Fenchel-type Representations and Large-scale Problems with Convex Structure on Difficult Geometry Domains

Arkadi Nemirovski
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Joint research with Bruce Cox[†] & Anatoli Juditsky[‡]
[†]US AirForce [‡]University J. Fourier, Grenoble

Optimization and Big Data 2015
Edinburgh May 2015

Problem of Interest: large-scale problems with convex structure:

- **C**onvex **M**inimization: $\min_{x\in X} f(x)$
  - X: convex compact set in Euclidean space
  - $f : X \to \mathbb{R}$: convex & Lipschitz
- Convex-Concave **S**addle **P**oint: $\min_{u\in U} \max_{v\in V} f(u, v)$
- **V**ariational **I**nequality with monotone operator:

$$\text{Find } x \in X : \langle F(y), y - x\rangle \geq 0\ \forall y \in X$$

Note: CM and **SP** reduce to **VI**

- **CM** $\Rightarrow$ **VI**: F: bounded section of $x \mapsto \partial_X f$
- **SP** $\Rightarrow$ **VI**: $X := U \times V$, F: bounded section of

$$[u; v] \mapsto [\partial_u f(u, v); \partial_v[-f(u, v)]]$$

$\Rightarrow$ *Every problem with convex structure induces a monotone vector field on problem's domain.*

Problem of Interest: large-scale problems with convex structure:

- **C**onvex **M**inimization: $\min_{x \in X} f(x)$
- Convex-Concave **S**addle **P**oint: $\min_{u \in U} \max_{v \in V} f(u, v)$
  - U, V: convex compact sets in Euclidean spaces
  - $f(\cdot, \cdot) : X \to \mathbb{R}$: convex-concave & Lipschitz
- **V**ariational **I**nequality with monotone operator:

$$\text{Find } x \in X : \langle F(y), y - x \rangle \geq 0 \; \forall y \in X$$

Note: CM and **SP** reduce to **VI**

- **CM** $\Rightarrow$ **VI**: F: bounded section of $x \mapsto \partial_X f$
- **SP** $\Rightarrow$ **VI**: $X := U \times V$, F: bounded section of

$$[u; v] \mapsto [\partial_u f(u, v); \partial_v [-f(u, v)]]$$

$\Rightarrow$ *Every problem with convex structure induces a monotone vector field on problem's domain.*

Problem of Interest: large-scale problems with convex structure:

- **C**onvex **M**inimization: $\min_{x \in X} f(x)$
- Convex-Concave **S**addle **P**oint: $\min_{u \in U} \max_{v \in V} f(u, v)$
- **V**ariational **I**nequality with monotone operator:

$$\text{Find } x \in X : \langle F(y), y - x \rangle \geq 0 \; \forall y \in X$$

  - X: convex compact set in Euclidean space E
  - $F : X \to E$: monotone & bounded

Note: **CM** and **SP** reduce to **VI**

- **CM** $\Rightarrow$ **VI**: F: bounded section of $x \mapsto \partial_X f$
- **SP** $\Rightarrow$ **VI**: $X := U \times V$, F: bounded section of

$$[u; v] \mapsto [\partial_u f(u, v); \partial_v [-f(u, v)]]$$

$\Rightarrow$ *Every problem with convex structure induces a monotone vector field on problem's domain.*

Problem of Interest: large-scale problems with convex structure:

- **C**onvex **M**inimization: $\min_{x \in X} f(x)$
- Convex-Concave **S**addle **P**oint: $\min_{u \in U} \max_{v \in V} f(u, v)$
- **V**ariational **I**nequality with monotone operator:

$$\text{Find } x \in X : \langle F(y), y - x \rangle \geq 0 \; \forall y \in X$$

Note: CM and **SP** reduce to **VI**

- **CM** $\Rightarrow$ **VI**: F: bounded section of $x \mapsto \partial_X f$
- **SP** $\Rightarrow$ **VI**: $X := U \times V$, F: bounded section of

$$[u; v] \mapsto [\partial_u f(u, v); \partial_v [-f(u, v)]]$$

$\Rightarrow$ *Every problem with convex structure induces a monotone vector field on problem's domain.*

Problem of Interest: large-scale problems with convex structure:

- **C**onvex **M**inimization: $\min_{x\in X} f(x)$
- Convex-Concave **S**addle **P**oint: $\min_{u\in U}\max_{v\in V} f(u,v)$
- **V**ariational **I**nequality with monotone operator:

$$\text{Find } x\in X : \langle F(y), y-x\rangle \geq 0\ \forall y\in X$$

Note: CM and **SP** reduce to **VI**

- **CM** $\Rightarrow$ **VI**: F: bounded section of $x\mapsto \partial_X f$
- **SP** $\Rightarrow$ **VI**: $X := U\times V$, F: bounded section of

$$[u;v]\mapsto[\partial_u f(u,v); \partial_v[-f(u,v)]]$$

$\Rightarrow$ *Every problem with convex structure induces a monotone vector field on problem's domain.*

Problem of Interest: large-scale problems with convex structure:

- **C**onvex **M**inimization: $\min_{x\in X} f(x)$
- Convex-Concave **S**addle **P**oint: $\min_{u\in U}\max_{v\in V} f(u,v)$
- **V**ariational **I**nequality with monotone operator:

$$\text{Find } x\in X : \langle F(y), y-x\rangle \geq 0\ \forall y\in X$$

Note: CM and **SP** reduce to **VI**

- **CM** $\Rightarrow$ **VI**: F: bounded section of $x\mapsto \partial_X f$
- **SP** $\Rightarrow$ **VI**: $X := U\times V$, F: bounded section of

$$[u;v]\mapsto [\partial_u f(u,v);\partial_v[-f(u,v)]]$$

$\Rightarrow$ *Every problem with convex structure induces a monotone vector field on problem's domain.*

Fact: At present, main tools of large-scale convex optimization are proximal **F**irst **O**rder **M**ethods where a linear perturbation of fixed *strongly convex* function is minimized over problem's domain X at every step.
$\Rightarrow$ Proximal **FOM**s require X to be *proximal friendly* – to allow for cheap minimization of all linear perturbations of a strongly convex function. In particular, *X should admit a cheap* **L***inear* **M***inimization* **O***racle* – a routine capable to minimize over X linear forms.

Fact: Some important high-dimensional domains admit cheap **LMO** and are not proximal friendly:

- nuclear/trace norm balls [low rank matrix recovery & SDP]
- total variation ball [imaging]
- some combinatorial polytopes

$\Rightarrow$ **Challenge:** *Design of* **LMO** *based First Order Methods*

Fact: At present, main tools of large-scale convex optimization are proximal **F**irst **O**rder **M**ethods where a linear perturbation of fixed *strongly convex* function is minimized over problem's domain X at every step.
$\Rightarrow$ Proximal **FOM**s require X to be *proximal friendly* – to allow for cheap minimization of all linear perturbations of a strongly convex function. In particular, *X should admit a cheap* ***L****inear* ***M****inimization* ***O****racle* – a routine capable to minimize over X linear forms.

Fact: Some important high-dimensional domains admit cheap **LMO** and are not proximal friendly:

- nuclear/trace norm balls [low rank matrix recovery & SDP]
- total variation ball [imaging]
- some combinatorial polytopes

$\Rightarrow$ **Challenge:** *Design of* **LMO** *based First Order Methods*

Fact: At present, main tools of large-scale convex optimization are proximal **F**irst **O**rder **M**ethods where a linear perturbation of fixed *strongly convex* function is minimized over problem's domain X at every step.
$\Rightarrow$ Proximal **FOM**s require X to be *proximal friendly* – to allow for cheap minimization of all linear perturbations of a strongly convex function. In particular, *X should admit a cheap* ***L****inear* ***M****inimization* ***O****racle* – a routine capable to minimize over X linear forms.

Fact: Some important high-dimensional domains admit cheap **LMO** and are not proximal friendly:

- nuclear/trace norm balls [low rank matrix recovery & SDP]
- total variation ball [imaging]
- some combinatorial polytopes

$\Rightarrow$ **Challenge:** *Design of* **LMO** *based First Order Methods*

Fact: At present, main tools of large-scale convex optimization are proximal **F**irst **O**rder **M**ethods where a linear perturbation of fixed *strongly convex* function is minimized over problem's domain X at every step.

$\Rightarrow$ Proximal **FOM**s require X to be *proximal friendly* – to allow for cheap minimization of all linear perturbations of a strongly convex function. In particular, *X should admit a cheap* ***L****inear* ***M****inimization* ***O****racle* – a routine capable to minimize over X linear forms.

Fact: Some important high-dimensional domains admit cheap **LMO** and are not proximal friendly:

- nuclear/trace norm balls [low rank matrix recovery & SDP]
- total variation ball [imaging]
- some combinatorial polytopes

$\Rightarrow$ **Challenge:** *Design of* **LMO** *based First Order Methods*

Fact: At present, main tools of large-scale convex optimization are proximal **F**irst **O**rder **M**ethods where a linear perturbation of fixed *strongly convex* function is minimized over problem's domain X at every step.
$\Rightarrow$ Proximal **FOM**s require X to be *proximal friendly* – to allow for cheap minimization of all linear perturbations of a strongly convex function. In particular, *X should admit a cheap **L**inear **M**inimization **O**racle* – a routine capable to minimize over X linear forms.

Fact: Some important high-dimensional domains admit cheap **LMO** and are not proximal friendly:

- nuclear/trace norm balls [low rank matrix recovery & SDP]
- total variation ball [imaging]
- some combinatorial polytopes

$\Rightarrow$ **Challenge:** *Design of **LMO** based First Order Methods*

Fact: The only traditional optimization technique capable to solve problems with convex structure on **LMO** represented domains is the classical **C***onditional* **G***radient* algorithm (Frank & Wolfe, 1958).

But: The *only* problem with convex structure in the scope of **CG** is *smooth* **CM**. There are *no* traditional LMO-based **FOM**s for (smooth and nonsmooth alike) **SP** and **VI** and *nonsmooth* **CM** problems.

Goal of the talk: **LMO** based **FOM**s for **SP**/**VI**/nonsmooth **CM**.

Fact: The only traditional optimization technique capable to solve problems with convex structure on **LMO** represented domains is the classical **C***onditional* **G***radient* algorithm (Frank & Wolfe, 1958).

But: The *only* problem with convex structure in the scope of **CG** is *smooth* **CM**. There are *no* traditional LMO-based **FOM**s for (smooth and nonsmooth alike) **SP** and **VI** and *nonsmooth* **CM** problems.

Goal of the talk: **LMO** based **FOM**s for **SP**/**VI**/nonsmooth **CM**.

Fact: The only traditional optimization technique capable to solve problems with convex structure on **LMO** represented domains is the classical **C***onditional* **G***radient* algorithm (Frank & Wolfe, 1958).

But: The *only* problem with convex structure in the scope of **CG** is *smooth* **CM**. There are *no* traditional LMO-based **FOM**s for (smooth and nonsmooth alike) **SP** and **VI** and *nonsmooth* **CM** problems.

Goal of the talk: **LMO** based **FOM**s for **SP**/**VI**/nonsmooth **CM**.

Strategy:

1. Use *Fenchel-type representation* of the monotone vector field associated with the problem of interest to build the *dual* problem, which again will be a problem with convex structure.
2. Solve the dual problem to a desired accuracy ε by a **FOM** (e.g., a proximal one, provided the domain of the dual is proximal friendly).
3. Utilize *accuracy certificates* to build an ε-solution to the problem of interest from information acquired when processing the dual problem.

Strategy:

1. Use *Fenchel-type representation* of the monotone vector field associated with the problem of interest to build the *dual* problem, which again will be a problem with convex structure.
2. Solve the dual problem to a desired accuracy ε by a **FOM** (e.g., a proximal one, provided the domain of the dual is proximal friendly).
3. Utilize *accuracy certificates* to build an ε-solution to the problem of interest from information acquired when processing the dual problem.

Strategy:

1. Use *Fenchel-type representation* of the monotone vector field associated with the problem of interest to build the *dual* problem, which again will be a problem with convex structure.
2. Solve the dual problem to a desired accuracy ε by a **FOM** (e.g., a proximal one, provided the domain of the dual is proximal friendly).
3. Utilize *accuracy certificates* to build an ε-solution to the problem of interest from information acquired when processing the dual problem.

Fenchel representation of a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is
$$f(x) = \sup_y \{\langle x, y\rangle - f^*(y)\}$$
• f^*: convex, proper, l.sc.

F.r. of a convex proper l.sc. f exists and is unique.

However: Fenchel representation "exists in the nature," but, aside of a fistful of simple cases, is not available for numerical use.

