Faster Lagrangian-Based Methods in Convex Optimization

Shoham Sabach

Faculty of Industrial Engineering and Management

Technion - Israel Institute of Technology

Joint work with

Marc Teboulle (Tel Aviv University)

One World Optimization Seminar
April 4, 2022

The Linearly Constrained Convex Optimization Model

We focus on the **linearly constrained** convex optimization problem defined by

$$(\mathrm{P}) \qquad \min_{x \in \mathbb{R}^n} \left\{ \Psi\left(x\right) : \; \mathcal{A}x = b \right\},$$

where

- $\Psi : \mathbb{R}^n \to (-\infty, +\infty]$ is proper, lsc and σ-strongly convex with $\sigma \geq 0$.
- $\mathcal{A} : \mathbb{R}^n \to \mathbb{R}^m$ is a linear mapping.
- $b \in \mathbb{R}^m$.
- The feasible set of problem (P) is denoted by $\mathcal{F} = \left\{ x \in \mathbb{R}^n : \; \mathcal{A}x = b \right\} \neq \emptyset$.

The Linearly Constrained Convex Optimization Model

We focus on the **linearly constrained** convex optimization problem defined by

$$\text{(P)} \qquad \min_{x \in \mathbb{R}^n} \left\{ \Psi\left(x\right) : \; \mathcal{A}x = b \right\},$$

where

- $\Psi : \mathbb{R}^n \to (-\infty, +\infty]$ is proper, lsc and σ-strongly convex with $\sigma \geq 0$.
- $\mathcal{A} : \mathbb{R}^n \to \mathbb{R}^m$ is a linear mapping.
- $b \in \mathbb{R}^m$.
- The feasible set of problem (P) is denoted by $\mathcal{F} = \{x \in \mathbb{R}^n : \; \mathcal{A}x = b\} \neq \emptyset$.

Main goal. To unify, simplify, and improve the convergence rate analysis of Lagrangian-based methods for solving model (P).

The Linearly Constrained Convex Optimization Model

We focus on the **linearly constrained** convex optimization problem defined by

$$(\text{P}) \qquad \min_{x \in \mathbb{R}^n} \left\{ \Psi\left(x\right) : \; \mathcal{A}x = b \right\},$$

where

- $\Psi : \mathbb{R}^n \rightarrow (-\infty, +\infty]$ is proper, lsc and σ-strongly convex with $\sigma \geq 0$.
- $\mathcal{A} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a linear mapping.
- $b \in \mathbb{R}^m$.
- The feasible set of problem (P) is denoted by $\mathcal{F} = \{x \in \mathbb{R}^n : \; \mathcal{A}x = b\} \neq \emptyset$.

Main goal. To unify, simplify, and improve the convergence rate analysis of Lagrangian-based methods for solving model (P).

Main contribution. A framework of Faster LAGrangian (FLAG) methods with new non-ergodic rate of convergence results!

Important Particular Instances of Model (P)

Linear composite model

$$\min_{u\in\mathbb{R}^p}\{f(u)+g(Au)\}=\min_{u\in\mathbb{R}^p,v\in\mathbb{R}^q}\{f(u)+g(v):\ Au=v\},$$

where $f:\mathbb{R}^p\rightarrow(-\infty,+\infty]$ and $g:\mathbb{R}^q\rightarrow(-\infty,+\infty]$ are proper, lower semi-continuous and convex functions, and $A:\mathbb{R}^p\rightarrow\mathbb{R}^q$ is a linear mapping.

Important Particular Instances of Model (P)

Linear composite model

$$\min_{u\in\mathbb{R}^p}\left\{f(u)+g(Au)\right\}=\min_{u\in\mathbb{R}^p,v\in\mathbb{R}^q}\left\{f(u)+g(v):\ Au=v\right\},$$

where $f:\mathbb{R}^p\to(-\infty,+\infty]$ and $g:\mathbb{R}^q\to(-\infty,+\infty]$ are proper, lower semi-continuous and convex functions, and $A:\mathbb{R}^p\to\mathbb{R}^q$ is a linear mapping.

