}) - f^* \leq \frac{2C_f}{t+2} \ .$$

Proof. Let C_f be a constant s.t.

$$f(\boldsymbol{y}) \leq f(\boldsymbol{x}) + \gamma \underbrace{\langle \boldsymbol{s} - \boldsymbol{x}, \nabla f(\boldsymbol{x}) \rangle}_{-g(\boldsymbol{x})} + \frac{\gamma^2}{2} C_f \qquad \text{for all } \boldsymbol{x}, \boldsymbol{s} \in \mathcal{D}, \boldsymbol{y} := \boldsymbol{x} + \gamma(\boldsymbol{s} - \boldsymbol{x}), \gamma \in [0, 1]$$

Writing $h(\boldsymbol{x}^{(t)}) := f(\boldsymbol{x}^{(t)}) - f^*$ for the objective error, we have

$$\begin{aligned} h(\boldsymbol{x}^{(t+1)}) &\leq h(\boldsymbol{x}^{(t)}) - \gamma g(\boldsymbol{x}^{(t)}) + \frac{\gamma^2}{2} C_f && \text{definition of } C_f \\ &\leq h(\boldsymbol{x}^{(t)}) - \gamma h(\boldsymbol{x}^{(t)}) + \frac{\gamma^2}{2} C_f && h \leq g \text{ by convexity} \\ &= (1 - \gamma) h(\boldsymbol{x}^{(t)}) + \frac{\gamma^2}{2} C_f \ , \end{aligned}$$

From here, the decrease rate follows from a simple induction as in Lemma 2. □

Lemma 2. *Suppose a sequence of numbers h_t satisfies*

$$h_{t+1} \leq (1 - \gamma^{(t)}) h_t + {\gamma^{(t)}}^2 C \ ,$$

for $\gamma^{(t)} = \frac{2}{t+2}$, and $t = 0, 1, \ldots$, and a constant C. Then

$$h_t \leq \frac{4C}{t+2} \qquad t = 0, 1, \ldots$$

A Simple Duality Gap

Original Problem

$$\min_{\boldsymbol{x} \in \mathcal{D}} f(\boldsymbol{x})$$

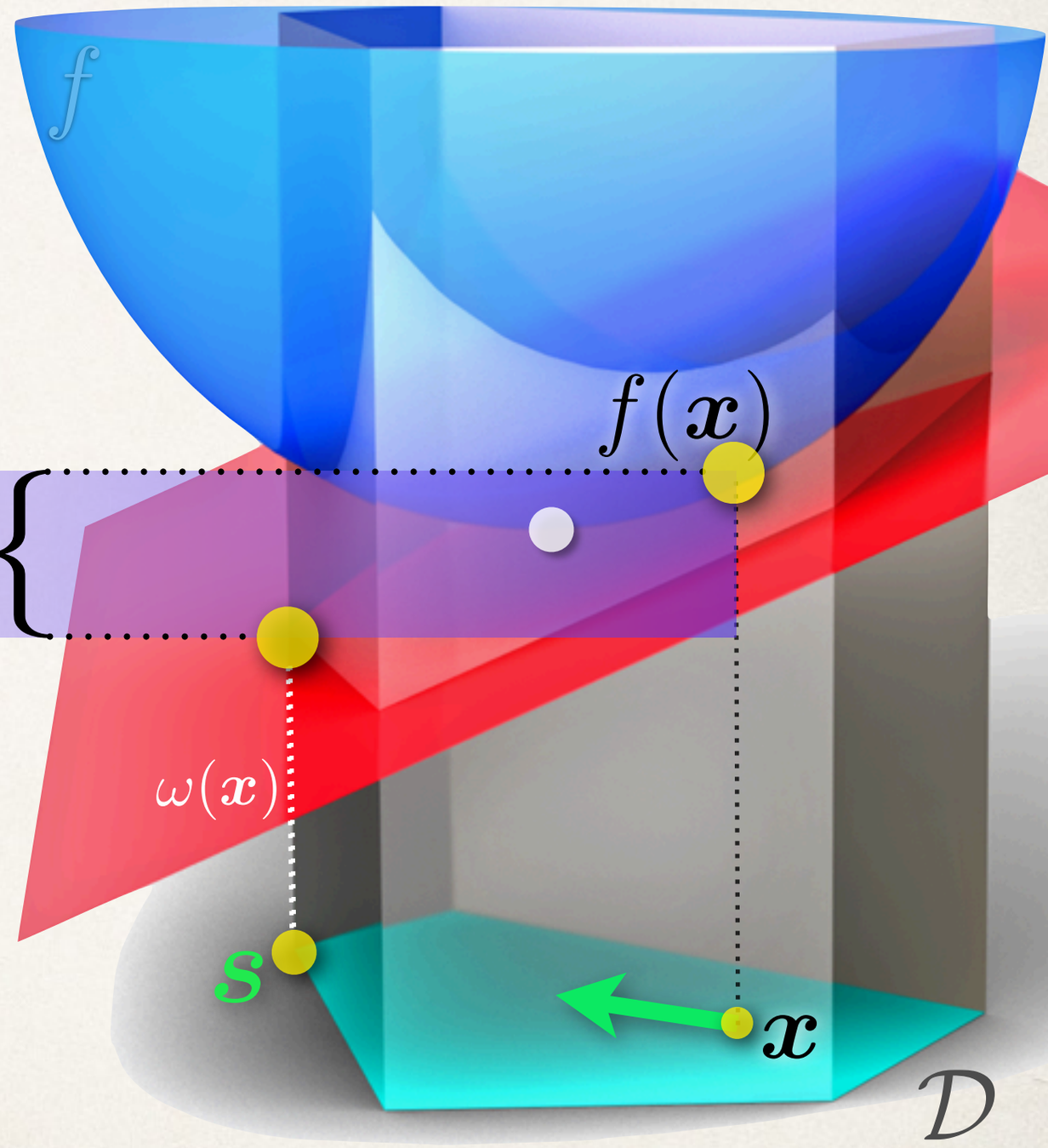
The Dual Function

$$\omega(\boldsymbol{x}) :=$$

$$\min_{\boldsymbol{s} \in \mathcal{D}} \; f(\boldsymbol{x}) + \langle \nabla f(\boldsymbol{x}), \boldsymbol{s} - \boldsymbol{x} \rangle$$