Fenchel-type representation of a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is
$$f(x) = \sup_{y\in Y} \{\langle x, Ay + a\rangle - \phi(y)\}$$
• Y: convex • $\phi : Y \to \mathbb{R}$: convex.

• *bilinear* **F.-t.r.**: Y is closed & ϕ is affine.

Fact: **F.-t.r.**'s (even bilinear ones) admit *fully algorithmic* calculus: as applied to **F.-t.** represented convex operands, basic convexity preserving operations straightforwardly yield an **F.-t.r.** of the result.

$\Rightarrow$ *Explicit* **F.-t.r.***'s of convex functions are "common commodity."*

Fenchel representation of a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is

$$f(x) = \sup_y \{\langle x, y\rangle - f^*(y)\}$$

• f^*: convex, proper, l.sc.

F.r. of a convex proper l.sc. f exists and is unique.

However: Fenchel representation "exists in the nature," but, aside of a fistful of simple cases, is not available for numerical use.

Fenchel-type representation of a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is

$$f(x) = \sup_{y \in Y} \{\langle x, Ay + a\rangle - \phi(y)\}$$

• Y: convex • $\phi : Y \to \mathbb{R}$: convex.

• *bilinear* **F.-t.r.**: Y is closed & ϕ is affine.

Fact: **F.-t.r.**'s (even bilinear ones) admit *fully algorithmic* calculus: as applied to **F.-t.** represented convex operands, basic convexity preserving operations straightforwardly yield an **F.-t.r.** of the result.

$\Rightarrow$ *Explicit* **F.-t.r.** *'s of convex functions are "common commodity."*

Fenchel representation of a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is

$$f(x) = \sup_y \{\langle x, y\rangle - f^*(y)\}$$

• f^*: convex, proper, l.sc.

F.r. of a convex proper l.sc. f exists and is unique.

However: Fenchel representation "exists in the nature," but, aside of a fistful of simple cases, is not available for numerical use.

Fenchel-type representation of a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is

$$f(x) = \sup_{y \in Y} \{\langle x, Ay + a\rangle - \phi(y)\}$$

• Y: convex • $\phi : Y \to \mathbb{R}$: convex.

• *bilinear* **F.-t.r.**: Y is closed & ϕ is affine.

Fact: **F.-t.r.**'s (even bilinear ones) admit *fully algorithmic* calculus: as applied to **F.-t.** represented convex operands, basic convexity preserving operations straightforwardly yield an **F.-t.r.** of the result.

$\Rightarrow$ *Explicit* **F.-t.r.** *'s of convex functions are "common commodity."*

Fenchel representation of a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is

$$f(x) = \sup_y \{\langle x, y\rangle - f^*(y)\}$$

• f^*: convex, proper, l.sc.

F.r. of a convex proper l.sc. f exists and is unique.

However: Fenchel representation "exists in the nature," but, aside of a fistful of simple cases, is not available for numerical use.

Fenchel-type representation of a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is

$$f(x) = \sup_{y \in Y} \{\langle x, Ay + a\rangle - \phi(y)\}$$

• Y: convex • $\phi : Y \to \mathbb{R}$: convex.

• *bilinear* **F.-t.r.**: Y is closed & ϕ is affine.

Fact: **F.-t.r.**'s (even bilinear ones) admit *fully algorithmic* calculus: as applied to **F.-t.** represented convex operands, basic convexity preserving operations straightforwardly yield an **F.-t.r.** of the result.

$\Rightarrow$ *Explicit* **F.-t.r.***'s of convex functions are "common commodity."*

Fenchel representation of a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is

$$f(x) = \sup_y \{\langle x, y\rangle - f^*(y)\}$$

• f^*: convex, proper, l.sc.

F.r. of a convex proper l.sc. f exists and is unique.

However: Fenchel representation "exists in the nature," but, aside of a fistful of simple cases, is not available for numerical use.

Fenchel-type representation of a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ is

$$f(x) = \sup_{y \in Y} \{\langle x, Ay + a\rangle - \phi(y)\}$$

• Y: convex • $\phi : Y \to \mathbb{R}$: convex.

• *bilinear* **F.-t.r.**: Y is closed & ϕ is affine.

Fact: **F.-t.r.**'s (even bilinear ones) admit *fully algorithmic* calculus: as applied to **F.-t.** represented convex operands, basic convexity preserving operations straightforwardly yield an **F.-t.r.** of the result.

$\Rightarrow$ *Explicit* **F.-t.r.***'s of convex functions are "common commodity."*

Example: F.r. of the sum $f_1(x) + f_2(x)$ of two functions given by explicit **F.r.'s** is given by computationally demanding inf-convolution:

$$(f_1 + f_2)^*(y) = \inf_{y_1+y_2=y}[f_1^*(y_1) + f_2^*(y_2)]$$

F.-t.r. of the sum $\sum_i f_i$ of functions given by **F.-t.r.**'s is immediate:

$$\alpha_i \geq 0,\ f_i(x) = \sup_{y_i \in Y_i}[\langle x, A_i y_i + a_i\rangle - \phi_i(y_i)) \Rightarrow$$

$$\sum_i \alpha_i f_i(x) = \sup_{y:=[y_1;...;y_k]\in Y_1\times...\times Y_k}\left[\langle x, \underbrace{\textstyle\sum_i \alpha_i[A_i y_i + a_i]}_{Ay+a}\rangle - \underbrace{\textstyle\sum_i \alpha_i \phi_i(y)}_{\phi(y)}\right]$$

Example: $f(x) := \|\mathcal{A}x - b\| = \max_{y\in Y}\{\langle x, \mathcal{A}^* y\rangle - \langle b, y\rangle\}$

- Y: unit ball of the norm $\|\cdot\|_*$ conjugate to $\|\cdot\|$.

Fact: *Strictly feasible conic representation of a function f:*

$$t \geq f(x) \Leftrightarrow \exists u : \mathcal{A}(x, t, u) \in \mathbf{K}$$

[• $\mathcal{A}(x, t, u)$: affine • $\mathbf{K}$: product of second order and semidefinite cones]

can be straightforwardly converted into an explicit **F.-t.r.** *of f.*

Example: **F.r.** of the sum $f_1(x) + f_2(x)$ of two functions given by explicit **F.r.'s** is given by computationally demanding inf-convolution:

$$(f_1 + f_2)^*(y) = \inf_{y_1+y_2=y}[f_1^*(y_1) + f_2^*(y_2)]$$

F.-t.r.of the sum $\sum_i f_i$ of functions given by **F.-t.r.**'s is immediate:

$$\alpha_i \geq 0,\ f_i(x) = \sup_{y_i \in Y_i}[\langle x, A_i y_i + a_i\rangle - \phi_i(y_i) \Rightarrow$$

$$\sum_i \alpha_i f_i(x) = \sup_{y:=[y_1;...;y_k]\in Y_1\times...\times Y_k} \left[\langle x, \underbrace{\sum_i \alpha_i [A_i y_i + a_i]\rangle}_{Ay+a}\rangle - \underbrace{\sum_i \alpha_i \phi_i(y)}_{\phi(y)}\right]$$

Example: $f(x) := \|\mathcal{A}x - b\| = \max_{y\in Y}\{\langle x, \mathcal{A}^* y\rangle - \langle b, y\rangle\}$

- Y: unit ball of the norm $\|\cdot\|_*$ conjugate to $\|\cdot\|$.

Fact: *Strictly feasible conic representation of a function f:*

$$t \geq f(x) \Leftrightarrow \exists u : \mathcal{A}(x,t,u) \in \mathbf{K}$$

[• $\mathcal{A}(x,t,u)$: affine • $\mathbf{K}$: product of second order and semidefinite cones]

can be straightforwardly converted into an explicit **F.-t.r.** *of f.*

Example: F.r. of the sum $f_1(x) + f_2(x)$ of two functions given by explicit **F.r.'s** is given by computationally demanding inf-convolution:

$$(f_1 + f_2)^*(y) = \inf_{y_1+y_2=y}[f_1^*(y_1) + f_2^*(y_2)]$$

F.-t.r. of the sum $\sum_i f_i$ of functions given by **F.-t.r.**'s is immediate:

$$\alpha_i \geq 0,\ f_i(x) = \sup_{y_i \in Y_i}[\langle x, A_i y_i + a_i\rangle - \phi_i(y_i)) \Rightarrow$$

$$\sum_i \alpha_i f_i(x) = \sup_{y:=[y_1;\ldots;y_k]\in Y_1\times\ldots\times Y_k} \left[\langle x, \underbrace{\sum_i \alpha_i [A_i y_i + a_i]\rangle}_{Ay+a}\rangle - \underbrace{\sum_i \alpha_i \phi_i(y)}_{\phi(y)}\right]$$

Example: $f(x) := \|\mathcal{A}x - b\| = \max_{y\in Y}\{\langle x, \mathcal{A}^* y\rangle - \langle b, y\rangle\}$

- Y: unit ball of the norm $\|\cdot\|_*$ conjugate to $\|\cdot\|$.

Fact: *Strictly feasible conic representation of a function f:*

$$t \geq f(x) \Leftrightarrow \exists u : \mathcal{A}(x, t, u) \in \mathbf{K}$$

[• $\mathcal{A}(x, t, u)$: affine • $\mathbf{K}$: product of second order and semidefinite cones]

can be straightforwardly converted into an explicit **F.-t.r.** *of f.*

Example: **F.r.** of the sum $f_1(x)+f_2(x)$ of two functions given by explicit **F.r.'s** is given by computationally demanding inf-convolution:

$$(f_1+f_2)^*(y)=\inf_{y_1+y_2=y}[f_1^*(y_1)+f_2^*(y_2)]$$

F.-t.r.of the sum $\sum_i f_i$ of functions given by **F.-t.r.**'s is immediate:

$$\alpha_i\geq 0,\ f_i(x)=\sup_{y_i\in Y_i}[\langle x,A_iy_i+a_i\rangle-\phi_i(y_i)\rangle\Rightarrow$$

$$\sum_i\alpha_if_i(x)=\sup_{y:=[y_1;...;y_k]\in Y_1\times...\times Y_k}\left[\langle x,\underbrace{\sum_i\alpha_i[A_iy_i+a_i]\rangle}_{Ay+a}\rangle-\underbrace{\sum_i\alpha_i\phi_i(y)}_{\phi(y)}\right]$$

Example: $f(x):=\|\mathcal{A}x-b\|=\max_{y\in Y}\{\langle x,\mathcal{A}^*y\rangle-\langle b,y\rangle\}$

- Y: unit ball of the norm $\|\cdot\|_*$ conjugate to $\|\cdot\|$.

Fact: *Strictly feasible conic representation of a function f:*

$$t\geq f(x)\Leftrightarrow\exists u:\mathcal{A}(x,t,u)\in\mathbf{K}$$

[• $\mathcal{A}(x,t,u)$: affine • $\mathbf{K}$: product of second order and semidefinite cones]

can be straightforwardly converted into an explicit **F.-t.r.** *of f.*

$$\mathrm{Opt}(P) = \min_{x\in X} f(x) \qquad \text{(P)}$$

• X: convex compact • f: convex & Lipschitz

Given an **F.-t.r.** of f:

$$f(x) = \max_{y\in Y} [\langle x, Ay + a\rangle - \phi(y)]$$

with compact Y, we define the problem dual to (P) as

$$\mathrm{Opt}(D) = \min_{y\in Y} [f^*(y) := \phi(y) - \min_{x\in X}\langle x, Ay + a\rangle] \qquad \text{(D)}$$

$$[\mathrm{Opt}(P) = -\mathrm{Opt}(D)]$$

Note: **LMO** for X and **F**irst **O**rder **O**racle for ϕ induce an **FOO** for f^*:

$$y \in Y, \phi'(y) \in \partial_Y \phi(y), x(y) \in \mathrm{Argmin}_X \langle x, Ay + a\rangle \Rightarrow \phi'(y) - A^*x(y) \in \partial_Y f^*(y)$$

Fact: *Assuming* **FOO** *for ϕ available and Y proximal friendly, we can solve* (D) *to a desired accuracy ε by proximal* **FOM** *producing "accuracy certificates" and use these certificates to get an ε-solution to* (P).

$$\mathrm{Opt}(P) = \min_{x\in X} f(x) \qquad (P)$$

• X: convex compact • f: convex & Lipschitz

Given an **F.-t.r.** of f:

$$f(x) = \max_{y\in Y} [\langle x, Ay + a\rangle - \phi(y)]$$

with compact Y, we define the problem dual to (P) as

$$\mathrm{Opt}(D) = \min_{y\in Y} [f^*(y) := \phi(y) - \min_{x\in X}\langle x, Ay + a\rangle] \qquad (D)$$

$$[\mathrm{Opt}(P) = -\mathrm{Opt}(D)]$$

Note: LMO for X and **F**irst **O**rder **O**racle for ϕ induce an **FOO** for f^*:

$$y \in Y, \phi'(y) \in \partial_Y \phi(y), x(y) \in \mathrm{Argmin}_X \langle x, Ay + a\rangle \Rightarrow \phi'(y) - A^* x(y) \in \partial_Y f^*(y)$$

Fact: *Assuming* **FOO** *for ϕ available and Y proximal friendly, we can solve (D) to a desired accuracy ε by proximal* **FOM** *producing "accuracy certificates" and use these certificates to get an ε-solution to (P).*

$$\mathrm{Opt}(P) = \min_{x\in X} f(x) \qquad \text{(P)}$$

• X: convex compact • f: convex & Lipschitz

Given an **F.-t.r.** of f:

$$f(x) = \max_{y\in Y} [\langle x, Ay + a\rangle - \phi(y)]$$

with compact Y, we define the problem dual to (P) as

$$\mathrm{Opt}(D) = \min_{y\in Y} [f^*(y) := \phi(y) - \min_{x\in X}\langle x, Ay + a\rangle] \qquad \text{(D)}$$

$$[\mathrm{Opt}(P) = -\mathrm{Opt}(D)]$$

Note: **LMO** for X and **F**irst **O**rder **O**racle for ϕ induce an **FOO** for f^*:

$$y \in Y, \phi'(y) \in \partial_Y \phi(y), x(y) \in \mathrm{Argmin}_X \langle x, Ay + a\rangle \Rightarrow \phi'(y) - A^* x(y) \in \partial_Y f^*(y)$$

Fact: *Assuming* **FOO** *for* ϕ *available and* Y *proximal friendly, we can solve* (D) *to a desired accuracy* ε *by proximal* **FOM** *producing "accuracy certificates" and use these certificates to get an* ε*-solution to* (P).