Block linear constrained model

$$\min_{u\in\mathbb{R}^p,v\in\mathbb{R}^q}\left\{f(u)+g(v):\ Au+Bv=b\right\},$$

where $f:\mathbb{R}^p\to(-\infty,+\infty]$ and $g:\mathbb{R}^q\to(-\infty,+\infty]$ are proper, lower semi-continuous and convex functions, $A:\mathbb{R}^p\to\mathbb{R}^m$ and $B:\mathbb{R}^q\to\mathbb{R}^m$ are linear mappings. **It fits into model (P)**, with $x=\left(u^T,v^T\right)^T$, $\Psi(x):=f(u)+g(v)$ and $\mathcal{A}x=Au+Bv$.

Important Particular Instances of Model (P)

Linear composite model

$$\min_{u \in \mathbb{R}^p} \{f(u) + g(Au)\} = \min_{u \in \mathbb{R}^p, v \in \mathbb{R}^q} \{f(u) + g(v) : \; Au = v\},$$

where $f : \mathbb{R}^p \to (-\infty, +\infty]$ and $g : \mathbb{R}^q \to (-\infty, +\infty]$ are proper, lower semi-continuous and convex functions, and $A : \mathbb{R}^p \to \mathbb{R}^q$ is a linear mapping.

Block linear constrained model

$$\min_{u \in \mathbb{R}^p, v \in \mathbb{R}^q} \{f(u) + g(v) : \; Au + Bv = b\},$$

where $f : \mathbb{R}^p \to (-\infty, +\infty]$ and $g : \mathbb{R}^q \to (-\infty, +\infty]$ are proper, lower semi-continuous and convex functions, $A : \mathbb{R}^p \to \mathbb{R}^m$ and $B : \mathbb{R}^q \to \mathbb{R}^m$ are linear mappings. **It fits into model (P)**, with $x = \left(u^T, v^T\right)^T$, $\Psi(x) := f(u) + g(v)$ and $\mathcal{A}x = Au + Bv$.

Additive smooth/non-smooth composite objective

$$\min_{x \in \mathbb{R}^n} \{f(x) + h(x) : \; \mathcal{A}x = b\},$$

where $f : \mathbb{R}^n \to (-\infty, +\infty]$ is a proper, lsc and convex function, while $h : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable function with a Lipschitz continuous gradient.

Preliminaries on the Convex Model (P)

We recall problem (P)

$$(P) \qquad \min_{x \in \mathbb{R}^n} \left\{ \Psi(x) : \mathcal{A}x = b \right\},$$

The corresponding Lagrangian and augmented Lagrangian, are given by:

$$\mathcal{L}(x, y) = \Psi(x) + \langle y, \mathcal{A}x - b \rangle,$$

and, for any $\rho > 0$,

$$\mathcal{L}_\rho(x, y) = \mathcal{L}(x, y) + \frac{\rho}{2} \|\mathcal{A}x - b\|^2.$$

Preliminaries on the Convex Model (P)

We recall problem (P)

$$\text{(P)} \qquad \min_{x \in \mathbb{R}^n} \left\{ \Psi\left(x\right) : \, \mathcal{A}x = b \right\},$$

The corresponding Lagrangian and augmented Lagrangian, are given by:

$$\mathcal{L}\left(x, y\right) = \Psi\left(x\right) + \left\langle y, \mathcal{A}x - b \right\rangle,$$

and, for any $\rho > 0$,

$$\mathcal{L}_\rho\left(x, y\right) = \mathcal{L}\left(x, y\right) + \frac{\rho}{2}\left\|\mathcal{A}x - b\right\|^2.$$

Assumption

The Lagrangian $\mathcal{L}$ has a saddle point, that is, there exists $\left(x^*, y^*\right)$ such that

$$\mathcal{L}\left(x^*, y\right) \leq \mathcal{L}\left(x^*, y^*\right) \leq \mathcal{L}\left(x, y^*\right), \quad \forall\, x \in \mathbb{R}^n, \quad \forall\, y \in \mathbb{R}^m.$$

It can be warranted, for instance, **under standard CQ on the problem's data**.

Lagrangian-based Methods for Model (P)

The starting point is that **all Lagrangian-based methods** update a couple (x, y) by

$$x^+ \in \mathcal{P}(x, y),$$
$$y^+ = y + \mu\rho\left(\mathcal{A}x^+ - b\right),$$

where $\mu > 0$ is a scaling parameter and $\mathcal{P}$ is a primal algorithmic map.