Weak Duality

$$\omega(\boldsymbol{x}) \leq f^* \leq f(\boldsymbol{x}')$$



Affine Invariance

(Geometry Independence)

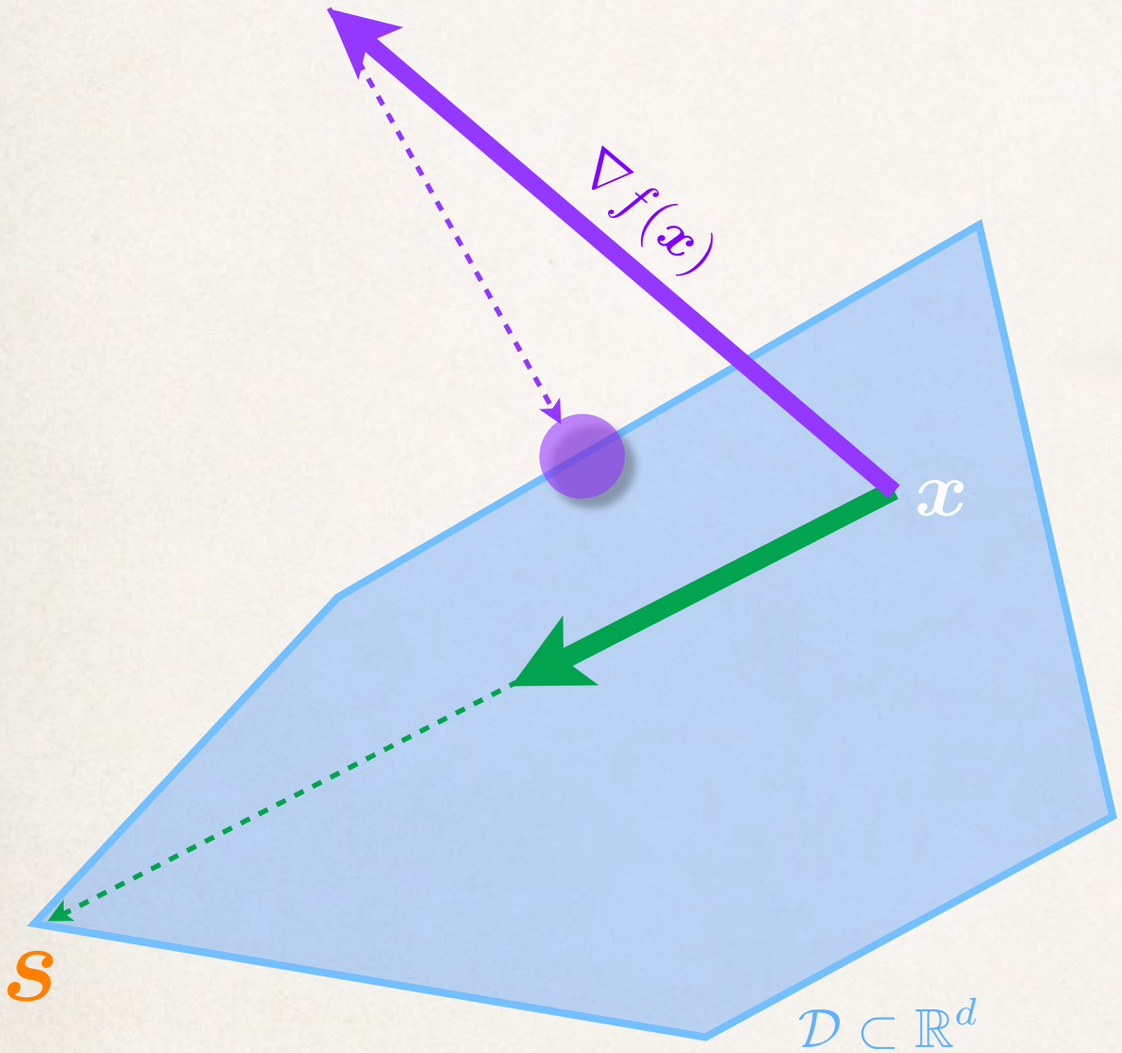
Convergence

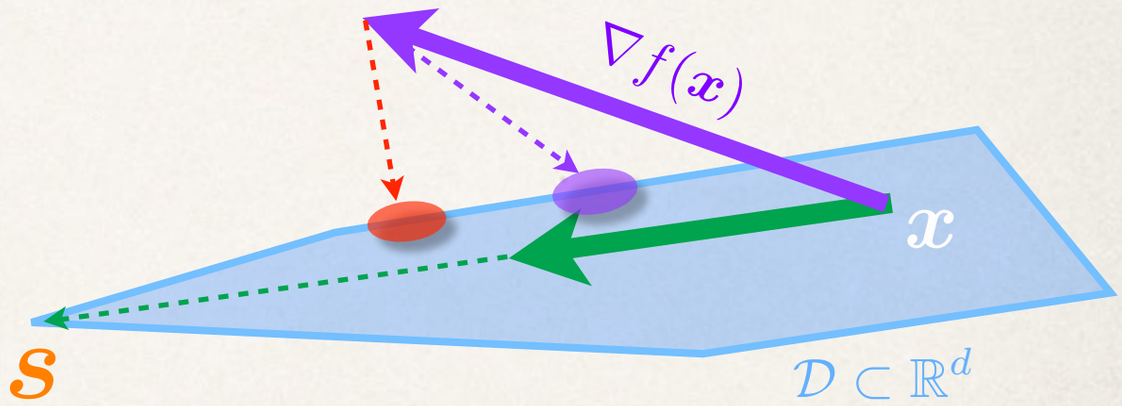
$$f(\boldsymbol{x}^{(t)}) - f^* \leq \frac{2C_f}{t+2}$$