Situation:

- We are processing vector field $F(y) : Y \to F$ on convex compact set Y in Euclidean space H.
- after t of the process, we have built a sequence of *search points* $y_i \in Y$ and have computed $F(y_i)$, $1 \le i \le t$.

Definition: An *accuracy certificate* for *t-step execution protocol* $\mathcal{I}_t = \{y_i, F(y_i) : 1 \le i \le t\}$ is a vector $\lambda \in \mathbb{R}^t_+$ with $\sum_{i=1}^t \lambda_i = 1$.

The *residual* and the *approximate solution* associated with $(\mathcal{I}_t, \lambda)$ are

$$\begin{array}{rcll} \mathrm{Res}(\mathcal{I}_t, \lambda | Y) & = & \max_{y \in Y} \sum_{i=1}^t \lambda_i \langle F(y_i), y_i - y \rangle & \text{[residual]} \\ y^t & := & y^t(\mathcal{I}_t, \lambda) = \sum_{i=1}^t \lambda_i y_i & \text{[approximate solution]} \end{array}$$

Note: $y^t \in Y$.

Theorem A. *Let $F : Y \to H$ be vector field induced by* **CM/SP/VI** *problem with convex structure and problem's domain $Y \subset H$. Inaccuracy of the approximate solution y^t associated with t-step execution protocol $\mathcal{I}_t = \{y_i, F(y_i) : 1 \le i \le t\}$ and accuracy certificate λ is upper-bounded by the residual.*

Situation:

• We are processing vector field $F(y) : Y \to F$ on convex compact set Y in Euclidean space H.

• after t of the process, we have built a sequence of *search points* $y_i \in Y$ and have computed $F(y_i)$, $1 \le i \le t$.

Definition: An *accuracy certificate* for *t-step execution protocol* $\mathcal{I}_t = \{y_i, F(y_i) : 1 \le i \le t\}$ is a vector $\lambda \in \mathbb{R}^t_+$ with $\sum_{i=1}^t \lambda_i = 1$.

The *residual* and the *approximate solution* associated with $(\mathcal{I}_t, \lambda)$ are

$$\begin{array}{rcll} \mathrm{Res}(\mathcal{I}_t, \lambda | Y) & = & \max_{y \in Y} \sum_{i=1}^t \lambda_i \langle F(y_i), y_i - y \rangle & \text{[residual]} \\ y^t & := & y^t(\mathcal{I}_t, \lambda) = \sum_{i=1}^t \lambda_i y_i & \text{[approximate solution]} \end{array}$$

Note: $y^t \in Y$.

Theorem A. *Let $F : Y \to H$ be vector field induced by* **CM/SP/VI** *problem with convex structure and problem's domain $Y \subset H$. Inaccuracy of the approximate solution y^t associated with t-step execution protocol $\mathcal{I}_t = \{y_i, F(y_i) : 1 \le i \le t\}$ and accuracy certificate λ is upper-bounded by the residual.*

Situation:

• We are processing vector field $F(y) : Y \to F$ on convex compact set Y in Euclidean space H.

• after t of the process, we have built a sequence of *search points* $y_i \in Y$ and have computed $F(y_i)$, $1 \leq i \leq t$.

Definition: An *accuracy certificate* for *t-step execution protocol* $\mathcal{I}_t = \{y_i, F(y_i) : 1 \leq i \leq t\}$ is a vector $\lambda \in \mathbb{R}^t_+$ with $\sum_{i=1}^t \lambda_i = 1$.

The *residual* and the *approximate solution* associated with $(\mathcal{I}_t, \lambda)$ are

$$\begin{array}{rcll} \mathrm{Res}(\mathcal{I}_t, \lambda | Y) & = & \max_{y \in Y} \sum_{i=1}^t \lambda_i \langle F(y_i), y_i - y \rangle & \text{[residual]} \\ y^t & := & y^t(\mathcal{I}_t, \lambda) = \sum_{i=1}^t \lambda_i y_i & \text{[approximate solution]} \end{array}$$

Note: $y^t \in Y$.

Theorem A. *Let $F : Y \to H$ be vector field induced by* **CM/SP/VI** *problem with convex structure and problem's domain $Y \subset H$. Inaccuracy of the approximate solution y^t associated with t-step execution protocol $\mathcal{I}_t = \{y_i, F(y_i) : 1 \leq i \leq t\}$ and accuracy certificate λ is upper-bounded by the residual.*

Situation:

• We are processing vector field $F(y) : Y \to F$ on convex compact set Y in Euclidean space H.

• after t of the process, we have built a sequence of *search points* $y_i \in Y$ and have computed $F(y_i)$, $1 \leq i \leq t$.

Definition: An *accuracy certificate* for *t-step execution protocol* $\mathcal{I}_t = \{y_i, F(y_i) : 1 \leq i \leq t\}$ is a vector $\lambda \in \mathbb{R}^t_+$ with $\sum_{i=1}^t \lambda_i = 1$.

The *residual* and the *approximate solution* associated with $(\mathcal{I}_t, \lambda)$ are

$$\begin{array}{rcll} \mathrm{Res}(\mathcal{I}_t, \lambda | Y) & = & \max_{y \in Y} \sum_{i=1}^t \lambda_i \langle F(y_i), y_i - y \rangle & \text{[residual]} \\ y^t & := & y^t(\mathcal{I}_t, \lambda) = \sum_{i=1}^t \lambda_i y_i & \text{[approximate solution]} \end{array}$$

Note: $y^t \in Y$.

Theorem A. *Let $F : Y \to H$ be vector field induced by* **CM/SP/VI** *problem with convex structure and problem's domain $Y \subset H$. Inaccuracy of the approximate solution y^t associated with t-step execution protocol $\mathcal{I}_t = \{y_i, F(y_i) : 1 \leq i \leq t\}$ and accuracy certificate λ is upper-bounded by the residual.*

Situation:

• We are processing vector field $F(y): Y \to F$ on convex compact set Y in Euclidean space H.

• after t of the process, we have built a sequence of *search points* $y_i \in Y$ and have computed $F(y_i)$, $1 \le i \le t$.

Definition: An *accuracy certificate* for *t-step execution protocol* $\mathcal{I}_t = \{y_i, F(y_i) : 1 \le i \le t\}$ is a vector $\lambda \in \mathbb{R}^t_+$ with $\sum_{i=1}^t \lambda_i = 1$.

The *residual* and the *approximate solution* associated with $(\mathcal{I}_t, \lambda)$ are

$$\begin{array}{rcll} \mathrm{Res}(\mathcal{I}_t, \lambda | Y) & = & \max_{y \in Y} \sum_{i=1}^t \lambda_i \langle F(y_i), y_i - y \rangle & \text{[residual]} \\ y^t & := & y^t(\mathcal{I}_t, \lambda) = \sum_{i=1}^t \lambda_i y_i & \text{[approximate solution]} \end{array}$$

Note: $y^t \in Y$.

Theorem A. *Let $F: Y \to H$ be vector field induced by* **CM/SP/VI** *problem with convex structure and problem's domain $Y \subset H$. Inaccuracy of the approximate solution y^t associated with t-step execution protocol $\mathcal{I}_t = \{y_i, F(y_i) : 1 \le i \le t\}$ and accuracy certificate λ is upper-bounded by the residual.*

Theorem A [Nem.,Onn,Rothblum,'10] *Let $F : Y \to H$ be vector field induced by* **CM/SP/VI** *problem with domain $Y \subset H$. Inaccuracy of the approximate solution associated with protocol $\mathcal{I}_t = \{y_i, F(y_i) : 1 \leq i \leq t\}$ and certificate λ is upper-bounded by the residual:*

- *for* **CM** *problem* $\min_Y g(y)$*:*

$$\epsilon_{\rm CM}(y^t|g, Y) := g(y^t) - \min_Y g \leq {\rm Res}(\mathcal{I}_t, \lambda|Y)$$

  - $\epsilon_{\rm CM}$ is the standard *optimality gap*

Theorem A [Nem.,Onn,Rothblum,'10] *Let $F : Y \to H$ be vector field induced by* **CM/SP/VI** *problem with domain $Y \subset H$. Inaccuracy of the approximate solution associated with protocol $\mathcal{I}_t = \{y_i, F(y_i) : 1 \leq i \leq t\}$ and certificate λ is upper-bounded by the residual:*

-
- *for* **SP** *problem* $\min_{u \in U} \max_{v \in V} g(u, v)$:
$$\epsilon_{\rm SP}(y^t = [u^t; v^t] | g, U, V) := \max_{v \in V} g(u^t, v) - \min_{u \in U} g(u, v^t) \leq \mathrm{Res}(\mathcal{I}_t, \lambda | Y)$$
  - $\epsilon_{\rm SP}([\bar{u}; \bar{v}] | g, U, V)$ is the *duality gap* – the sum of non-optimalities of $\bar{u}$ and $\bar{v}$ as approximate solutions to the associated with **SP** primal and dual optimization problems
$$\begin{array}{rcl} \mathrm{Opt(Pr)} & = & \min_{u \in U} [\overline{g}(u) := \max_{v \in V} g(u, v)] \\ \mathrm{Opt(Dl)} & = & \max_{v \in V} \left[\underline{g}(v) := \min_{u \in U} g(u, v)\right] \end{array}$$
-

Theorem A [Nem.,Onn,Rothblum,'10] *Let $F : Y \to H$ be vector field induced by* **CM/SP/VI** *problem with domain $Y \subset H$. Inaccuracy of the approximate solution associated with protocol $\mathcal{I}_t = \{y_i, F(y_i) : 1 \le i \le t\}$ and certificate λ is upper-bounded by the residual:*

-
-
- *for* **VI** *"find $y \in Y : \langle F(z), z - y \rangle \ge 0 \, \forall z \in Y$:"*
  $$\epsilon_{\mathrm{VI}}(y^t | F, Y) := \sup_{z \in Y} \langle F(z), y^t - z \rangle \le \mathrm{Res}(\mathcal{I}_t, \lambda | Y)$$
  - $\epsilon_{\mathrm{VI}}(\cdot|\cdot,\cdot)$ is the *dual gap function* associated with **VI**.

$$\begin{array}{rcll} \mathrm{Opt}(P) & = & \min_{x\in X}\left\{f(x) := \max_{y\in Y}\left[\langle x, Ay+a\rangle - \phi(y)\right]\right\} & \text{(P)} \\ -\mathrm{Opt}(P) = \mathrm{Opt}(D) & = & \min_{y\in Y}\left\{f^*(y) := \phi(y) - \min_{x\in X}\langle x, Ay+a\rangle\right\} & \text{(D)} \end{array}$$

Theorem B [Cox,Juditsky,Nem.,'13] *Let*

- $F(y) = \phi'(y) - A^* x(y), x(y) \in \mathrm{Argmin}_X \langle x, Ay + a\rangle$ *be the monotone operator associated with the dual problem* (D),
- λ *be an accuracy certificate for t-step execution protocol*
$$\{y_i \in Y, F(y_i) = \phi'(y_i) - A^* x(y_i)\}$$
produced by an **FOM** *as applied to the dual problem.*

