Mirror Prox Algorithm for Multi-Term Composite Minimization

Niao He

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

Joint work with Anatoli Juditsky and Arkadi Nemirovski

DOS, October 29, 2014
Motivational Example I

Matrix Completion Problem

\[ \min_{y \in \mathbb{R}^{n \times n}} \left\{ \frac{1}{2} \| P_{\Omega} y - b \|_2^2 + \mu_1 \| y \|_1 + \mu_2 \| y \|_{\text{nuc}} \right\} \]

Regularizations:
- $\ell_1$ norm $\rightarrow$ sparsity
- nuclear norm $\rightarrow$ low-rank

Social Network

Low-rank and Sparse Matrix
Motivational Example II

**Image Decomposition Problem**

\[
\min_{y_1, y_2, y_3 \in \mathbb{R}^{n \times n}} \left\{ \|A(y_1 + y_2 + y_3) - b\|_2^2 + \mu_1 \|Dy_1\|_1 + \mu_2 \cdot TV(y_2) + \mu_3 \|y_3\|_1 \right\}
\]

\[
\approx \text{image} + \text{texture } (y_1) + \text{smooth } (y_2) + \text{noise } (y_3) + \ldots
\]
Multi-Term Composite Minimization

\[
\min_{y \in Y} \left\{ \sum_{k=1}^{K} \left[ \phi_k(A_k y) + \Psi_k(A_k y) \right] \right\}
\]

- \(K \geq 2\)
- \(\phi_k, 1 \leq k \leq K\): smooth convex functions
- \(\Psi_k, 1 \leq k \leq K\): “simple” nonsmooth functions

**Simplicity of \(\Psi(\cdot)\):**
- Proximal operator is easy to compute:
  \[
  \text{prox}_\alpha(\Psi)(u) = \argmin_y \left\{ \frac{1}{2} \|y - u\|_2^2 + \alpha \Psi(y) \right\}
  \]
  - \(\Psi(\cdot) = \|\cdot\|_1\), \(\text{prox}_\alpha(\Psi)(u) = sgn(u) \circ \max\{|u| - \alpha, 0\}\) (-shrinkage-)
  - \(\Psi(\cdot) = \|\cdot\|_{\text{nuc}}\), \(\text{prox}_\alpha(\Psi)(u) = U^T \text{diag}(sgn(\sigma) \circ \max\{\sigma| - \alpha, 0\}) V\)

Lots and lots of applications ...
Existing Algorithms

\[
\min_{y \in Y} \left\{ \sum_{k=1}^{K} [\phi_k(A_ky) + \Psi_k(A_ky)] \right\}
\]

- **Fast Gradient Method** [Y. Nesterov, 2004]
  \[ \text{prox}_\alpha(\| \cdot \|_1) \text{ is easy,} \]
  \[ \text{prox}_\alpha(\| \cdot \|_{\text{nuc}}) \text{ is easy,} \]
  \[ \text{but } \text{prox}_\alpha(\mu_1 \| \cdot \|_1 + \mu_2 \| \cdot \|_{\text{nuc}}) ? \]

- **Proximal-Average Method** [Y. Yu, 2013]
  \[ \text{prox}_\alpha(\mu_1 \| \cdot \|_1 + \mu_2 \| \cdot \|_{\text{nuc}}) \approx \mu_1 \text{prox}_\alpha(\| \cdot \|_1) + \mu_2 \text{prox}_\alpha(\| \cdot \|_{\text{nuc}}) \]

- **Alternating Direction Method of Multipliers (ADMM)**
  😞 Difficult to solve subproblems explicitly for many cases.

Lots and lots of algorithms...
Outline

0  Motivation and Background

1  Preliminaries: Variational Inequalities

2  Working Horse: Composite Mirror Prox algorithm

3  Application: Multi-Term Composite Minimization

4  Extension to Other Situations
Outline

0 Motivation and Background

1 **Preliminaries:** Variational Inequalities

2 **Working Horse:** Composite Mirror Prox algorithm

3 **Application:** Multi-Term Composite Minimization

4 Extension to Other Situations
Variational Inequalities — Facts you already know

Variational Inequality \( \text{VI}(X, F) \)

Find \( x_\ast \in X : \langle F(x), x - x_\ast \rangle \geq 0, \forall x \in X \) (weak solution)

Find \( x_\ast \in X : \langle F(x_\ast), x - x_\ast \rangle \geq 0, \forall x \in X \) (strong solution)