Lagrangian-based Methods for Model (P)

The starting point is that **all Lagrangian-based methods** update a couple (x, y) by

$$x^+ \in \mathcal{P}(x, y),$$
$$y^+ = y + \mu\rho\left(\mathcal{A}x^+ - b\right),$$

where $\mu > 0$ is a scaling parameter and $\mathcal{P}$ is a primal algorithmic map.

Augmented Lagrangian (Hestenes (69), Powell (69))

1. **Input:** $\mu > 0$.
2. **Initialization:** Start with any $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$.
3. **Main step:** Given (x, y), update the new point $\left(x^+, y^+\right)$ via:

$$x^+ \in \operatorname{argmin}\left\{\mathcal{L}_\rho(\xi, y) : \xi \in \mathbb{R}^n\right\},$$
$$y^+ = y + \mu\rho\left(\mathcal{A}x^+ - b\right).$$

In this case, $\mathcal{P}$ is an **exact minimization applied on the augmented Lagrangian**.

Lagrangian-based Methods for Model (P)

The starting point is that **all Lagrangian-based methods** update a couple (x, y) by

$$x^+ \in \mathcal{P}(x, y),$$
$$y^+ = y + \mu\rho\left(\mathcal{A}x^+ - b\right),$$

where $\mu > 0$ is a scaling parameter and $\mathcal{P}$ is a primal algorithmic map.

Proximal Linearized Augmented Lagrangian

1. **Input:** $M \succeq 0$ and $\mu > 0$.
2. **Initialization:** Start with any $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$.
3. **Main step:** Given (x, y), update the new point $\left(x^+, y^+\right)$ via:

$$x^+ \in \operatorname{argmin}\left\{\Psi(\xi) + \left\langle \xi, \mathcal{A}^T\left(y + \rho\left(\mathcal{A}x - b\right)\right)\right\rangle + \frac{1}{2}\|\xi - x\|_M^2 : \xi \in \mathbb{R}^n\right\},$$
$$y^+ = y + \mu\rho\left(\mathcal{A}x^+ - b\right).$$

In this case, $\mathcal{P}$ is a **proximal gradient applied on the augmented Lagrangian**.

Lagrangian-based Methods for Block Models

As discussed above, Model (P) covers the following block model

$$\min_{(u,v)\in\mathbb{R}^n} \{f(u) + g(v) \, : \, Au + Bv = b\}.$$

Note: Here, we only need to assume that **one** of the functions is strongly convex.

Lagrangian-based Methods for Block Models

As discussed above, Model (P) covers the following block model

$$\min_{(u,v)\in\mathbb{R}^n} \left\{ f(u) + g(v) \,:\, Au + Bv = b \right\}.$$

Note: Here, we only need to assume that **one** of the functions is strongly convex.

The block structure can be exploited in designing Lagrangian-based methods.

Alternating Direction Method of Multipliers (ADMM) (Glowinski and Marroco (75), Gabay and Mercier (76))

1. **Input:** $\mu > 0$.
2. **Initialization:** Start with any $(u, v, y) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$.
3. **Main step:** Given (u, v, y), update the new point $\left(u^+, v^+, y^+\right)$ via:

$$\begin{aligned} u^+ &= \operatorname{argmin}\left\{ \mathcal{L}_\rho(\xi, v, y) \,:\, \xi \in \mathbb{R}^n \right\}, \\ v^+ &= \operatorname{argmin}\left\{ \mathcal{L}_\rho\left(u^+, \eta, y\right) \,:\, \eta \in \mathbb{R}^m \right\}, \\ y^+ &= y + \mu\rho\left(Au^+ + Bv^+ - b\right). \end{aligned}$$

In this case, $\mathcal{P}$ is an **alternating minimization applied on $\mathcal{L}_\rho$**.