Setting $x^t = \sum_{i=1}^t \lambda_i x(y_i)$, we get a feasible solution to the primal problem (P) *such that*

$$\begin{array}{rcl} \epsilon_{\rm CM}(x^t|f,X) & \leq & \mathrm{Gap}(\mathcal{I}_t,\lambda|Y) \\ & := & \left[\min_{i\leq t} f^*(y_i) - \sum_{i=1}^t \lambda_i f^*(y_i)\right] + \mathrm{Res}(\mathcal{I}_t,\lambda|Y) \\ & \leq & \mathrm{Res}(\mathcal{I}_t,\lambda|Y) \end{array}$$

$\mathrm{Opt}(P)$	$=$	$\min_{x\in X}\{f(x) := \max_{y\in Y}[\langle x, Ay+a\rangle - \phi(y)]\}$	(P)
$-\mathrm{Opt}(P) = \mathrm{Opt}(D)$	$=$	$\min_{y\in Y}\{f^*(y) := \phi(y) - \min_{x\in X}\langle x, Ay+a\rangle\}$	(D)

Theorem B [Cox,Juditsky,Nem.,'13] *Let*

- $F(y) = \phi'(y) - A^*x(y), x(y) \in \mathrm{Argmin}_X\langle x, Ay+a\rangle$ *be the monotone operator associated with the dual problem* (D),
- λ *be an accuracy certificate for t-step execution protocol*
$$\{y_i \in Y, F(y_i) = \phi'(y_i) - A^*x(y_i)\}$$
produced by an **FOM** *as applied to the dual problem.*

Setting $x^t = \sum_{i=1}^t \lambda_i x(y_i)$, *we get a feasible solution to the primal problem* (P) *such that*

$$\begin{array}{rcl}\epsilon_{\mathrm{CM}}(x^t|f,X) & \leq & \mathrm{Gap}(\mathcal{I}_t,\lambda|Y)\\ & := & \left[\min_{i\leq t} f^*(y_i) - \sum_{i=1}^t \lambda_i f^*(y_i)\right] + \mathrm{Res}(\mathcal{I}_t,\lambda|Y)\\ & \leq & \mathrm{Res}(\mathcal{I}_t,\lambda|Y)\end{array}$$

Fact: Standard FOMs:

- Subradient/Mirror Descent and its Bundle versions,
- Cutting plane algorithms, including Ellipsoid/Inscribed Ellipsoid methods,
- Nesterov's Fast Gradients,...

produce certificates with gap/residual obeying the standard efficiency estimates of the algorithms.

Example: Let F be a vector field on convex domain Y of diameter D w.r.t. a norm $\|\cdot\|$.

- t-step *Mirror Descent* as applied to (F, Y) ensures $\mathrm{Res} \le \mathcal{C}_{\|\cdot\|}\frac{LD}{\sqrt{t}}$, provided $\|F(y)\|_* \le L \forall y \in Y$.
  For good norms (ℓ_p with $1 \le p \le 2$, nuclear,...), $\mathcal{C}_{\|\cdot\|}$ is (nearly) dimension independent.
- t-step *Mirror Prox* as applied to (F, Y) ensures $\mathrm{Res} \le \mathcal{C}^2_{\|\cdot\|}\frac{LD^2}{t}$, provided $\|F(y) - F(z)\|_* \le L\|y - z\| \forall y, z \in Y$

$\Rightarrow$ **F.-t.r.**'s *with proximal friendly and bounded dual domains induce $O(1/\sqrt{t})$-converging* **LMO**-*based algorithms for Nonsmooth* **CM**

Fact: Standard FOMs:

- Subradient/Mirror Descent and its Bundle versions,
- Cutting plane algorithms, including Ellipsoid/Inscribed Ellipsoid methods,
- Nesterov's Fast Gradients,...

produce certificates with gap/residual obeying the standard efficiency estimates of the algorithms.

Example: Let F be a vector field on convex domain Y of diameter D w.r.t. a norm $\|\cdot\|$.

- t-step *Mirror Descent* as applied to (F,Y) ensures $\mathrm{Res} \leq \mathcal{C}_{\|\cdot\|}\frac{LD}{\sqrt{t}}$, provided $\|F(y)\|_* \leq L \forall y \in Y$.
  For good norms (ℓ_p with $1 \leq p \leq 2$, nuclear,...), $\mathcal{C}_{\|\cdot\|}$ is (nearly) dimension independent.

- t-step *Mirror Prox* as applied to (F,Y) ensures $\mathrm{Res} \leq \mathcal{C}^2_{\|\cdot\|}\frac{LD^2}{t}$, provided $\|F(y)-F(z)\|_* \leq L\|y-z\| \forall y,z \in Y$

$\Rightarrow$ **F.-t.r.**'*s with proximal friendly and bounded dual domains induce $O(1/\sqrt{t})$-converging* **LMO**-*based algorithms for Nonsmooth* **CM**

Fact: Standard FOMs:

- Subradient/Mirror Descent and its Bundle versions,
- Cutting plane algorithms, including Ellipsoid/Inscribed Ellipsoid methods,
- Nesterov's Fast Gradients,...

produce certificates with gap/residual obeying the standard efficiency estimates of the algorithms.

Example: Let F be a vector field on convex domain Y of diameter D w.r.t. a norm $\|\cdot\|$.

- t-step *Mirror Descent* as applied to (F, Y) ensures $\mathrm{Res} \leq \mathcal{C}_{\|\cdot\|} \frac{LD}{\sqrt{t}}$, provided $\|F(y)\|_* \leq L \forall y \in Y$.
  For good norms (ℓ_p with $1 \leq p \leq 2$, nuclear,...), $\mathcal{C}_{\|\cdot\|}$ is (nearly) dimension independent.
- t-step *Mirror Prox* as applied to (F, Y) ensures $\mathrm{Res} \leq \mathcal{C}^2_{\|\cdot\|} \frac{LD^2}{t}$, provided $\|F(y) - F(z)\|_* \leq L\|y - z\| \forall y, z \in Y$

$\Rightarrow$ **F.-t.r.**'*s with proximal friendly and bounded dual domains induce* $O(1/\sqrt{t})$*-converging* **LMO**-*based algorithms for Nonsmooth* **CM**

Fact: Standard FOMs:

- Subradient/Mirror Descent and its Bundle versions,
- Cutting plane algorithms, including Ellipsoid/Inscribed Ellipsoid methods,
- Nesterov's Fast Gradients,...

produce certificates with gap/residual obeying the standard efficiency estimates of the algorithms.

Example: Let F be a vector field on convex domain Y of diameter D w.r.t. a norm $\|\cdot\|$.

- t-step *Mirror Descent* as applied to (F, Y) ensures $\mathrm{Res} \leq \mathcal{C}_{\|\cdot\|} \frac{LD}{\sqrt{t}}$, provided $\|F(y)\|_* \leq L \forall y \in Y$.
  For good norms (ℓ_p with $1 \leq p \leq 2$, nuclear,...), $\mathcal{C}_{\|\cdot\|}$ is (nearly) dimension independent.
- t-step *Mirror Prox* as applied to (F, Y) ensures $\mathrm{Res} \leq \mathcal{C}^2_{\|\cdot\|} \frac{LD^2}{t}$, provided $\|F(y) - F(z)\|_* \leq L\|y - z\| \forall y, z \in Y$

$\Rightarrow$ **F.-t.r.***'s with proximal friendly and bounded dual domains induce $O(1/\sqrt{t})$-converging* **LMO***-based algorithms for Nonsmooth* **CM**

Fact: Standard FOMs:

- Subradient/Mirror Descent and its Bundle versions,
- Cutting plane algorithms, including Ellipsoid/Inscribed Ellipsoid methods,
- Nesterov's Fast Gradients,...

produce certificates with gap/residual obeying the standard efficiency estimates of the algorithms.

Example: Let F be a vector field on convex domain Y of diameter D w.r.t. a norm $\|\cdot\|$.

- t-step *Mirror Descent* as applied to (F, Y) ensures $\mathrm{Res} \leq \mathcal{C}_{\|\cdot\|} \frac{LD}{\sqrt{t}}$, provided $\|F(y)\|_* \leq L \forall y \in Y$.
  For good norms (ℓ_p with $1 \leq p \leq 2$, nuclear,...), $\mathcal{C}_{\|\cdot\|}$ is (nearly) dimension independent.

- t-step *Mirror Prox* as applied to (F, Y) ensures $\mathrm{Res} \leq \mathcal{C}^2_{\|\cdot\|} \frac{LD^2}{t}$, provided $\|F(y) - F(z)\|_* \leq L\|y - z\| \,\forall y, z \in Y$

$\Rightarrow$ **F.-t.r.***'s with proximal friendly and bounded dual domains induce $O(1/\sqrt{t})$-converging* **LMO***-based algorithms for Nonsmooth* **CM***.*

Observation: **F.-t.r.** $f(x) = \max_{y\in Y}\{\langle x, Ay + a\rangle - \phi(x)\}$ *of a convex function f with proximal friendly Y and affine ϕ induces easy to compute smooth approximation of f*

$$f_\epsilon(x) = \max_{y\in Y}\left[\langle x, Ay + a\rangle - \phi(y) + \epsilon d(y)\right]$$

$d(\cdot)$: strongly convex function with easy-to-minimize linear perturbations

and one can minimize this approximation over an **LMO**-*represented domain X by Conditional Gradients, resulting in the same $O(1/\sqrt{t})$ convergence as* **F.-t.r.**/*Accuracy Certificates machinery.*

Note: This is nothing but *Nesterov's smoothing* with Conditional Gradients in the role of Fast Gradients.

Observation: **F.-t.r.** $f(x) = \max_{y\in Y}\{\langle x, Ay + a\rangle - \phi(x)\}$ *of a convex function f with* ***proximal friendly*** Y *and* ***affine*** ϕ *induces easy to compute smooth approximation of f*

$$f_\epsilon(x) = \max_{y\in Y}\left[\langle x, Ay + a\rangle - \phi(y) + \epsilon d(y)\right]$$

$d(\cdot)$: strongly convex function with easy-to-minimize linear perturbations

and one can minimize this approximation over an **LMO**-*represented domain X by Conditional Gradients, resulting in the same* $O(1/\sqrt{t})$ *convergence as* **F.-t.r.**/*Accuracy Certificates machinery.*

Note: This is nothing but *Nesterov's smoothing* with Conditional Gradients in the role of Fast Gradients.

Definition [Juditsky,Nem.,'14] Let

- $\Phi(\cdot) : X \to E$ be a monotone operator on convex domain X in Euclidean space E
- H be Euclidean space and $y \mapsto Ay + a : H \to E$ be an affine mapping
- $G(\cdot) : Y \to H$ be monotone operator on convex domain $Y \subset H$, and for all $x \in X$ the **VI** "find $y \in Y : \langle G(z) - A^* x, z - y\rangle \geq 0 \ \forall z \in Y$" has a *strong* solution $y(x) \in Y$: $\langle G(y(x)) - A^* x, z - y(x)\rangle \geq 0 \, \forall z \in Y$.
- It holds $\Phi(x) = Ay(x) + a \ \forall x \in X$

We say that

- $H, A, a, Y, G(\cdot), y(\cdot)$ form a Fenchel-type representation of Φ on X. **Informally: F.-t.r.** of Φ is
$$\Phi(x) = AG^{-1}(A^* x) + a$$
- the **VI** "find $y \in Y : \langle \Psi(v), v - y\rangle \geq 0 \, \forall v \in Y$" with
$$\Psi(v) = G(v) - A^* x(v), \;\; x(v) \in \mathrm{Argmin}_{x \in X} \langle x, Av + a\rangle$$
is the *dual* of the *primal* **VI** "find $x \in X : \langle \Phi(u), u - x\rangle \geq 0 \ \forall u \in X$"

Note: $\Psi(\cdot)$ is monotone on Y, and computing $\Psi(\cdot)$ reduces to computing $G(\cdot)$ and a call to **LMO** for X.