Curvature constant

$$C_f := \sup_{\substack{\boldsymbol{x},\boldsymbol{s}\in\mathcal{D},\\ \gamma\in[0,1],\\ \boldsymbol{y}=\boldsymbol{x}+\gamma(\boldsymbol{s}-\boldsymbol{x})}} \frac{2}{\gamma^2}\big(f(\boldsymbol{y}) - f(\boldsymbol{x}) - \langle \boldsymbol{y}-\boldsymbol{x}, \nabla f(\boldsymbol{x})\rangle\big)$$

$$C_f \leq \mathrm{diam}_{\|.\|}(\mathcal{D})^2 L$$





Lower Bound

for Sparsity, and Optimality of the Convergence Rate

Error

$$\geq \frac{C_f}{t}$$

For an ***optimization accuracy*** of $\frac{C_f}{t}$,
any x must have ***at least t non-zeros*** (use ***at least t corners***)

Trade-Off: Optimization Accuracy vs **Sparsity**

$$f(x) := \|\boldsymbol{x}\|^2$$
$$\mathcal{D} := \Delta$$

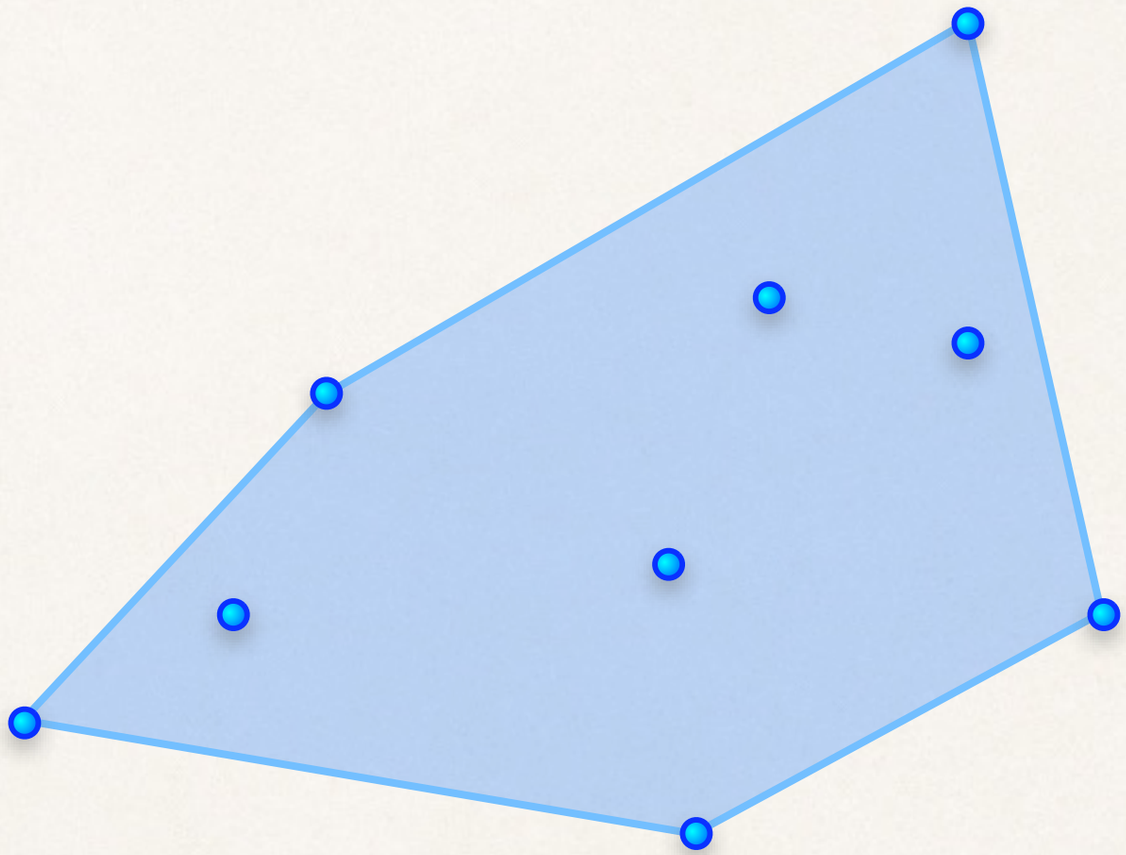
[Jaggi 2011]

Greedy Optimization

over Atomic Sets

$$\min_{\boldsymbol{x} \in \mathcal{D}} f(\boldsymbol{x})$$

Atoms $\mathcal{A}$



$$\mathcal{D} := \operatorname{conv}(\mathcal{A})$$

Frank-Wolfe:

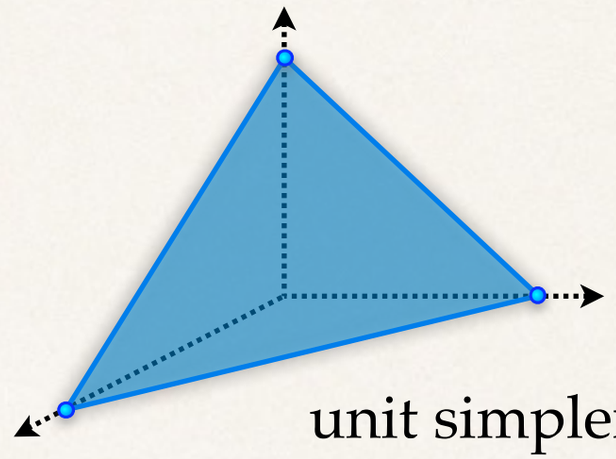
$$\boldsymbol{x}^{(t+1)} := (1-\gamma)\boldsymbol{x}^{(t)} + \gamma \boldsymbol{s} \quad \text{for} \quad \boldsymbol{s} \in \mathcal{A}$$

[Chandrasekaran et al. 2012]

Sparse Approximation

$$\min_{\boldsymbol{x}\in\Delta} \; f(\boldsymbol{x})$$

$$\mathcal{D} := \mathrm{conv}(\{\mathbf{e}_i \mid i \in [n]\})$$



unit simplex

Corollary:
Obtain $\frac{2C_f}{t}$ -approximate solution of **sparsity** t .