Fact I:

weak solution \( \xrightarrow{\text{F is continuous}} \) strong solution

Fact II:

- Convex minimization
  \[
  \min_{x \in X} f(x) \iff \text{VI}(X, F) \text{ with } F = \nabla f(x)
  \]

- Convex-concave saddle point problem
  \[
  \min_{x_1 \in X_1} \max_{x_2 \in X_2} \phi(x_1, x_2) \iff \text{VI}(X_1 \times X_2, F) \text{ with } F = \begin{bmatrix} \nabla_{x_1} \phi(x_1, x_2) \\ -\nabla_{x_2} \phi(x_1, x_2) \end{bmatrix}
  \]

- Nash equilibrium / complementary / fixed-point problems
Variational Inequalities — Facts you already know

Variational Inequality $VI(X, F)$

Find $x^* \in X : \langle F(x), x - x^* \rangle \geq 0, \forall x \in X$

Accuracy measure of candidate solution $\hat{x} \in X$:

$$\epsilon_{VI}(\hat{x}|X, F) := \sup_{x \in X} \langle F(x), \hat{x} - x \rangle.$$ 

Fact III: optimal rate of convergence [Nemirovski, 2004]

- $O(M/\sqrt{t})$ if $\|F(x)\|_* \leq M, \forall x \in X$;
- $O(L/t)$ if $\|F(x) - F(x')\|_* \leq L \|x - x'\|, \forall x, x' \in X$

can be achieved by Mirror Prox algorithm

\[
\begin{align*}
    x_1 &\in X \\
y_\tau &= \operatorname{Argmin}_{x \in X} \{V_\omega(x, x_\tau) + \langle \gamma_\tau F(x_\tau), x \rangle\} \\
x_{\tau+1} &= \operatorname{Argmin}_{x \in X} \{V_\omega(x, x_\tau) + \langle \gamma_\tau F(y_\tau), x \rangle\} \\
x^t &= \left[\sum_{\tau=1}^t \gamma_\tau\right]^{-1} \sum_{\tau=1}^t \gamma_\tau y_\tau
\end{align*}
\]

- Gradient Descent
  $$V_\omega(x, y) = \frac{1}{2} \|x - y\|_2^2$$
  (Euclidean distance)

- Mirror Descent
  $$V_\omega(x, y) = \omega(x) - \omega(y) - \langle \omega'(y), x - y \rangle$$
  (Bregman distance)
Back to Multi-Term Composite Minimization

$$\min_{y \in Y} \left\{ \sum_{k=1}^{K} [\phi_k(A_ky) + \Psi_k(A_ky)] \right\}$$

- $\text{VI}(Y, F)$ with $F(y) = \sum_{k=1}^{K} A_k^T [\nabla \phi(A_ky) + \partial \Psi_k(A_ky)]$

- Directly applying Mirror Prox algorithm would render $O(1/\sqrt{t})$ rate of convergence.

Can we do better?
Outline

0 Motivation and Background

1 Preliminaries: Variational Inequalities

2 Working Horse: Composite Mirror Prox algorithm

3 Application: Multi-Term Composite Minimization

4 Extension to Other Situations
Composite Mirror Prox — Goal and Situation

Special Variational Inequality VI\((X, F)\)

Find \(x_* \in X : \langle F(x), x - x_* \rangle \geq 0, \forall x \in X\)

- **Domain X**
  - \(x = [u; v] \in X\) and \(X\) is closed convex;
  - **Bregman distance on** \(P_u X\):
    \(V_\omega(u, u') = \omega(u) - \omega(u') - \langle \nabla \omega(u'), u - u' \rangle\)
  - generated by strongly convex function \(\omega(u)\).

- **Operator** \(F\)
  - \(F(x = [u, v]) = [F_u(u); F_v]\) is independent of \(v\),
  - \(\forall u, u' \in P_u X : \|F_u(u) - F_u(u')\|_* \leq L\|u - u'\| + M\);
Composite Mirror Prox — The Algorithm

General VI \( (X, F) \)

\[
x_1 \in X
y_{\tau} = \operatorname{Argmin}_{x \in X} \{ V_\omega(x, x_{\tau}) + \langle \gamma_{\tau} F(x_{\tau}), x \rangle \}
x_{\tau + 1} = \operatorname{Argmin}_{x \in X} \{ V_\omega(x, x_{\tau}) + \langle \gamma_{\tau} F(y_{\tau}), x \rangle \}
x^t = \left[ \sum_{\tau=1}^{t} \gamma_{\tau} \right]^{-1} \sum_{\tau=1}^{t} \gamma_{\tau} y_{\tau}
\]