Definition [Juditsky,Nem.,'14] Let

- $\Phi(\cdot) : X \to E$ be a monotone operator on convex domain X in Euclidean space E
- H be Euclidean space and $y \mapsto Ay + a : H \to E$ be an affine mapping
- $G(\cdot) : Y \to H$ be monotone operator on convex domain $Y \subset H$, and for all $x \in X$ the **VI** "find $y \in Y : \langle G(z) - A^* x, z - y\rangle \geq 0 \; \forall z \in Y$" has a *strong* solution $y(x) \in Y$: $\langle G(y(x)) - A^* x, z - y(x)\rangle \geq 0 \, \forall z \in Y$.
- It holds $\Phi(x) = Ay(x) + a \; \forall x \in X$

We say that

- $H, A, a, Y, G(\cdot), y(\cdot)$ form a Fenchel-type representation of Φ on X. **Informally: F.-t.r.** of Φ is
$$\Phi(x) = AG^{-1}(A^* x) + a$$
- the **VI** "find $y \in Y : \langle \Psi(v), v - y\rangle \geq 0 \, \forall v \in Y$" with
$$\Psi(v) = G(v) - A^* x(v), \;\; x(v) \in \mathrm{Argmin}_{x \in X} \langle x, Av + a\rangle$$
is the *dual* of the *primal* **VI** "find $x \in X : \langle \Phi(u), u - x\rangle \geq 0 \; \forall u \in X$"

Note: $\Psi(\cdot)$ is monotone on Y, and computing $\Psi(\cdot)$ reduces to computing $G(\cdot)$ and a call to **LMO** for X.

Definition [Juditsky,Nem.,'14] Let

- $\Phi(\cdot) : X \to E$ be a monotone operator on convex domain X in Euclidean space E
- H be Euclidean space and $y \mapsto Ay + a : H \to E$ be an affine mapping
- $G(\cdot) : Y \to H$ be monotone operator on convex domain $Y \subset H$, and for all $x \in X$ the **VI** "find $y \in Y : \langle G(z) - A^* x, z - y\rangle \geq 0 \; \forall z \in Y$" has a *strong* solution $y(x) \in Y$: $\langle G(y(x)) - A^* x, z - y(x)\rangle \geq 0 \, \forall z \in Y$.
- It holds $\Phi(x) = Ay(x) + a \; \forall x \in X$

We say that

- $H, A, a, Y, G(\cdot), y(\cdot)$ form a Fenchel-type representation of Φ on X.
  Informally: F.-t.r. of Φ is
  $$\Phi(x) = AG^{-1}(A^* x) + a$$
- the **VI** "find $y \in Y : \langle \Psi(v), v - y\rangle \geq 0 \, \forall v \in Y$" with
  $$\Psi(v) = G(v) - A^* x(v), \;\; x(v) \in \mathrm{Argmin}_{x \in X}\langle x, Av + a\rangle$$
  is the *dual* of the *primal* **VI** "find $x \in X : \langle \Phi(u), u - x\rangle \geq 0 \; \forall u \in X$"

Note: $\Psi(\cdot)$ is monotone on Y, and computing $\Psi(\cdot)$ reduces to computing $G(\cdot)$ and a call to **LMO** for X.

Definition [Juditsky,Nem.,'14] Let

- $\Phi(\cdot) : X \to E$ be a monotone operator on convex domain X in Euclidean space E
- H be Euclidean space and $y \mapsto Ay + a : H \to E$ be an affine mapping
- $G(\cdot) : Y \to H$ be monotone operator on convex domain $Y \subset H$, and for all $x \in X$ the **VI** "find $y \in Y : \langle G(z) - A^*x, z - y\rangle \geq 0 \ \forall z \in Y$" has a *strong* solution $y(x) \in Y$: $\langle G(y(x)) - A^*x, z - y(x)\rangle \geq 0 \, \forall z \in Y$.
- It holds $\Phi(x) = Ay(x) + a \ \forall x \in X$

We say that

- $H, A, a, Y, G(\cdot), y(\cdot)$ form a Fenchel-type representation of Φ on X.
  Informally: F.-t.r. of Φ is
  $$\Phi(x) = AG^{-1}(A^*x) + a$$
- the **VI** "find $y \in Y : \langle \Psi(v), v - y\rangle \geq 0 \, \forall v \in Y$" with
  $$\Psi(v) = G(v) - A^*x(v), \;\; x(v) \in \mathrm{Argmin}_{x \in X}\langle x, Av + a\rangle$$
  is the *dual* of the *primal* **VI** "find $x \in X : \langle \Phi(u), u - x\rangle \geq 0 \ \forall u \in X$"

Note: $\Psi(\cdot)$ is monotone on Y, and computing $\Psi(\cdot)$ reduces to computing $G(\cdot)$ and a call to **LMO** for X.

Example: Affine monotone operator $\Phi(x) = Sx + s : E \to E$ on a convex compact domain $X \subset E$ admits **F.-t.r.** with the data

$$H = E, Ay + a \equiv Sy + s, Y \supset X, G(y) = S^* y, y(x) = x$$

Fact:

- **F.-t.r.**'s of monotone operators admit *fully algorithmic* calculus: as applied to **F.-t.** represented monotone operands, basic monotonicity-preserving operations with vector fields:
  - summation with nonnegative weights,
  - direct summation,
  - affine substitution of argument,

  straightforwardly yield an **F.-t.r.** of the result.

$\Rightarrow$ **F.-t.r.** *'s of monotone operators form a "common commodity."*

Example: Affine monotone operator $\Phi(x) = Sx + s : E \to E$ on a convex compact domain $X \subset E$ admits **F.-t.r.** with the data

$$H = E, Ay + a \equiv Sy + s, Y \supset X, G(y) = S^* y, y(x) = x$$

Fact:

- **F.-t.r.**'s of monotone operators admit *fully algorithmic* calculus: as applied to **F.-t.** represented monotone operands, basic monotonicity-preserving operations with vector fields:
  - summation with nonnegative weights,
  - direct summation,
  - affine substitution of argument,

  straightforwardly yield an **F.-t.r.** of the result.

$\Rightarrow$ **F.-t.r.**'s *of monotone operators form a "common commodity."*

Example: Affine monotone operator $\Phi(x) = Sx + s : E \to E$ on a convex compact domain $X \subset E$ admits **F.-t.r.** with the data

$$H = E, Ay + a \equiv Sy + s, Y \supset X, G(y) = S^* y, y(x) = x$$

Fact:

- **F.-t.r.**'s of monotone operators admit *fully algorithmic* calculus: as applied to **F.-t.** represented monotone operands, basic monotonicity-preserving operations with vector fields:
  - summation with nonnegative weights,
  - direct summation,
  - affine substitution of argument,

  straightforwardly yield an **F.-t.r.** of the result.

$\Rightarrow$ **F.-t.r.***'s of monotone operators form a "common commodity."*

$$\begin{array}{lr} \text{find } x \in X : \langle \Phi(u), u - x \rangle \geq 0 \;\; \forall u \in X & \text{(P)} \\ \Phi(x) = Ay(x) + a \;\&\; \langle G(y(x)) - A^*x, v - y(x) \rangle \geq 0 \; \forall v \in Y & (1) \\ \Psi(y) := G(y) - A^*x(y) \;\&\; x(y) \in \mathrm{Argmin}_{x \in X} \langle x, Ay + a \rangle & (2) \end{array}$$

$$\Rightarrow \text{find } y \in Y : \langle \Psi(v), v - y \rangle \geq 0 \;\; \forall v \in Y \qquad \text{(D)}$$

Theorem C [Juditsky,Nem.,'14] *Let*

- $\Phi(\cdot) : X \to E$ *be a monotone vector field with* **F.-t.r.** (1) *on convex compact subset* X *of Euclidean space* E*, and*
- λ *be an accuracy certificate for* t*-step execution protocol*
$$\mathcal{I}_t = \{y_i, \Psi(y_i) : 1 \leq i \leq t\}$$
produced by an **FOM** *as applied to the* *dual* **VI** (D).

The point $x^t = \sum_{i=1}^t \lambda_i x(y_i)$ *is a feasible solution to the primal* **VI** (P)*, and the inaccuracy of* x^t *w.r.t.* (P) *can be upper-bounded as*

$$\epsilon_{\rm VI}(x^t|F, X) \leq \mathrm{Res}(\mathcal{I}_t, \lambda|\Psi, Y).$$

When Φ *is the affine monotone vector field associated with* **SP** *problem*

$$\min_{u \in U} \max_{v \in V} g(u, v)$$

with bilinear $g(u, v)$ *and convex compact* U, V*, we have also*

$$\epsilon_{\rm SP}(x^t|g, U, V) \leq \mathrm{Res}(\mathcal{I}_t, \lambda|\Psi, Y).$$

find $x \in X : \langle \Phi(u), u - x\rangle \geq 0 \;\; \forall u \in X$	(P)
$\Phi(x) = Ay(x) + a \;\&\; \langle G(y(x)) - A^* x, v - y(x)\rangle \geq 0 \; \forall v \in Y$	(1)
$\Psi(y) := G(y) - A^* x(y) \;\&\; x(y) \in \mathrm{Argmin}_{x \in X} \langle x, Ay + a\rangle$	(2)
$\Rightarrow$ find $y \in Y : \langle \Psi(v), v - y\rangle \geq 0 \;\; \forall v \in Y$	(D)

Theorem C [Juditsky,Nem.,'14] *Let*

- $\Phi(\cdot) : X \to E$ *be a monotone vector field with* **F.-t.r.** (1) *on convex compact subset* X *of Euclidean space* E*, and*
- λ *be an accuracy certificate for* t*-step execution protocol*
$$\mathcal{I}_t = \{y_i, \Psi(y_i) : 1 \leq i \leq t\}$$
produced by an **FOM** *as applied to the dual* **VI** (D).

The point $x^t = \sum_{i=1}^t \lambda_i x(y_i)$ *is a feasible solution to the primal* **VI** (P)*, and the inaccuracy of* x^t *w.r.t.* (P) *can be upper-bounded as*
$$\epsilon_{\mathrm{VI}}(x^t | F, X) \leq \mathrm{Res}(\mathcal{I}_t, \lambda | \Psi, Y).$$

When Φ *is the affine monotone vector field associated with* **SP** *problem*
$$\min_{u \in U} \max_{v \in V} g(u, v)$$
with bilinear $g(u, v)$ *and convex compact* U, V*, we have also*
$$\epsilon_{\mathrm{SP}}(x^t | g, U, V) \leq \mathrm{Res}(\mathcal{I}_t, \lambda | \Psi, Y).$$

$$\text{find } x \in X : \langle \Phi(u), u - x\rangle \geq 0 \;\; \forall u \in X \qquad \text{(P)}$$

$$\Phi(x) = Ay(x) + a \;\&\; \langle G(y(x)) - A^*x, v - y(x)\rangle \geq 0 \;\forall v \in Y \qquad (1)$$

$$\Psi(y) := G(y) - A^*x(y) \;\&\; x(y) \in \mathrm{Argmin}_{x\in X}\langle x, Ay + a\rangle \qquad (2)$$

$$\Rightarrow \text{find } y \in Y : \langle \Psi(v), v - y\rangle \geq 0 \;\; \forall v \in Y \qquad \text{(D)}$$

Theorem C [Juditsky,Nem.,'14] *Let*

- $\Phi(\cdot) : X \to E$ *be a monotone vector field with* **F.-t.r.** (1) *on convex compact subset* X *of Euclidean space* E, *and*
- λ *be an accuracy certificate for* t*-step execution protocol*
$$\mathcal{I}_t = \{y_i, \Psi(y_i) : 1 \leq i \leq t\}$$
produced by an **FOM** *as applied to the dual* **VI** (D).

The point $x^t = \sum_{i=1}^t \lambda_i x(y_i)$ *is a feasible solution to the primal* **VI** (P), *and the inaccuracy of* x^t *w.r.t.* (P) *can be upper-bounded as*
$$\epsilon_{\mathrm{VI}}(x^t|F, X) \leq \mathrm{Res}(\mathcal{I}_t, \lambda|\Psi, Y).$$
When Φ *is the affine monotone vector field associated with* **SP** *problem*
$$\min_{u\in U}\max_{v\in V} g(u, v)$$
with bilinear $g(u, v)$ *and convex compact* U, V, *we have also*
$$\epsilon_{\mathrm{SP}}(x^t|g, U, V) \leq \mathrm{Res}(\mathcal{I}_t, \lambda|\Psi, Y).$$

$$\boxed{\begin{array}{lr} \text{find } x \in X : \langle \Phi(u), u - x\rangle \geq 0 \;\; \forall u \in X & (\mathrm{P}) \\ \Phi(x) = Ay(x) + a \;\&\; \langle G(y(x)) - A^* x, v - y(x)\rangle \geq 0 \;\forall v \in Y & (1) \\ \Psi(y) := G(y) - A^* x(y) \;\&\; x(y) \in \mathrm{Argmin}_{x\in X}\langle x, Ay + a\rangle & (2) \end{array}}$$

$$\Downarrow$$

$$\boxed{\text{find } y \in Y : \langle \Psi(v), v - y\rangle \geq 0 \;\; \forall v \in Y \qquad (\mathrm{D})}$$

Corollary: *When Y is proximal friendly and $G(\cdot)$ is bounded on Y,* **F.-t.r.** (1) *induces $O(1/\sqrt{t})$-converging* **LMO**-*based algorithm for* (P).