[Clarkson 2008]

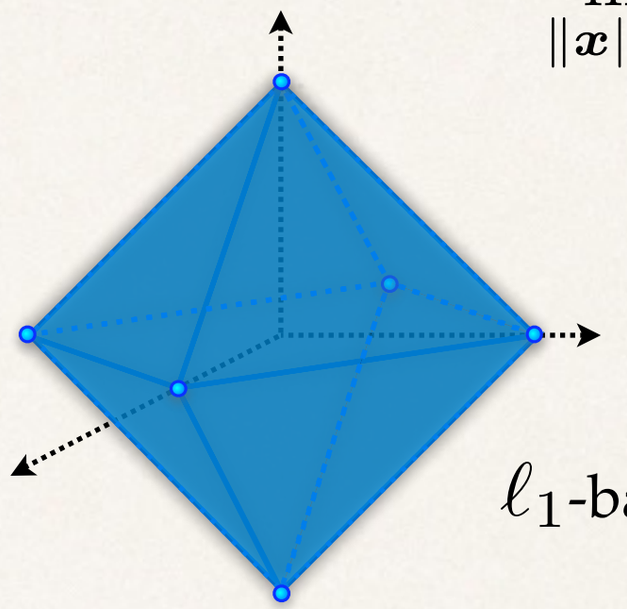
Trade-Off:
Optimization accuracy vs **sparsity**

lower bound:
$\frac{C_f}{t}$

Sparse Approximation

$$\min_{\|\boldsymbol{x}\|_1 \leq 1} \quad f(\boldsymbol{x})$$

$$\mathcal{D} := \operatorname{conv}(\{\pm \mathbf{e}_i \mid i \in [n]\})$$



ℓ_1-ball

Corollary:
Obtain $\frac{2C_f}{t}$ -approximate solution of **sparsity** t .

lower bound:
$\frac{C_f}{t}$

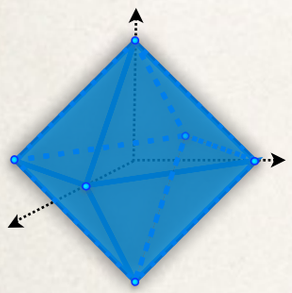
Trade-Off:
Optimization accuracy vs **sparsity**

Greedy Optimization Meets Frank-Wolfe

Convex optimization

methods applied to

$$\min_{\boldsymbol{x}\in\mathcal{D}} \|A\boldsymbol{x}-\boldsymbol{b}\|^2$$



ℓ_1-ball

Signal processing

sparse/direct recovery methods

recover a sparse **x** from a noisy measurement **b**

Frank-Wolfe ⟷ (selects the same atom per step) ⟷ **Matching Pursuit**

$$i := \arg\max_i |\nabla f(x)_i|$$

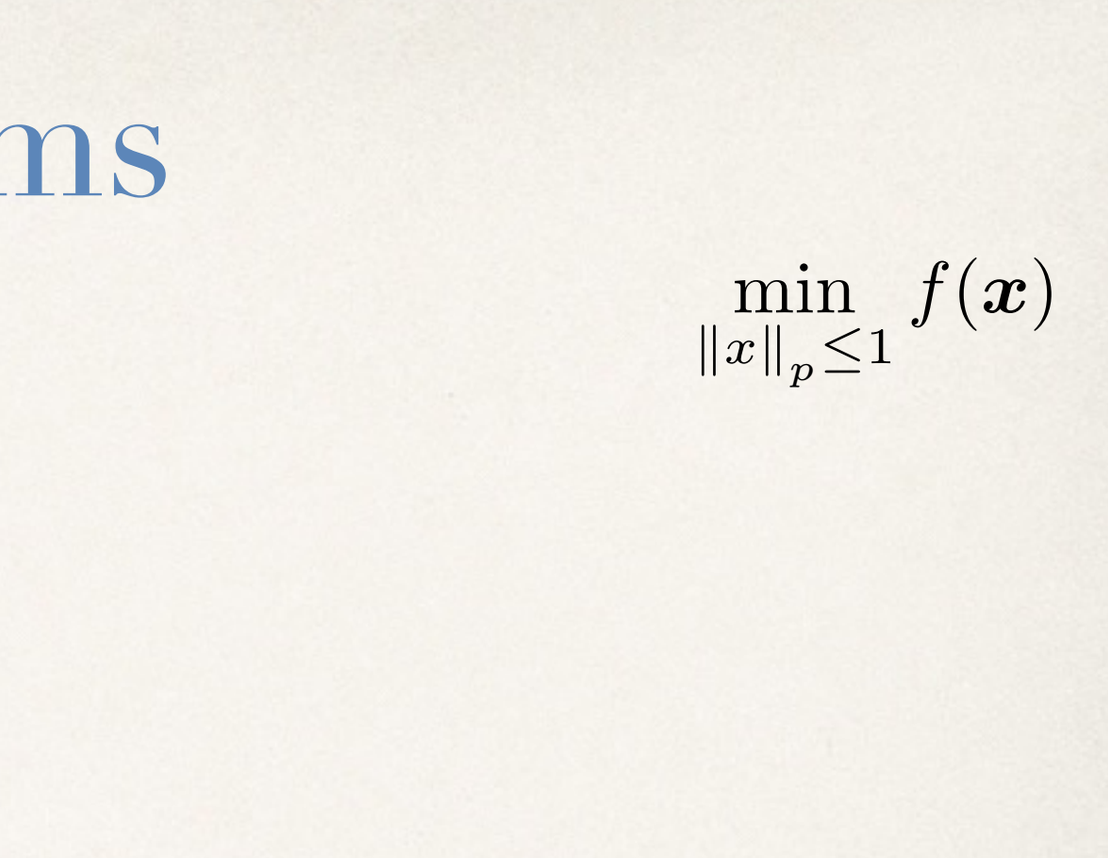
fully corrective **Frank-Wolfe** ⟷ (equivalent to) ⟷ **OMP**

ℓ_p - Norm Problems

$$\min_{\|x\|_p \leq 1} f(\boldsymbol{x})$$

$\mathcal{D} := \ell_p$-ball

$1 < p < \infty$



Projection:	*LMO:*
unknown?	linear time

Low Rank Approximation

$$\min_{\|X\|_* \leq 1} \quad f(X)$$

$$\mathcal{D} := \operatorname{conv}\left(\left\{ \boldsymbol{u}\boldsymbol{v}^T \;\middle|\; \begin{matrix} \boldsymbol{u}\in\mathbb{R}^n,\, \|\boldsymbol{u}\|_2=1 \\ \boldsymbol{v}\in\mathbb{R}^m,\, \|\boldsymbol{v}\|_2=1 \end{matrix} \right\} \right)$$

nuclear-norm-ball

Corollary:
Obtain $\frac{2C_f}{t}$ -approximate solution of **rank** t .