Mirror Prox algorithm

Special structured VI \( (X, F') \) with \( F(x) = [F_u(u); F_v] \)

\[
x_1 := [u_1; v_1] \in X
y_{\tau} := [p_{\tau}; q_{\tau}] = \operatorname{Argmin}_{x=[u; v] \in X} \{ V_\omega(u, u_{\tau}) + \langle \gamma_{\tau} F_u(u_{\tau}), u \rangle + \langle \gamma_{\tau} F_v, v \rangle \}
x_{\tau + 1} := [u_{\tau + 1}; v_{\tau + 1}] = \operatorname{Argmin}_{x=[u; v] \in X} \{ V_\omega(u, u_{\tau}) + \langle \gamma_{\tau} F_u(p_{\tau}), u \rangle + \langle \gamma_{\tau} F_v, v \rangle \}
x^t = \left[ \sum_{\tau=1}^{t} \gamma_{\tau} \right]^{-1} \sum_{\tau=1}^{t} \gamma_{\tau} y_{\tau}
\]

Composite Mirror Prox algorithm
**Theorem**

Let $x^t$ be generated by the Composite Mirror Prox algorithm by setting $0 < \gamma \leq \frac{1}{\sqrt{2L}}$. Then

$$
\epsilon_{VI}(x^t \mid X, F) \leq \frac{\Theta[X] + M^2 \sum_{\tau=1}^{t} \gamma^2}{\sum_{\tau=1}^{t} \gamma},
$$

particularly, when $M = 0$, one can ensure that

$$
\epsilon_{VI}(x^t \mid X, F') \leq \frac{\Theta[X]L}{t},
$$

where $\Theta[X] = \sup_{[u\nu] \in X} V_\omega(u, u_1)$.

- $O(M/\sqrt{t})$ if $\|F_u(u)\|_* \leq M$;
- $O(L/t)$ if $\|F_u(u) - F_u(u')\|_* \leq L\|u - u'\|$.
Motivation and Background

1 Preliminaries: Variational Inequalities

2 Working Horse: Composite Mirror Prox algorithm

3 Application: Multi-Term Composite Minimization

4 Extension to Other Situations
Back Again to Multi-Term Composite Minimization

\[ \min_{y \in Y} \left\{ \sum_{k=1}^{K} [\phi_k(A_k y) + \Psi_k(A_k y)] \right\} \]

- \( K \geq 2 \)
- \( \phi_k(\cdot), 1 \leq k \leq K \): smooth convex functions on \( Y_k \);
- \( \Psi_k(\cdot), 1 \leq k \leq K \): "simple" convex functions on \( Y_k \);

**Simplicity of \( \Psi_k(\cdot) \):**
- Proximal operator is easy to compute:
  \( \xi^k \mapsto \arg\min_{y^k \in Y_k} \left[ \frac{1}{2} \|y^k - \xi^k\|_2^2 + \alpha \Psi_k(y^k) \right] \)

What does this have to do with the special variational inequality?
Problem of Interest

\[ \text{Opt} = \min_{y^0 \in Y_0} \left\{ v(y^0) := \sum_{k=0}^{K} \left[ \phi_k(A_ky^0) + \Psi_k(A_ky^0) \right] \right\} \quad (P.1) \]

\[ = \min_{x^1 \in X_1} \left\{ \Upsilon(x^1) := \sum_{k=0}^{K} [\phi_k(y^k) + \eta^k] : y^k = A_ky^0, 1 \leq k \leq K \right\} \quad (P.2) \]

where \( X_1 := \{ x^1 = \{ [y^k; \eta^k] \}_{k=0}^{K} : y^k \in Y_k, \eta^k \geq \Psi_k(y^k), 0 \leq k \leq K \} \).