Corollary: **F.-t.r.** (1) *with bounded $Y \subset \mathbb{R}^n$ induces* **LMO**-*based linearly converging algorithm for* (P) *with convergence ratio ϑ depending solely on n* (e.g., $\vartheta = 1 - O(1/n)$ or $\vartheta = 1 - O(1/n^2)$)

Note: In some applications, $\dim X$ is huge and $\dim Y$ is small (tens/hundreds)...

$$\boxed{\begin{array}{lr} \text{find } x \in X : \langle \Phi(u), u - x\rangle \geq 0 \;\; \forall u \in X & (\mathrm{P}) \\ \Phi(x) = Ay(x) + a \;\&\; \langle G(y(x)) - A^* x, v - y(x)\rangle \geq 0 \;\forall v \in Y & (1) \\ \Psi(y) := G(y) - A^* x(y) \;\&\; x(y) \in \mathrm{Argmin}_{x \in X}\langle x, Ay + a\rangle & (2) \end{array}}$$

$$\Downarrow$$

$$\boxed{\text{find } y \in Y : \langle \Psi(v), v - y\rangle \geq 0 \;\; \forall v \in Y \qquad (\mathrm{D})}$$

Corollary: *When Y is proximal friendly and $G(\cdot)$ is bounded on Y,* **F.-t.r.** (1) *induces $O(1/\sqrt{t})$-converging* **LMO**-*based algorithm for* (P).

Corollary: **F.-t.r.** (1) *with bounded $Y \subset \mathbb{R}^n$ induces* **LMO**-*based linearly converging algorithm for* (P) *with convergence ratio ϑ depending solely on n* (e.g., $\vartheta = 1 - O(1/n)$ or $\vartheta = 1 - O(1/n^2)$)

Note: In some applications, $\dim X$ is huge and $\dim Y$ is small (tens/hundreds)...

$$\boxed{\begin{array}{lr}\text{find } x\in X:\langle\Phi(u),u-x\rangle\geq 0\ \forall u\in X & (\mathrm{P})\\ \Phi(x)=Ay(x)+a\ \&\ \langle G(y(x))-A^*x,v-y(x)\rangle\geq 0\ \forall v\in Y & (1)\\ \Psi(y):=G(y)-A^*x(y)\ \&\ x(y)\in\mathrm{Argmin}_{x\in X}\langle x,Ay+a\rangle & (2)\end{array}}$$

$$\Downarrow$$

$$\boxed{\text{find } y\in Y:\langle\Psi(v),v-y\rangle\geq 0\ \forall v\in Y \qquad (\mathrm{D})}$$

Corollary: *When Y is proximal friendly and $G(\cdot)$ is bounded on Y,* **F.-t.r.** (1) *induces $O(1/\sqrt{t})$-converging* **LMO***-based algorithm for* (P).

Corollary: **F.-t.r.** (1) *with bounded $Y\subset\mathbb{R}^n$ induces* **LMO***-based linearly converging algorithm for* (P) *with convergence ratio ϑ depending solely on n* (e.g., $\vartheta=1-O(1/n)$ or $\vartheta=1-O(1/n^2)$)
Note: In some applications, $\dim X$ is huge and $\dim Y$ is small (tens/hundreds)...

Fact: *Nonsmooth* **CM** problems *with smooth* **F.-t.r.***'s* and *smooth* **SP**/**VI**'s on *proximal friendly* domains admit $O(1/t)$-converging *proximal* **FOM**s.

In contrast, *the best known convergence rate of* **LMO***-based* **FOM***s for nonsmooth* **CM** *and smooth/nonsmooth* **SP/VI** *problems is* $O(1/\sqrt{t})$.

Reason: For a typical **LMO**-based **FOM**, *approximate solution generated in t calls to* **LMO** *belongs to the linear span of the minimizers of the t linear forms processed by LMO in these calls. Under this restriction, already pretty simple nonsmooth* **CM***'s and smooth/nonsmooth* **SP***'s/***VI***'s on domains like ℓ_1/nuclear norm balls provably cannot be solved faster than at $O(1/\sqrt{t})$ rate.*

Fact: *Nonsmooth* **CM** problems *with smooth* **F.-t.r.***'s* and *smooth* **SP**/**VI**'s on *proximal friendly* domains admit $O(1/t)$-converging *proximal* **FOM**s.

In contrast, *the best known convergence rate of* **LMO***-based* **FOM***s for nonsmooth* **CM** *and smooth/nonsmooth* **SP/VI** *problems is* $O(1/\sqrt{t})$.

Reason: For a typical **LMO**-based **FOM**, *approximate solution generated in t calls to* **LMO** *belongs to the linear span of the minimizers of the t linear forms processed by LMO in these calls. Under this restriction, already pretty simple nonsmooth* **CM***'s and smooth/nonsmooth* **SP***'s/***VI***'s on domains like ℓ_1/nuclear norm balls provably cannot be solved faster than at $O(1/\sqrt{t})$ rate.*

Fact: *Nonsmooth* **CM** problems *with smooth* **F.-t.r.***'s* and *smooth* **SP**/**VI**'s on *proximal friendly* domains admit $O(1/t)$-converging *proximal* **FOM**s.

In contrast, *the best known convergence rate of* **LMO***-based* **FOM***s for nonsmooth* **CM** *and smooth/nonsmooth* **SP**/**VI** *problems is* $O(1/\sqrt{t})$.

Reason: For a typical **LMO**-based **FOM**, *approximate solution generated in t calls to* **LMO** *belongs to the linear span of the minimizers of the t linear forms processed by LMO in these calls.* *Under this restriction, already pretty simple nonsmooth* **CM***'s and smooth/nonsmooth* **SP***'s/***VI***'s on domains like ℓ_1/nuclear norm balls provably cannot be solved faster than at $O(1/\sqrt{t})$ rate.*

Fact: *Nonsmooth* **CM** problems *with smooth* **F.-t.r.***'s* and *smooth* **SP**/**VI**'s on *proximal friendly* domains admit $O(1/t)$-converging *proximal* **FOM**s.

In contrast, *the best known convergence rate of* **LMO***-based* **FOM***s for nonsmooth* **CM** *and smooth/nonsmooth* **SP/VI** *problems is* $O(1/\sqrt{t})$.

Reason: For a typical **LMO**-based **FOM**, *approximate solution generated in t calls to* **LMO** *belongs to the linear span of the minimizers of the t linear forms processed by LMO in these calls.* *Under this restriction, already pretty simple nonsmooth* **CM***'s and smooth/nonsmooth* **SP***'s/***VI***'s on domains like ℓ_1/nuclear norm balls provably cannot be solved faster than at $O(1/\sqrt{t})$ rate.*

Problem: *Given noisy observations a_{ij}, $(i,j) \in \Omega$, of N entries in a $p \times p$ matrix, find matrix of nuclear norm ≤ 1 which is the best fit to the observations in the uniform norm.*

Note:

- **LMO** for the domain X of the problem of interest (unit Nuclear norm ball in $\mathbb{R}^{p\times p}$) is readily given by Power method.
- The objective admits bilinear **F.-t.r.**, the domain Y of the dual problem is the proximal-friendly unit ℓ_1-ball in $\mathbb{R}^N$.

Worst-case efficiency: $\epsilon_{\rm CM}(x^t) \leq O(1)\frac{\ln(N)[1+\max_{i,j}|a_{ij}|]}{\sqrt{t}}$

Results, I: Restricted Memory Bundle-Level algorithm on the dual of a low size ($p = 512, N = 512$) Matrix Completion:

Memory depth	1	33	65	129
$\mathrm{Gap}_1/\mathrm{Gap}_{1024}$	114	164	350	3253

Results, II: Subgradient Descent on the dual of Matrix Completion:

p	N	Gap_1	$\mathrm{Gap}_1/\mathrm{Gap}_{32}$	$\mathrm{Gap}_1/\mathrm{Gap}_{128}$	$\mathrm{Gap}_1/\mathrm{Gap}_{1024}$	CPU, sec
2048	8192	1.81e-1	171.2	213.8	451.4	521.3
4096	16384	3.74e-1	335.4	1060.8	1287.3	1524.8
8192	16384	2.54e-1	37.8	875.8	1183.6	3644.0

Platform: desktop PC with 4×3.40 GHz *Intel Core2* CPU and 16 GB RAM, Windows 7-64 OS.

Problem: *Given noisy observations a_{ij}, $(i,j) \in \Omega$, of N entries in a $p \times p$ matrix, find matrix of nuclear norm ≤ 1 which is the best fit to the observations in the uniform norm.*

Note:

• **LMO** for the domain X of the problem of interest (unit Nuclear norm ball in $\mathbb{R}^{p\times p}$) is readily given by Power method.

• The objective admits bilinear **F.-t.r.**, the domain Y of the dual problem is the proximal-friendly unit ℓ_1-ball in $\mathbb{R}^N$.

Worst-case efficiency: $\epsilon_{\rm CM}(x^t) \leq O(1)\frac{\ln(N)[1+\max_{i,j}|a_{ij}|]}{\sqrt{t}}$

Results, I: Restricted Memory Bundle-Level algorithm on the dual of a low size ($p = 512$, $N = 512$) Matrix Completion:

Memory depth	1	33	65	129
${\rm Gap}_1/{\rm Gap}_{1024}$	114	164	350	3253

Results, II: Subgradient Descent on the dual of Matrix Completion:

p	N	${\rm Gap}_1$	${\rm Gap}_1/{\rm Gap}_{32}$	${\rm Gap}_1/{\rm Gap}_{128}$	${\rm Gap}_1/{\rm Gap}_{1024}$	CPU, sec
2048	8192	1.81e-1	171.2	213.8	451.4	521.3
4096	16384	3.74e-1	335.4	1060.8	1287.3	1524.8
8192	16384	2.54e-1	37.8	875.8	1183.6	3644.0

Platform: desktop PC with 4 $\times$ 3.40 GHz *Intel Core2* CPU and 16 GB RAM, Windows 7-64 OS.

Problem: *Given noisy observations a_{ij}, $(i,j) \in \Omega$, of N entries in a $p \times p$ matrix, find matrix of nuclear norm ≤ 1 which is the best fit to the observations in the uniform norm.*

Note:

• **LMO** for the domain X of the problem of interest (unit Nuclear norm ball in $\mathbb{R}^{p\times p}$) is readily given by Power method.

• The objective admits bilinear **F.-t.r.**, the domain Y of the dual problem is the proximal-friendly unit ℓ_1-ball in $\mathbb{R}^N$.

Worst-case efficiency: $\epsilon_{\mathrm{CM}}(x^t) \leq O(1)\frac{\ln(N)[1+\max_{i,j}|a_{ij}|]}{\sqrt{t}}$

Results, I: Restricted Memory Bundle-Level algorithm on the dual of a low size ($p = 512, N = 512$) Matrix Completion:

Memory depth	1	33	65	129
$\mathrm{Gap}_1/\mathrm{Gap}_{1024}$	114	164	350	3253

Results, II: Subgradient Descent on the dual of Matrix Completion:

p	N	Gap_1	$\mathrm{Gap}_1/\mathrm{Gap}_{32}$	$\mathrm{Gap}_1/\mathrm{Gap}_{128}$	$\mathrm{Gap}_1/\mathrm{Gap}_{1024}$	CPU, sec
2048	8192	1.81e-1	171.2	213.8	451.4	521.3
4096	16384	3.74e-1	335.4	1060.8	1287.3	1524.8
8192	16384	2.54e-1	37.8	875.8	1183.6	3644.0

Platform: desktop PC with 4×3.40 GHz *Intel Core2* CPU and 16 GB RAM, Windows 7-64 OS.

Problem: *Given noisy observations a_{ij}, $(i,j) \in \Omega$, of N entries in a $p \times p$ matrix, find matrix of nuclear norm ≤ 1 which is the best fit to the observations in the uniform norm.*

Note:

• **LMO** for the domain X of the problem of interest (unit Nuclear norm ball in $\mathbb{R}^{p\times p}$) is readily given by Power method.

• The objective admits bilinear **F.-t.r.**, the domain Y of the dual problem is the proximal-friendly unit ℓ_1-ball in $\mathbb{R}^N$.