[Jaggi & Sulovský 2010]

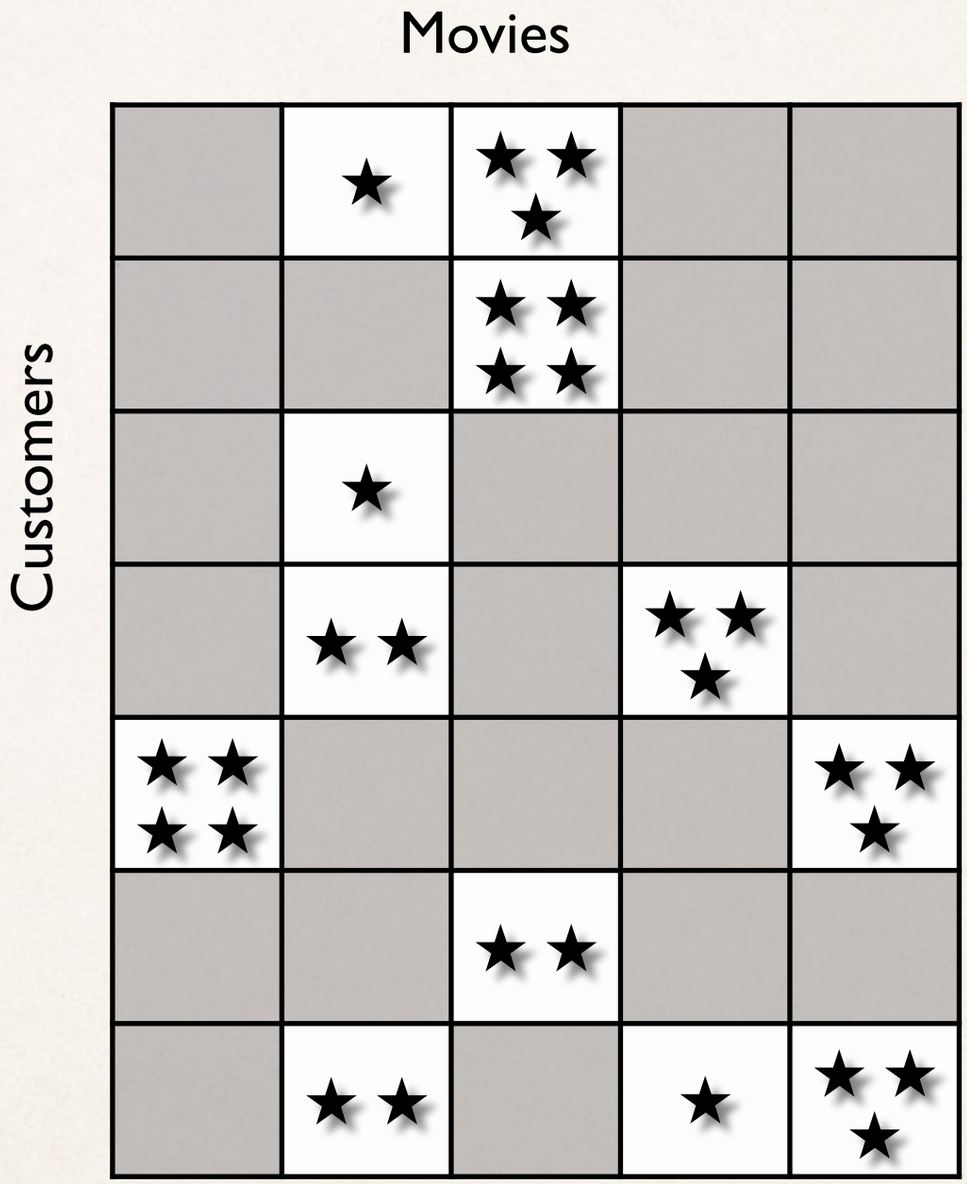
lower bound:
$\frac{C_f}{t}$

Trade-Off:
Optimization accuracy vs **rank**

Projection:	*LMO:*
full SVD	approx. top singular vector

Low Rank Approximation

$$\min_{\|X\|_* \leq 1} \quad f(X)$$



$$\approx UV^T$$

[Jaggi & Sulovský 2010]

Factorized Matrix Domains

(and *tensor*)

$$\mathcal{D} := \operatorname{conv}\left(\left\{\boldsymbol{u}\boldsymbol{v}^T \;\middle|\; \begin{matrix}\boldsymbol{u}\in\mathbb{R}^n,\ \|\boldsymbol{u}\|_2=1\\ \boldsymbol{v}\in\mathbb{R}^m,\ \|\boldsymbol{v}\|_2=1\end{matrix}\right\}\right)$$

(nuclear norm)

$$\mathcal{D} := \operatorname{conv}\left(\left\{\boldsymbol{u}\boldsymbol{v}^T \;\middle|\; \begin{matrix}\boldsymbol{u}\in\mathcal{A}_{\text{left}}\\ \boldsymbol{v}\in\mathcal{A}_{\text{right}}\end{matrix}\right\}\right) \quad \subset \mathbb{R}^{n\times m}$$