Saddle Point Approximation

\[ \hat{\text{Opt}} = \min_{x^1 \in X_1} \max_{x^2 \in X_2} \left\{ \sum_{k=0}^{K} [\phi_k(y^k) + \eta^k] + \sum_{k=1}^{K} \rho_k \langle y^k - A_ky^0, w^k \rangle \right\} \quad (S.1) \]

\[ = \min_{x^1 \in X_1} \left\{ \overline{\Phi}(x^1) := \sum_{k=0}^{K} [\phi_k(y^k) + \eta^k] + \sum_{k=1}^{K} \rho_k \| y^k - A_ky^0 \|_{k,*} \right\} \quad (S.2) \]

where \( X_2 := \{ x^2 = [w^1; \ldots; w^K] : \| w^k \|_k \leq 1, 1 \leq k \leq K \} \).

Special Structured VI \((X, F)\) Meets The Situation

Set \( x = [u = [y^0; \ldots; y^K; w^1; \ldots; w^K]; v = [\eta^0; \ldots; \eta^K]] \)

- \( X = \{ x = [u; v] : y^k \in Y_k, \| w^k \|_k \leq 1, \eta^k \geq \Psi_k(y^k), 0 \leq k \leq K \} \)
- \( F = [F_u(u); F_v], \) where \( F_v = [1; 1; \ldots; 1] \).
Multi-Term Composite Minimization — An Example

Example:
\[
\min_{y \in \mathbb{R}^{n \times n}} \left\{ \frac{1}{2} \|Py - b\|_2^2 + \mu_1 \|y\|_1 + \mu_2 \|Ay\|_{\text{nuc}} \right\}
\]

\[\uparrow \text{ (move nonsmooth terms)}\]

\[
\min_{\tau_1 \geq \mu_1 \|y\|_1, \tau_2 \geq \mu_2 \|z\|_{\text{nuc}}} \left\{ \frac{1}{2} \|Py - b\|_2^2 + \tau_1 + \tau_2 : Ay = z \right\}
\]

\[\uparrow \text{ (penalize constraints with proper } \rho \text{)}\]

\[
\min_{\tau_1 \geq \mu_1 \|y\|_1, \tau_2 \geq \mu_2 \|z\|_{\text{nuc}}} \left\{ \frac{1}{2} \|Py - b\|_2^2 + \tau_1 + \tau_2 + \rho \|Ay - z\|_2 \right\}
\]

\[\uparrow \text{ (use dual norm)}\]

Convex-Concave Saddle Point Problem:
\[
\min_{\tau_1 \geq \mu_1 \|y\|_1, \tau_2 \geq \mu_2 \|z\|_{\text{nuc}}} \max_{\|w\|_2 \leq 1} \left\{ \frac{1}{2} \|Py - b\|_2^2 + \tau_1 + \tau_2 + \rho \langle Ay - z, w \rangle \right\}
\]
Multi-Term Composite Minimization — An Example

Convex-Concave Saddle Point Problem:

\[
\begin{align*}
\min_{\tau_1 \geq \|y\|_1, \tau_2 \geq \mu_2 \|z\|_{\text{nuc}}} & \quad \max_{\|w\|_2 \leq 1} \left\{ \frac{1}{2} \|Py - b\|_2^2 + \tau_1 + \tau_2 + \rho \langle Ay - z, w \rangle \right\} \\
\end{align*}
\]

Variational Inequality VI(\(X, F\)):

- \(X = \{ x = [y, z, w; \tau_1, \tau_2] : \tau_1 \geq \mu_1 \|y\|_1, \tau_2 \geq \mu_2 \|z\|_{\text{nuc}}, \|w\|_2 \leq 1 \}\)
  - \(x = [u; v] \) with \(u = [y, z, w], v = [\tau_1, \tau_2], X \) is closed convex
  - Bregman distance generated by \(\omega(u) = \alpha_1 \omega_1(y) + \alpha_2 \omega_2(z) + \alpha_3 \omega_3(w)\)
- \(F = [P^T(Py - b) + \rho A^T w; -\rho w; \rho(z - Ay); 1; 1]\)
  - \(F(x) = [F_u(u); F_v]\) of special structure
  - \(F_u(u)\) is Lipschitz continuous
Multi-Term Composite Minimization — Summary

\[
\min_{y \in Y} \left\{ \sum_{k=1}^{K} [\phi_k(A_k y) + \Psi_k(A_k y)] \right\}
\]

\[\iff\]
\[
\min_{y \in Y} \left\{ \sum_{k=1}^{K} [\phi_k(y_k) + \Psi_k(y_k)] : A_k y = y_k, 1 \leq k \leq K \right\}
\]