Worst-case efficiency: $\epsilon_{\rm CM}(x^t) \leq O(1)\frac{\ln(N)[1+\max_{i,j}|a_{ij}|]}{\sqrt{t}}$

Results, I: Restricted Memory Bundle-Level algorithm on the dual of a low size ($p = 512, N = 512$) Matrix Completion:

Memory depth	1	33	65	129
$\mathrm{Gap}_1/\mathrm{Gap}_{1024}$	114	164	350	3253

Results, II: Subgradient Descent on the dual of Matrix Completion:

p	N	Gap_1	$\mathrm{Gap}_1/\mathrm{Gap}_{32}$	$\mathrm{Gap}_1/\mathrm{Gap}_{128}$	$\mathrm{Gap}_1/\mathrm{Gap}_{1024}$	CPU, sec
2048	8192	1.81e-1	171.2	213.8	451.4	521.3
4096	16384	3.74e-1	335.4	1060.8	1287.3	1524.8
8192	16384	2.54e-1	37.8	875.8	1183.6	3644.0

Platform: desktop PC with 4 $\times$ 3.40 GHz *Intel Core2* CPU and 16 GB RAM, Windows 7-64 OS.

Problem: *Given a noisy observation* $b \in \mathbb{R}^{q\times q}$ *of the linear image* $u \mapsto \mathcal{A}u : \mathbb{R}^{p\times p} \to \mathbb{R}^{q\times q}$, *find* $p \times p$ *matrix of nuclear norm* ≤ 1 *which is the best fit to the observation in the spectral norm.*

Note:

- The problem reduces to **VI** with affine monotone operator Φ on **LMO**-represented domain (product of two nuclear norm balls)
- Φ admits explicit **F.-t.r.** with affine $G(\cdot)$ and proximal-friendly dual domain (product of two Euclidean balls)

Worst-case efficiency: $\epsilon_{\rm CM}(v^t|f,U_p) \leq \epsilon_{\rm SP}([v^t;w^t]|g,U_p,U_q) \leq O(1)L[\mathcal{A}]/\sqrt{t}$

- $L[\mathcal{A}]$: Lipschitz constant of $u \mapsto \mathcal{A}u$ w.r.t. Frobenius norms.

Accuracy of primal solutions yielded by Subgradient Descent applied to the dual **VI**:

		Iteration count t						
		1	65	129	257	385	449	512
$p=8192$, $q=4048$	$\epsilon^t_{\rm SP}$	0.1193	0.0232	0.0134	0.0054	0.0035	0.0034	0.0034
	$\epsilon^1_{\rm SP}/\epsilon^t_{\rm SP}$	1.00	5.14	8.90	22.00	33.93	34.85	35.14
	cpu, sec	6.5	289.9	683.8	1816.0	3648.3	4572.2	5490.8
$p=16384$, $q=8192$	$\epsilon^t_{\rm SP}$	0.1196	0.0214	0.0146	0.0085			
	$\epsilon^1_{\rm SP}/\epsilon^t_{\rm SP}$	1.00	5.60	8.19	14.01			
	cpu, sec	21.7	920.4	2050.2	4902.2			

Platform: 4 x 3.40 GHz desktop with 16 GB RAM, 64 bit Windows 7 OS.

Note: The design dimension of the largest instance is $2^{28} = 268\,435\,456$.

Problem: *Given a noisy observation $b \in \mathbb{R}^{q\times q}$ of the linear image $u \mapsto \mathcal{A}u : \mathbb{R}^{p\times p} \to \mathbb{R}^{q\times q}$, find $p \times p$ matrix of nuclear norm ≤ 1 which is the best fit to the observation in the spectral norm.*

Note:

- The problem reduces to **VI** with affine monotone operator Φ on **LMO**-represented domain (product of two nuclear norm balls)
- Φ admits explicit **F.-t.r.** with affine $G(\cdot)$ and proximal-friendly dual domain (product of two Euclidean balls)

Worst-case efficiency: $\epsilon_{\rm CM}(v^t|f, U_p) \leq \epsilon_{\rm SP}([v^t; w^t]|g, U_p, U_q) \leq O(1)L[\mathcal{A}]/\sqrt{t}$

- $L[\mathcal{A}]$: Lipschitz constant of $u \mapsto \mathcal{A}u$ w.r.t. Frobenius norms.

Accuracy of primal solutions yielded by Subgradient Descent applied to the dual **VI**:

		Iteration count t						
		1	65	129	257	385	449	512
$p = 8192$	$\epsilon^t_{\rm SP}$	0.1193	0.0232	0.0134	0.0054	0.0035	0.0034	0.0034
$q = 4048$	$\epsilon^1_{\rm SP}/\epsilon^t_{\rm SP}$	1.00	5.14	8.90	22.00	33.93	34.85	35.14
	cpu, sec	6.5	289.9	683.8	1816.0	3648.3	4572.2	5490.8
$p = 16384$	$\epsilon^t_{\rm SP}$	0.1196	0.0214	0.0146	0.0085			
$q = 8192$	$\epsilon^1_{\rm SP}/\epsilon^t_{\rm SP}$	1.00	5.60	8.19	14.01			
	cpu, sec	21.7	920.4	2050.2	4902.2			

Platform: 4 x 3.40 GHz desktop with 16 GB RAM, 64 bit Windows 7 OS.

Note: The design dimension of the largest instance is $2^{28} = 268\,435\,456$.

Problem: *Given a noisy observation $b \in \mathbb{R}^{q\times q}$ of the linear image $u \mapsto \mathcal{A}u : \mathbb{R}^{p\times p} \to \mathbb{R}^{q\times q}$, find $p\times p$ matrix of nuclear norm ≤ 1 which is the best fit to the observation in the spectral norm.*

Note:

- The problem reduces to **VI** with affine monotone operator Φ on **LMO**-represented domain (product of two nuclear norm balls)
- Φ admits explicit **F.-t.r.** with affine $G(\cdot)$ and proximal-friendly dual domain (product of two Euclidean balls)

Worst-case efficiency: $\epsilon_{\rm CM}(v^t|f,U_p) \leq \epsilon_{\rm SP}([v^t;w^t]|g,U_p,U_q) \leq O(1)L[\mathcal{A}]/\sqrt{t}$

- $L[\mathcal{A}]$: Lipschitz constant of $u\mapsto \mathcal{A}u$ w.r.t. Frobenius norms.

Accuracy of primal solutions yielded by Subgradient Descent applied to the dual **VI**:

		Iteration count t						
		1	65	129	257	385	449	512
$p=8192$	$\epsilon^t_{\rm SP}$	0.1193	0.0232	0.0134	0.0054	0.0035	0.0034	0.0034
$q=4048$	$\epsilon^1_{\rm SP}/\epsilon^t_{\rm SP}$	1.00	5.14	8.90	22.00	33.93	34.85	35.14
	cpu, sec	6.5	289.9	683.8	1816.0	3648.3	4572.2	5490.8
$p=16384$	$\epsilon^t_{\rm SP}$	0.1196	0.0214	0.0146	0.0085			
$q=8192$	$\epsilon^1_{\rm SP}/\epsilon^t_{\rm SP}$	1.00	5.60	8.19	14.01			
	cpu, sec	21.7	920.4	2050.2	4902.2			

Platform: 4 x 3.40 GHz desktop with 16 GB RAM, 64 bit Windows 7 OS.

Note: The design dimension of the largest instance is $2^{28} = 268\,435\,456$.

Problem: *Given a noisy observation $b \in \mathbb{R}^{q\times q}$ of the linear image $u \mapsto \mathcal{A}u : \mathbb{R}^{p\times p} \to \mathbb{R}^{q\times q}$, find $p \times p$ matrix of nuclear norm ≤ 1 which is the best fit to the observation in the spectral norm.*

Note:

- The problem reduces to **VI** with affine monotone operator Φ on **LMO**-represented domain (product of two nuclear norm balls)
- Φ admits explicit **F.-t.r.** with affine $G(\cdot)$ and proximal-friendly dual domain (product of two Euclidean balls)

Worst-case efficiency: $\epsilon_{\rm CM}(v^t|f,U_p) \leq \epsilon_{\rm SP}([v^t;w^t]|g,U_p,U_q) \leq O(1)L[\mathcal{A}]/\sqrt{t}$

- $L[\mathcal{A}]$: Lipschitz constant of $u \mapsto \mathcal{A}u$ w.r.t. Frobenius norms.

Accuracy of primal solutions yielded by Subgradient Descent applied to the dual **VI**:

		Iteration count t						
		1	65	129	257	385	449	512
$p = 8192$, $q = 4048$	$\epsilon_{\rm SP}^t$	0.1193	0.0232	0.0134	0.0054	0.0035	0.0034	0.0034
	$\epsilon_{\rm SP}^1/\epsilon_{\rm SP}^t$	1.00	5.14	8.90	22.00	33.93	34.85	35.14
	cpu, sec	6.5	289.9	683.8	1816.0	3648.3	4572.2	5490.8
$p = 16384$, $q = 8192$	$\epsilon_{\rm SP}^t$	0.1196	0.0214	0.0146	0.0085			
	$\epsilon_{\rm SP}^1/\epsilon_{\rm SP}^t$	1.00	5.60	8.19	14.01			
	cpu, sec	21.7	920.4	2050.2	4902.2			

Platform: 4 x 3.40 GHz desktop with 16 GB RAM, 64 bit Windows 7 OS.

Note: The design dimension of the largest instance is $2^{28} = 268\,435\,456$.

Problem: *Given a noisy observation $b \in \mathbb{R}^{q\times q}$ of the linear image $u \mapsto \mathcal{A}u : \mathbb{R}^{p\times p} \to \mathbb{R}^{q\times q}$, find $p \times p$ matrix of nuclear norm ≤ 1 which is the best fit to the observation in the spectral norm.*

Note:

- The problem reduces to **VI** with affine monotone operator Φ on **LMO**-represented domain (product of two nuclear norm balls)
- Φ admits explicit **F.-t.r.** with affine $G(\cdot)$ and proximal-friendly dual domain (product of two Euclidean balls)

Worst-case efficiency: $\epsilon_{\rm CM}(v^t|f,U_p) \leq \epsilon_{\rm SP}([v^t;w^t]|g,U_p,U_q) \leq O(1)L[\mathcal{A}]/\sqrt{t}$

- $L[\mathcal{A}]$: Lipschitz constant of $u \mapsto \mathcal{A}u$ w.r.t. Frobenius norms.

Accuracy of primal solutions yielded by Subgradient Descent applied to the dual **VI**:

		Iteration count t						
		1	65	129	257	385	449	512
$p = 8192$	$\epsilon^t_{\rm SP}$	0.1193	0.0232	0.0134	0.0054	0.0035	0.0034	0.0034
$q = 4048$	$\epsilon^1_{\rm SP}/\epsilon^t_{\rm SP}$	1.00	5.14	8.90	22.00	33.93	34.85	35.14
	cpu, sec	6.5	289.9	683.8	1816.0	3648.3	4572.2	5490.8
$p = 16384$	$\epsilon^t_{\rm SP}$	0.1196	0.0214	0.0146	0.0085			
$q = 8192$	$\epsilon^1_{\rm SP}/\epsilon^t_{\rm SP}$	1.00	5.60	8.19	14.01			
	cpu, sec	21.7	920.4	2050.2	4902.2			

Platform: 4 x 3.40 GHz desktop with 16 GB RAM, 64 bit Windows 7 OS.

Note: The design dimension of the largest instance is $2^{28} = 268\,435\,456$.

Observation: Structured Convex-Concave "master" SP problem

$$\mathrm{SadVal}(M) = \min_{[u;w]\in U\times W}\ \max_{[v;z]\in V\times Z} \Phi(u,w;v,z) \qquad \text{(M)}$$

gives rise to *primal-dual pair* of Convex-Concave SP's

$$\mathrm{SadVal}(\mathrm{Pr}) = \min_{u\in U}\max_{v\in V}\left[\phi(u;v) := \min_{w\in W}\max_{z\in Z}\Phi(u,w;v,z)\right] \quad \text{(Pr)}$$
$$\mathrm{SadVal}(\mathrm{Dl}) = \min_{w\in W}\max_{z\in Z}\left[\psi(w;z) = \min_{u\in U}\max_{v\in V}\Phi(u,w;v,z)\right] \quad \text{(Dl)}$$

with $\mathrm{SadVal}(M) = \mathrm{SadVal}(\mathrm{Pr}) = \mathrm{SadVal}(\mathrm{Dl})$.

Fact [Cox,Juditsky,Nem.'15]: *Assuming U, W, V, Z compact and Φ Lipschitz and differentiable, a t-step execution protocol $\mathcal{I}_t$ and accuracy certificate λ for (Pr) induce feasible approximate solution $[u^t; w^t; v^t; z^t]$ to (M) such that all three saddle point inaccuracies*

$$\epsilon_{\rm SP}([u^t;v^t]|\phi,U,V),\ \epsilon_{\rm SP}([w^t;z^t]|\psi,W,Z),\ \epsilon_{\rm SP}([u^t;w^t;v^t;z^t]|\Phi,U\times W,V\times Z)$$

do not exceed $\mathrm{Res}(\mathcal{I}_t,\lambda|U\times V)$.