$\mathcal{A}_{\text{left}} \subset \mathbb{R}^n$

$\mathcal{A}_{\text{right}} \subset \mathbb{R}^m$

r	$\mathcal{A}_{\text{left}} \subseteq \mathbb{R}^{m\times r}$	$\mathcal{A}_{\text{right}} \subseteq \mathbb{R}^{n\times r}$	$\Omega_{\text{conv}(\mathcal{A})}(M)$	$\Omega^*_{\mathcal{A}}(M)$	**FW step**
1	$\|.\|_2$-sphere	$\|.\|_2$-sphere	Trace norm $\|M\|_{tr}$	$\|M\|_{op}$	Lanczos, see Table 1
1	$\|.\|_1$-sphere	$\|.\|_1$-sphere	Vector ℓ_1-norm $\|\vec{M}\|_1$	$\|\vec{M}\|_\infty$	$O(nm)$
1	$\|.\|_\infty$-sphere	$\|.\|_\infty$-sphere		Cut-norm $\|.\|_{\infty\to 1}$	NP-hard (Alon & Naor, 2006)
n+m	$\|.\|_{2,\infty}$	$\|.\|_{2,\infty}$	Max-norm $\|M\|_{\max}$		SDP, see Table 1
1	$\|.\|_2 \cap \mathbb{R}^m_{\geq 0}$	$\|.\|_2 \cap \mathbb{R}^n_{\geq 0}$	"non-neg. trace norm"		NP-hard (Murty & Kabadi, 1987)
1	Simplex Δ_m	Simplex Δ_n	"non-neg. matrix ℓ_1-norm"		$O(nm)$

Table 2. Examples of some factorized matrix norms on $\mathbb{R}^{m\times n}$, each induced by two atomic norms (last two rows giving

Examples of *Atomic Domains* Suitable for Frank-Wolfe

$\mathcal{X}$	**Optimization Domain** Atoms $\mathcal{A}$	$\mathcal{D} = \text{conv}(\mathcal{A})$	**Complexity of one Frank-Wolfe Iteration** $\sup_{\boldsymbol{s}\in\mathcal{D}}\langle \boldsymbol{s}, \boldsymbol{y}\rangle$	Complexity
$\mathbb{R}^n$	Sparse Vectors	$\|.\|_1$-ball	$\|\boldsymbol{y}\|_\infty$	$O(n)$
$\mathbb{R}^n$	Sign-Vectors	$\|.\|_\infty$-ball	$\|\boldsymbol{y}\|_1$	$O(n)$
$\mathbb{R}^n$	ℓ_p-Sphere	$\|.\|_p$-ball	$\|\boldsymbol{y}\|_q$	$O(n)$
$\mathbb{R}^n$	Sparse Non-neg. Vectors	Simplex Δ_n	$\max_i\{\boldsymbol{y}_i\}$	$O(n)$
$\mathbb{R}^n$	Latent Group Sparse Vec.	$\|.\|_{\mathcal{G}}$-ball	$\max_{g\in\mathcal{G}} \left\|\boldsymbol{y}_{(g)}\right\|_g^*$	$\sum_{g\in\mathcal{G}} \|g\|$
$\mathbb{R}^{m\times n}$	Matrix Trace Norm	$\|.\|_{tr}$-ball	$\|\boldsymbol{y}\|_{op} = \sigma_1(\boldsymbol{y})$	$\tilde{O}\big(N_f/\sqrt{\varepsilon'}\big)$ (Lanczos)
$\mathbb{R}^{m\times n}$	Matrix Operator Norm	$\|.\|_{op}$-ball	$\|\boldsymbol{y}\|_{tr} = \|(\sigma_i(\boldsymbol{y}))\|_1$	SVD
$\mathbb{R}^{m\times n}$	Schatten Matrix Norms	$\|(\sigma_i(.))\|_p$-ball	$\|(\sigma_i(\boldsymbol{y}))\|_q$	SVD
$\mathbb{R}^{m\times n}$	Matrix Max-Norm	$\|.\|_{\max}$-ball		$\tilde{O}\big(N_f(n+m)^{1.5}/\varepsilon'^{2.5}\big)$
$\mathbb{R}^{n\times n}$	Permutation Matrices	Birkhoff polytope		$O(n^3)$
$\mathbb{R}^{n\times n}$	Rotation Matrices			SVD (Procrustes prob.)
$\mathbb{S}^{n\times n}$	Rank-1 PSD matrices of unit trace	$\{\boldsymbol{x}\succeq 0,\ \text{Tr}(\boldsymbol{x})=1\}$	$\lambda_{\max}(\boldsymbol{y})$	$\tilde{O}\big(N_f/\sqrt{\varepsilon'}\big)$ (Lanczos)
$\mathbb{S}^{n\times n}$	PSD matrices of bounded diagonal	$\{\boldsymbol{x}\succeq 0,\ \boldsymbol{x}_{ii}\leq 1\}$		$\tilde{O}\big(N_f\, n^{1.5}/\varepsilon'^{2.5}\big)$

Table 1: Some examples of atomic domains suitable for optimization using the Frank-Wolfe algorithm. *Here SVD refers to the complexity of computing a singular value decomposition, which is* $O(\min\{mn^2, m^2n\})$. N_f *is the number of non-zero entries in the gradient of the objective function* f, *and* $\varepsilon' = \frac{2\delta C_f}{k+2}$ *is the required accuracy for the linear subproblems. For any* $p \in [1,\infty]$, *the conjugate value* q *is meant to satisfy* $\frac{1}{p}+\frac{1}{q}=1$, *allowing* $q=\infty$ *for* $p=1$ *and vice versa.*

Examples of *Atomic Domains* Suitable for Frank-Wolfe

- shortest paths / network flows / transportation
- greedy selection and sparse optimization
- low-rank matrix factorizations, many other matrices
- wavelets (infinite-dimensional)
- structured sparsity and structured prediction
- total-variation-norm for image de-noising
- submodular optimization
- boosting
- training deep networks

Linear Convergence

for strongly convex f, under additional assumptions

[Guélat & Marcotte, 1986, Garber & Hazan 2013]