Basic Strategy

1. Penalize the linear constraints by \(\sum_{k=1}^{K} \rho_k \|A_k y - y_k\|_{k,*}\);
   - Smiley: Small magnitude of penalty parameters

2. Convert into a saddle point problem;

3. Apply Composite Mirror Prox to its variational inequality;
   - Smiley: Easy and separable updates at each step

4. “Correct” solution and obtain \(O(1/t)\) rate of convergence.
   - Smiley: Best rate of convergence known so far
Multi-Term Composite Minimization — How It Works

Test Problem I: Matrix Completion

\[
\text{Opt} = \min_{y \in \mathbb{R}^{n \times n}} \left\{ \frac{1}{2} \| P_\Omega y - b \|_2^2 + \mu_1 \| y \|_1 + \mu_2 \| y \|_{\text{nuc}} \right\}
\]

<table>
<thead>
<tr>
<th>$t$</th>
<th>16</th>
<th>32</th>
<th>64</th>
<th>128</th>
<th>256</th>
<th>512</th>
<th>1024</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU, sec</td>
<td>6.9</td>
<td>93.8</td>
<td>187.6</td>
<td>375.3</td>
<td>750.6</td>
<td>1501.2</td>
<td>3002.3</td>
</tr>
<tr>
<td>$v^t - v_t$</td>
<td>1.5e2</td>
<td>1.3e2</td>
<td>1.2e2</td>
<td>1.1e2</td>
<td>8.0e1</td>
<td>1.6e1</td>
<td>5.4e0</td>
</tr>
<tr>
<td>$\frac{v^t - v_t}{v_{1024}}$</td>
<td>2.2e-1</td>
<td>2.2e-1</td>
<td>1.9e-1</td>
<td>1.7e-01</td>
<td>1.2e-1</td>
<td>2.4e-2</td>
<td>8.1e-3</td>
</tr>
</tbody>
</table>
(a) $n = 1024$, $v_{1024} = 655.4 \leq \text{Opt} \leq v_{1024} = 660.8$

<table>
<thead>
<tr>
<th>$t$</th>
<th>1</th>
<th>7</th>
<th>8</th>
<th>128</th>
<th>256</th>
<th>512</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU, sec</td>
<td>8.9</td>
<td>48.1</td>
<td>51.9</td>
<td>392.7</td>
<td>752.1</td>
<td>1464.9</td>
</tr>
<tr>
<td>$v^t - \text{Opt}$</td>
<td>3.7e2</td>
<td>3.5e1</td>
<td>2.2e-1</td>
<td>2.1e-1</td>
<td>1.9e-1</td>
<td>1.6e-1</td>
</tr>
<tr>
<td>$\frac{v^t - \text{Opt}}{\text{Opt}}$</td>
<td>1.5e-1</td>
<td>1.5e-3</td>
<td>9.2e-5</td>
<td>9.0e-5</td>
<td>8.1e-5</td>
<td>7.0e-5</td>
</tr>
</tbody>
</table>
(b) $n = 1024$, Opt = 2401.2

(a) partial $P_\Omega$; (b) full $P_\Omega$; $v^t$ – upper bound; $v_t$ – lower bound;

Platform: Intel i7-3770 @2x3.40 GHz CPU, 16GB RAM, 64-bit Windows 7
Multi-Term Composite Minimization — How It Works

Test Problem II: Image Decomposition

\[
\min_{\substack{y_1, y_2, y_3 \\
\in \mathbb{R}^{n \times n}}} \left\{ \| A(y_1 + y_2 + y_3) - b \|_2^2 + \mu_1 \| Dy_1 \|_1 + \mu_2 \cdot TV(y_2) + \mu_3 \| y_3 \|_1 \right\}
\]

Observation (256 × 256)

\( y_1 \): texture

\( y_2 \): smooth

\( y_3 \): noise

Results by applying CMP of 1000 steps within 294.6 secs.