Observation: Structured Convex-Concave "master" SP problem

$$\mathrm{SadVal}(M) = \min_{[u;w]\in U\times W} \max_{[v;z]\in V\times Z} \Phi(u,w;v,z) \qquad \text{(M)}$$

gives rise to *primal-dual pair* of Convex-Concave SP's

$$\begin{array}{ll} \mathrm{SadVal}(\mathrm{Pr}) = \min\limits_{u\in U}\max\limits_{v\in V}\left[\phi(u;v) := \min\limits_{w\in W}\max\limits_{z\in Z}\Phi(u,w;v,z)\right] & \text{(Pr)} \\ \mathrm{SadVal}(\mathrm{Dl}) = \min\limits_{w\in W}\max\limits_{z\in Z}\left[\psi(w;z) = \min\limits_{u\in U}\max\limits_{v\in V}\Phi(u,w;v,z)\right] & \text{(Dl)} \end{array}$$

with $\mathrm{SadVal}(M) = \mathrm{SadVal}(\mathrm{Pr}) = \mathrm{SadVal}(\mathrm{Dl})$.

Fact [Cox,Juditsky,Nem.'15]: *Assuming U, W, V, Z compact and Φ Lipschitz and differentiable, a t-step execution protocol $\mathcal{I}_t$ and accuracy certificate λ for (Pr) induce feasible approximate solution $[u^t; w^t; v^t; z^t]$ to (M) such that all three saddle point inaccuracies*

$$\epsilon_{\mathrm{SP}}([u^t;v^t]|\phi,U,V),\ \epsilon_{\mathrm{SP}}([w^t;z^t]|\psi,W,Z),\ \epsilon_{\mathrm{SP}}([u^t;w^t;v^t;z^t]|\Phi,U\times W,V\times Z)$$

do not exceed $\mathrm{Res}(\mathcal{I}_t,\lambda|U\times V)$.

Example 1: To solve bilinear SP problem

$$\min_{w\in W}\max_{z\in Z}\left[\psi(w,z) := \langle p,w\rangle + \langle q,z\rangle + \langle z,Sw\rangle\right] \qquad (!)$$

with LMO-represented W, Z, set the master program as

$$\min_{[u;w]\in U\times W}\max_{[v;z]\in V\times Z}\left[\Phi(u,w;v,z) = \langle w, p+S^Tv\rangle + \langle z, q+Su\rangle - \langle v,Su\rangle\right]$$

$$[U\supset W, V\supset Z]$$

$\Rightarrow$ (!) is the dual SP problem induced by (M). The induced primal is

$$\min_{u\in U}\max_{v\in V}\left[\min_{w\in W}\langle w, p+S^Tv\rangle + \max_{z\in Z}\langle z,q+Su\rangle - \langle v,Su\rangle\right] \qquad (\mathrm{Pr})$$

Note:

- LMO's for W, Z induce First Order oracle for (Pr)
- We can make U, V proximal-friendly

$\Rightarrow$ *Solving* (Pr) *by proximal algorithm with certificates, we get $O(1/\sqrt{t})$ algorithm for bilinear SP on LMO-represented domains.*

Example 1: To solve bilinear SP problem

$$\min_{w\in W}\max_{z\in Z}\left[\psi(w,z) := \langle p,w\rangle + \langle q,z\rangle + \langle z,Sw\rangle\right] \qquad (!)$$

with LMO-represented W, Z, set the master program as

$$\min_{[u;w]\in U\times W}\max_{[v;z]\in V\times Z}\left[\Phi(u,w;v,z) = \langle w, p+S^Tv\rangle + \langle z, q+Su\rangle - \langle v, Su\rangle\right]$$
$$[U\supset W, V\supset Z]$$

$\Rightarrow$ (!) is the dual SP problem induced by (M). The induced primal is

$$\min_{u\in U}\max_{v\in V}\left[\min_{w\in W}\langle w, p+S^Tv\rangle + \max_{z\in Z}\langle z, q+Su\rangle - \langle v, Su\rangle\right] \qquad (\mathrm{Pr})$$

Note:

- LMO's for W, Z induce First Order oracle for (Pr)
- We can make U, V proximal-friendly

$\Rightarrow$ *Solving* (Pr) *by proximal algorithm with certificates, we get $O(1/\sqrt{t})$ algorithm for bilinear SP on LMO-represented domains.*

Example 1: To solve bilinear SP problem

$$\min_{w\in W}\max_{z\in Z}\left[\psi(w,z):=\langle p,w\rangle+\langle q,z\rangle+\langle z,Sw\rangle\right] \qquad (!)$$

with LMO-represented W, Z, set the master program as

$$\min_{[u;w]\in U\times W}\max_{[v;z]\in V\times Z}\left[\Phi(u,w;v,z)=\langle w,p+S^Tv\rangle+\langle z,q+Su\rangle-\langle v,Su\rangle\right]$$
$$[U\supset W, V\supset Z]$$

$\Rightarrow$ (!) is the dual SP problem induced by (M). The induced primal is

$$\min_{u\in U}\max_{v\in V}\left[\min_{w\in W}\langle w,p+S^Tv\rangle+\max_{z\in Z}\langle z,q+Su\rangle-\langle v,Su\rangle\right] \qquad (\mathrm{Pr})$$

Note:

- LMO's for W, Z induce First Order oracle for (Pr)
- We can make U, V proximal-friendly

$\Rightarrow$ *Solving* (Pr) *by proximal algorithm with certificates, we get $O(1/\sqrt{t})$ algorithm for bilinear SP on LMO-represented domains.*

Example 1: To solve bilinear SP problem

$$\min_{w\in W}\max_{z\in Z}\left[\psi(w,z):=\langle p,w\rangle+\langle q,z\rangle+\langle z,Sw\rangle\right] \qquad (!)$$

with LMO-represented W, Z, set the master program as

$$\min_{[u;w]\in U\times W}\max_{[v;z]\in V\times Z}\left[\Phi(u,w;v,z)=\langle w,p+S^Tv\rangle+\langle z,q+Su\rangle-\langle v,Su\rangle\right]$$
$$[U\supset W,\, V\supset Z]$$

$\Rightarrow$ (!) is the dual SP problem induced by (M). The induced primal is

$$\min_{u\in U}\max_{v\in V}\left[\min_{w\in W}\langle w,p+S^Tv\rangle+\max_{z\in Z}\langle z,q+Su\rangle-\langle v,Su\rangle\right] \qquad (\mathrm{Pr})$$

Note:

- LMO's for W, Z induce First Order oracle for (Pr)
- We can make U, V proximal-friendly

$\Rightarrow$ *Solving (Pr) by proximal algorithm with certificates, we get $O(1/\sqrt{t})$ algorithm for bilinear SP on LMO-represented domains.*

Example 2: To solve matrix game

$$\min_{w\in\Delta_M}\max_{z\in\Delta_N}[\psi(w,z)=\langle Az,Dw\rangle] \qquad (!)$$

$$\left[\Delta_k=\{x\in\mathbb{R}^k_+:\sum_i x_i=1\},\ D:K\times M,\ A:K\times N\right]$$

set the master program as

$$\min_{[u;w]\in U\times\Delta_M}\max_{[v;z]\in V\times\Delta_N}[\Phi(u,w;v,z)=\langle u,Az\rangle+\langle v,Dw\rangle-\langle v,u\rangle] \qquad (\mathrm{M})$$

$$[U\supset DW,\ V\supset AZ]$$

$\Rightarrow$ (!) is the dual SP problem induced by (M). The induced primal is

$$\min_{u\in U\subset\mathbb{R}^K}\max_{v\in V\subset\mathbb{R}^K}\left[\phi(u,v)=\mathrm{Max}(A^Tu)-\mathrm{Min}(D^Tv)-\langle u,v\rangle\right] \qquad (\mathrm{Pr})$$

Note: In some applications,

- the dimension K of (Pr) is small
- matrices A and D are *well organized*: it is easy to find a column making the largest inner product with a given $x\in\mathbb{R}^K$.

$\Rightarrow$ *Applying to* (Pr) *the Ellipsoid algorithm, we solve* (!) *to high accuracy with effort independent of the sizes M,N of* (!)

Example 2: To solve matrix game

$$\min_{w\in\Delta_M}\max_{z\in\Delta_N}[\psi(w,z)=\langle Az,Dw\rangle] \qquad (!)$$

$$\left[\Delta_k=\{x\in\mathbb{R}^k_+:\sum_i x_i=1\},\ D:K\times M,\ A:K\times N\right]$$

set the master program as

$$\min_{[u;w]\in U\times\Delta_M}\ \max_{[v;z]\in V\times\Delta_N}[\Phi(u,w;v,z)=\langle u,Az\rangle+\langle v,Dw\rangle-\langle v,u\rangle] \qquad (\mathrm{M})$$

$$[U\supset DW,\ V\supset AZ]$$

$\Rightarrow$ (!) is the dual SP problem induced by (M). The induced primal is

$$\min_{u\in U\subset\mathbb{R}^K}\ \max_{v\in V\subset\mathbb{R}^K}\left[\phi(u,v)=\mathrm{Max}(A^Tu)-\mathrm{Min}(D^Tv)-\langle u,v\rangle\right] \quad (\mathrm{Pr})$$

Note: In some applications,

- the dimension K of (Pr) is small
- matrices A and D are *well organized*: it is easy to find a column making the largest inner product with a given $x\in\mathbb{R}^K$.

$\Rightarrow$ *Applying to* (Pr) *the Ellipsoid algorithm, we solve* (!) *to high accuracy with effort independent of the sizes M,N of* (!)

Example 2: To solve matrix game

$$\min_{w\in\Delta_M}\max_{z\in\Delta_N}\left[\psi(w,z)=\langle Az,Dw\rangle\right] \qquad (!)$$

$$\left[\Delta_k=\{x\in\mathbb{R}^k_+:\sum_i x_i=1\},\ D:K\times M,\ A:K\times N\right]$$

set the master program as

$$\min_{[u;w]\in U\times\Delta_M}\max_{[v;z]\in V\times\Delta_N}\left[\Phi(u,w;v,z)=\langle u,Az\rangle+\langle v,Dw\rangle-\langle v,u\rangle\right] \qquad (\mathrm{M})$$

$$[U\supset DW,\ V\supset AZ]$$

$\Rightarrow$ (!) is the dual SP problem induced by (M). The induced primal is

$$\min_{u\in U\subset\mathbb{R}^K}\max_{v\in V\subset\mathbb{R}^K}\left[\phi(u,v)=\mathrm{Max}(A^Tu)-\mathrm{Min}(D^Tv)-\langle u,v\rangle\right] \quad (\mathrm{Pr})$$

Note: In some applications,

- the dimension K of (Pr) is small
- matrices A and D are *well organized*: it is easy to find a column making the largest inner product with a given $x\in\mathbb{R}^K$.

$\Rightarrow$ *Applying to* (Pr) *the Ellipsoid algorithm, we solve* (!) *to high accuracy with effort independent of the sizes M, N of* (!)

Example 2: To solve matrix game

$$\min_{w\in\Delta_M}\max_{z\in\Delta_N}[\psi(w,z)=\langle Az,Dw\rangle] \qquad (!)$$

$$\left[\Delta_k=\{x\in\mathbb{R}^k_+:\sum_i x_i=1\},\ D:K\times M,\ A:K\times N\right]$$

set the master program as

$$\min_{[u;w]\in U\times\Delta_M}\max_{[v;z]\in V\times\Delta_N}[\Phi(u,w;v,z)=\langle u,Az\rangle+\langle v,Dw\rangle-\langle v,u\rangle] \qquad (\mathrm{M})$$

$$[U\supset DW,\ V\supset AZ]$$

$\Rightarrow$ (!) is the dual SP problem induced by (M). The induced primal is

$$\min_{u\in U\subset\mathbb{R}^K}\max_{v\in V\subset\mathbb{R}^K}\left[\phi(u,v)=\mathrm{Max}(A^Tu)-\mathrm{Min}(D^Tv)-\langle u,v\rangle\right] \quad (\mathrm{Pr})$$

Note: In some applications,

- the dimension K of (Pr) is small
- matrices A and D are *well organized*: it is easy to find a column making the largest inner product with a given $x\in\mathbb{R}^K$.

$\Rightarrow$ *Applying to (Pr) the Ellipsoid algorithm, we solve (!) to high accuracy with effort independent of the sizes M,N of (!)*