Thm: *The iterates of* **standard FW** *(with line-search) satisfy*

$$f(\boldsymbol{x}^{(t)}) - f^* \leq h_0 \exp(-\frac{1}{2}\frac{\mu_f^{\text{int}}}{C_f} t)$$

[Lacoste-Julien & Jaggi 2013]

$$\frac{\mu_f^{\text{int}} \geq \mu \cdot min\, dist(x^*, \partial\mathcal{D})^2}{C_f \leq L \cdot \text{diam}(\mathcal{D})^2}$$

complexity measure *of function,* *of domain (and opt x)*

Faster Convergence

under additional assumptions

An affine invariant notion of *smoothness* and *strong convexity*

Curvature constant

$$C_f := \sup_{\substack{\boldsymbol{x},\boldsymbol{s}\in\mathcal{D},\\ \gamma\in[0,1],\\ \boldsymbol{y}=\boldsymbol{x}+\gamma(\boldsymbol{s}-\boldsymbol{x})}} \frac{2}{\gamma^2}\big(f(\boldsymbol{y}) - f(\boldsymbol{x}) - \langle \boldsymbol{y}-\boldsymbol{x}, \nabla f(\boldsymbol{x})\rangle\big)$$

$$C_f \leq L \cdot \operatorname{diam}(\mathcal{D})^2$$

Strong convexity

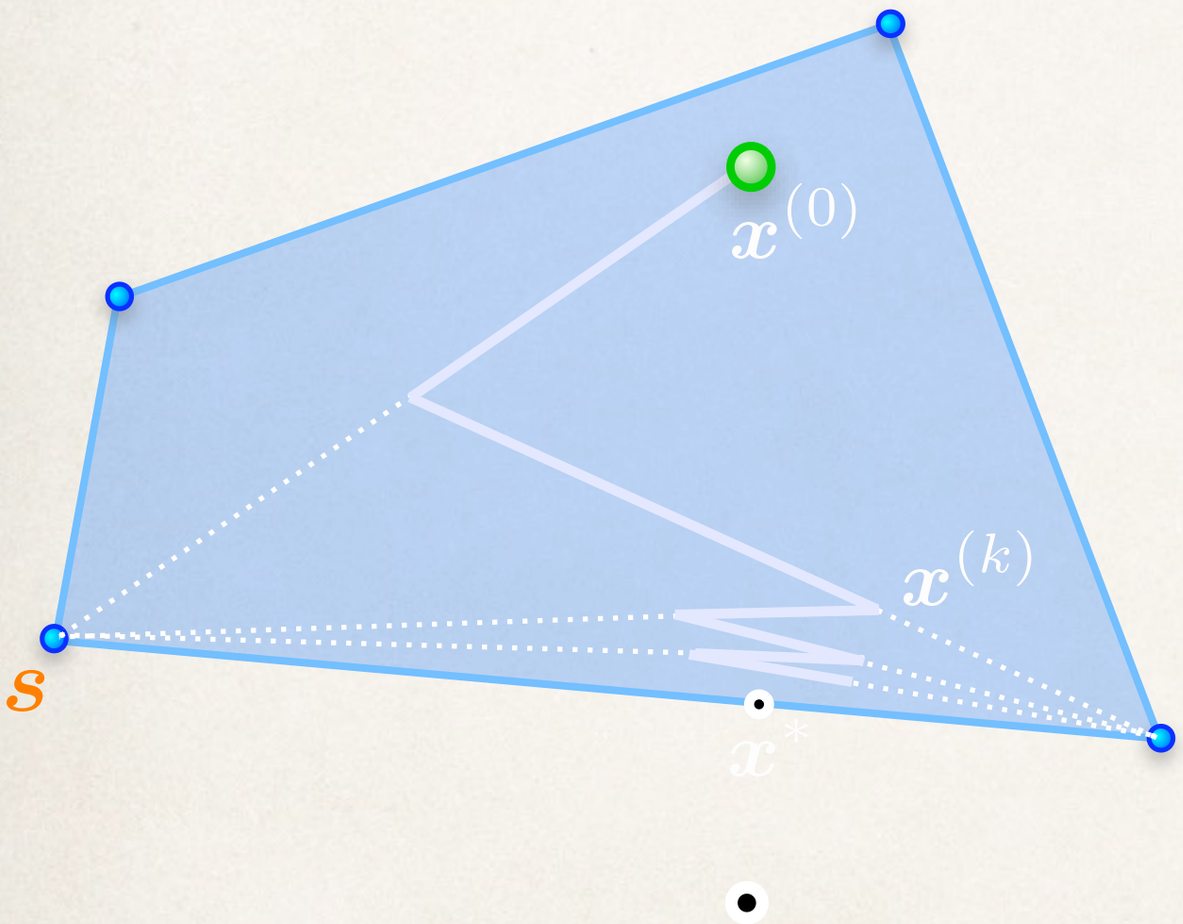
$$\mu_f^{\text{int}} := \inf_{\substack{\boldsymbol{x}\in\mathcal{D}\setminus\{\boldsymbol{x}^*\},\\ \gamma\in(0,1],\\ \overline{\boldsymbol{s}}:=\operatorname{ray}(\boldsymbol{x},\boldsymbol{x}^*)\cap\partial\mathcal{D},\\ \boldsymbol{y}=\boldsymbol{x}+\gamma(\overline{\boldsymbol{s}}-\boldsymbol{x})}} \frac{2}{\gamma^2}\big(f(\boldsymbol{y}) - f(\boldsymbol{x}) - \langle \boldsymbol{y}-\boldsymbol{x}, \nabla f(\boldsymbol{x})\rangle\big)$$