Platform: Intel i5-2400S @2.5GHz CPU, 4GB RAM, 64-bit Windows 7
Multi-Term Composite Minimization — How It Works

Test Problem III: Fused LASSO

\[
\min_{y \in \mathbb{R}^n} \left\{ \frac{1}{N} \sum_{i=1}^{N} \frac{1}{2} (s_i^T y - \ell_i)^2 + \mu \|y\|_1 + \nu \sum_{i=1}^{n-1} |y_{i+1} - y_i| \right\}
\]

\(N = 5000, \ n = 10000\)

Error vs. Iteration

Error vs. Time
Not the end of story yet...

\[
\min_{y \in Y} \left\{ \sum_{k=1}^{K} \left[ \phi_k(A_k y) + \Psi_k(A_k y) \right] \right\}
\]

- \( \phi_k(\cdot), 1 \leq k \leq K \): smooth convex functions;
- \( \Psi_k(\cdot), 1 \leq k \leq K \): simple nonsmooth convex functions;

Questions:
- What if \( \phi_k \) are non-smooth?
- What if \( \Psi_k \) does not admit cheap prox-mappings?
- What if there are linking constraints?
Outline

0 Motivation and Background

1 Preliminaries: Variational Inequalities

2 Working Horse: Composite Mirror Prox algorithm

3 Application: Multi-Term Composite Minimization

4 Extension to Other Situations
Extension I:  
— from **Smooth** functions to **Non-Smooth** functions

\[
\min_{y \in Y} \left\{ \sum_{k=1}^{K} [\phi_k(A_k y) + \Psi_k(A_k y)] \right\}
\]

**Situation:**  
\( \phi_k \) are given by saddle point representations:  
\[
\phi_k(y^k) = \max_{z^k \in Z_k} \{ \Phi_k(y^k, z^k) \}
\]
where \( \Phi_k(\cdot, \cdot) \) are smooth convex-concave.

**Example:**  
\[
\phi(y) = \max_{1 \leq i \leq m} \sigma_i(A y - b) = \max_{z : \|z\|_{\text{nuc}} \leq 1} \langle A y - b, z \rangle
\]

Total # of MP steps for a \( \epsilon \)-solution is at most \( O\left(\frac{1}{\epsilon}\right) \)!
Extension II:
— from **Full-Prox** setups to **Prox-Free** setups

\[
\min_{y \in Y} \left\{ \sum_{k=1}^{K} \left[ \phi_k(A_k y) + \Psi_k(A_k y) \right] \right\}
\]

**Situation:**
- Prox-mapping for \( (\Psi_k(\cdot), Y_k) \) is expensive;
  \[ \xi^k \mapsto \arg\min_{y^k \in Y_k} \left[ \omega_k(y^k) + \langle \xi^k, y^k \rangle + \alpha \Psi_k(y^k) \right] ; \]
- But Linear Minimization Oracle (LMO) is cheap;

**Example:** (nuclear norm) **full SVD** vs. **leading singular vector**

**Strategy:** mimic prox-mapping by *conditional gradient* method

Total # of Linear Minimization Oracle calls is at most \( \mathcal{O}(\frac{1}{\epsilon^2}) \)!
Extension III:

from Separable structures to Semi-Separable structures

\[
\begin{aligned}
\min_{[y^1; \ldots; y^K] \in Y_1 \times \cdots \times Y_K} & \left\{ \sum_{k=1}^{K} \left[ \phi_k(y^k) + \Psi_k(y^k) \right] \right\} \\
\text{s.t.} & \sum_{k=1}^{K} A_k y^k = b
\end{aligned}
\]

Situation:

- **Direct ADMM**
  
  😞 may not converge [Y. Ye et.al., 2013]

- **Variable Splitting ADMM** [S. Ma et.al., 2013]
  
  😞 need to add lots of variables and constraints

- **Proximal Jacobian ADMM** [W. Yin et.al., 2013]
  
  😞 could be difficult to solve subproblems

Strategy:

apply CMP to a sequence of saddle point problems

\[
\begin{aligned}
\min_{x_1 \in X_1} & \max_{w \in W} \left\{ \alpha_s \sum_{k=1}^{K} \left[ \phi_k(y^k) + \tau^k \right] + \left( 1 - \alpha_s \right) \langle \sum_{k=1}^{K} A_k y^k - b, w \rangle \right\}
\end{aligned}
\]

Total # of MP steps for a \( \epsilon \)-solution is at most \( \mathcal{O}\left( \frac{1}{\epsilon} \log\left( \frac{1}{\epsilon} \right) \right) \).
Key Takeaways

- Strategies for Multi-Term Composite Minimization:
  - Saddle-point-based preprocessing;
  - Composite Mirror Prox algorithm;
  - $O(1/t)$ rate of convergence;

- Always utilize the underlying structures of a problem:
  - Composite structure of variational inequalities;
  - Saddle point representations;
  - Proximal setups for different domains;
  - ....

Thanks! Questions?