[Lacoste-Julien & Jaggi 2013]

$$\mu_f^{\text{int}} \geq \mu \cdot min\, dist(x^*, \partial\mathcal{D})^2$$

Frank-Wolfe with Away-Steps

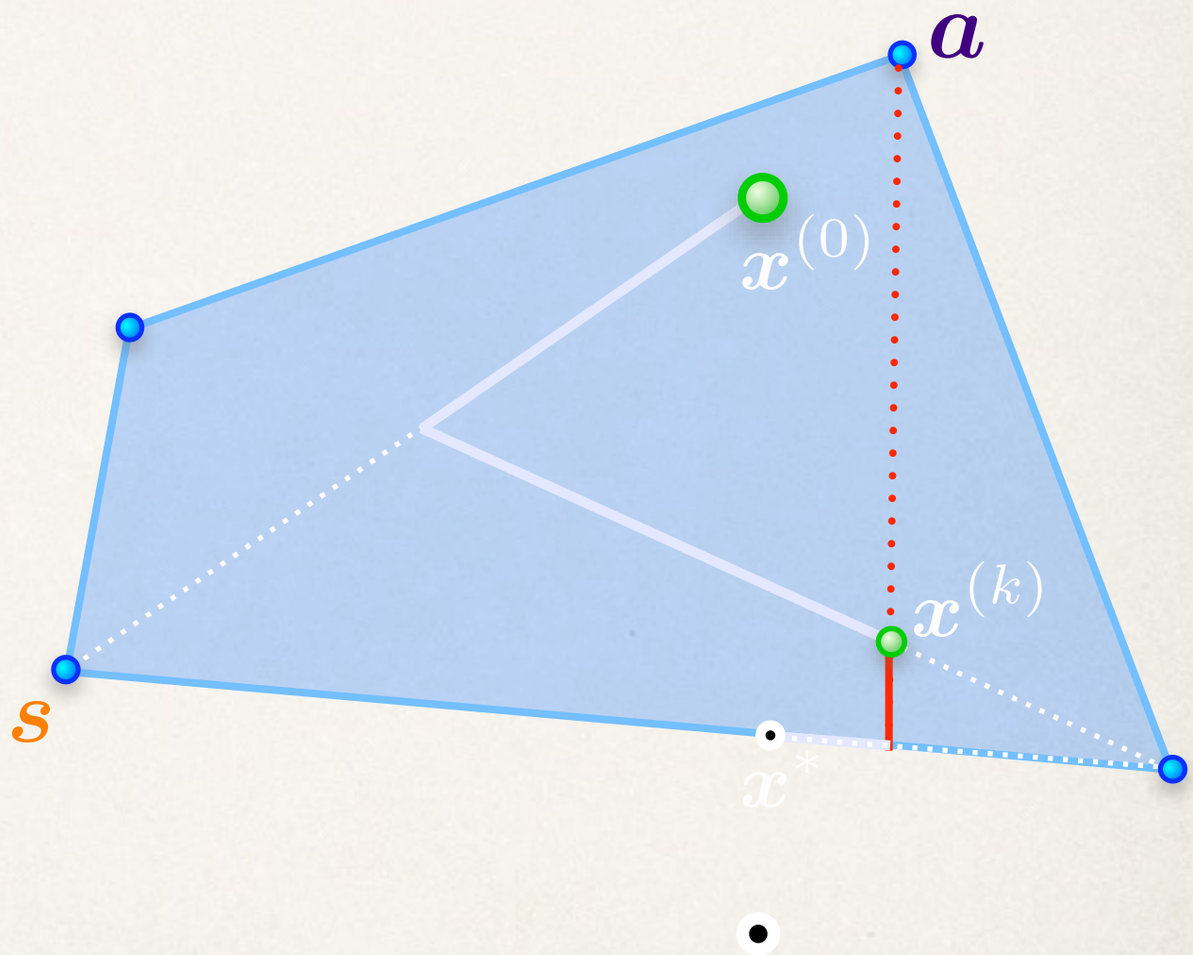
Standard FW



$$s := \mathrm{LMO}_{\mathcal{D}}(d) = \arg\min_{s \in \mathcal{D}} \langle d, s \rangle$$

FW with away steps

[Wolfe 1970, Guélat et al. 1986]



$$a := \arg\max_{a \in \{s^{(0)}, \ldots, s^{(t)}\}} \langle d, a \rangle$$

Linear Convergence

for strongly convex f, under additional assumptions

[Guélat & Marcotte, 1986, Garber & Hazan 2013]

Thm: *The iterates of* **standard FW** *(with line-search) satisfy*

$$f(\boldsymbol{x}^{(t)}) - f^* \leq h_0 \exp(-\frac{1}{2}\frac{\mu_f^{\text{int}}}{C_f} t)$$

[Lacoste-Julien & Jaggi 2013]

$$\frac{\mu_f^{\text{int}} \geq \mu \cdot min\, dist(x^*, \partial\mathcal{D})^2}{C_f \leq L \cdot \text{diam}(\mathcal{D})^2}$$

complexity measure *of function,* *of domain (and opt x)*

Linear Convergence

for strongly convex f, under additional assumptions

[Lacoste-Julien & Jaggi 2013, Beck & Shtern 2015]

Thm: *The iterates of* **FW with away-steps** *satisfy*

$$f(\boldsymbol{x}^{(t)}) - f^* \leq h_0 \exp(-\frac{1}{2}\frac{\mu_f^{\text{away}}}{C_f} t)$$

[Lacoste-Julien & Jaggi 2013]

$$\frac{\mu_f^{\text{away}} \geq \mu \cdot pyr\,width(\mathcal{D})^2}{C_f \leq L \cdot \text{diam}(\mathcal{D})^2}$$

complexity measure *of function,* *of domain* ~~*(and opt x)*~~

Extensions & Active Research

- Non-Smooth f

  [Lan 2013, Harchaoui et al. 2014, Hazan et al. 2012]

- Penalized Version

  [Harchaoui et al. 2014 , Zhang et al. 2012]

- Block-Wise Version

  [Lacoste-Julien et al. 2013, Beck et al. 2015]

- Submodular Minimization

  [Bach 2014]

- Faster Convergence under Additional Assumptions

  [Guélat & Marcotte, 1986, Garber & Hazan 2013,2014, Lacoste-Julien & Jaggi 2013, Beck & Shtern 2015]

More Information

https://sites.google.com/site/FrankWolfeGreedyTutorial

- Sample Code
- Bibliography
- Tutorial Slides
- Topics not covered