Lagrangian-based Methods for Block Models

As discussed above, Model (P) covers the following block model

$$\min_{(u,v)\in\mathbb{R}^n} \left\{ f(u) + g(v) \, : \, Au + Bv = b \right\}.$$

In this case, the augmented Lagrangian is given by

$$\mathcal{L}_\rho(u, v, y) = f(u) + g(v) + \langle y, Au + Bv - b \rangle + \frac{\rho}{2} \|Au + Bv - b\|^2.$$

Proximal Linearized ADMM (Chambolle and Pock (11), He and Yuan (12))

1. **Input:** $M_1, M_2 \succeq 0$ and $\mu > 0$.
2. **Initialization:** Start with any $(u, v, y) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$.
3. **Main step:** Given (u, v, y), update the new point $\left(u^+, v^+, y^+\right)$ via:

$$u^+ = \mathrm{argmin}_\xi \left\{ f(\xi) + \left\langle A^T \left(y + \rho(Au + Bv - b)\right), \xi - u \right\rangle + \frac{1}{2} \|\xi - u\|_{M_1}^2 \right\},$$

$$v^+ = \mathrm{argmin}_\eta \left\{ g(\eta) + \left\langle B^T \left(y + \rho\left(Au^+ + Bv - b\right)\right), \eta - v \right\rangle + \frac{1}{2} \|\eta - v\|_{M_2}^2 \right\},$$

$$y^+ = y + \mu\rho\left(Au^+ + Bv^+ - b\right).$$

In this case, $\mathcal{P}$ is a **alternating proximal gradient applied on $\mathcal{L}_\rho$**.

Nice Primal Algorithmic Map

Adopting the following **point of view of Lagrangian-based methods**:

$$
\begin{aligned}
x^+ &\in \mathcal{P}(x, y), \\
y^+ &= y + \mu\rho\left(\mathcal{A}x^+ - b\right).
\end{aligned}
$$

Nice Primal Algorithmic Map

Adopting the following **point of view of Lagrangian-based methods**:

$$x^{+} \in \mathcal{P}\left(x, y\right),$$
$$y^{+} = y + \mu\rho\left(\mathcal{A}x^{+} - b\right).$$

Definition (Nice primal algorithmic map)

Given $\rho, t > 0$. A primal algorithmic map $\mathrm{Prim}_t : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$ that generates $z^{+} \in \mathrm{Prim}_t\left(z, \lambda\right)$, is called **nice**, **if there exist** $\delta \in (0, 1]$ **and** $P, Q \succeq 0$,

Nice Primal Algorithmic Map

Adopting the following **point of view of Lagrangian-based methods**:

$$x^+ \in \mathcal{P}(x, y),$$
$$y^+ = y + \mu\rho\left(\mathcal{A}x^+ - b\right).$$

Definition (Nice primal algorithmic map)

Given $\rho, t > 0$. A primal algorithmic map $\text{Prim}_t : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ that generates $z^+ \in \text{Prim}_t(z, \lambda)$, is called **nice**, **if there exist** $\delta \in (0, 1]$ **and** $P, Q \succeq 0$, such that for any $\xi \in \mathcal{F}$ we have

$$\mathcal{L}_{\rho_t}\left(z^+, \lambda\right) - \mathcal{L}_{\rho_t}(\xi, \lambda) \leq \tau_t \Delta_P\left(\xi, z, z^+\right) - \frac{\tau_t}{2}\left\|z^+ - z\right\|_Q^2 - \frac{\sigma}{2}\left\|\xi - z^+\right\|^2 - \frac{\delta\rho_t}{2}\left\|\mathcal{A}z^+ - b\right\|^2$$

where $\rho_t = \rho$ and $\tau_t = 1$ (when $\sigma = 0$) or $\rho_t = \rho t$ and $\tau_t = t$ (when $\sigma > 0$).

Nice Primal Algorithmic Map

Adopting the following **point of view of Lagrangian-based methods**:

$$x^+ \in \mathcal{P}(x, y),$$
$$y^+ = y + \mu\rho\left(\mathcal{A}x^+ - b\right).$$

Definition (Nice primal algorithmic map)

Given $\rho, t > 0$. A primal algorithmic map $\text{Prim}_t : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ that generates $z^+ \in \text{Prim}_t(z, \lambda)$, is called **nice**, **if there exist $\delta \in (0, 1]$ and $P, Q \succeq 0$**, such that for any $\xi \in \mathcal{F}$ we have

$$\mathcal{L}_{\rho_t}\left(z^+, \lambda\right) - \mathcal{L}_{\rho_t}(\xi, \lambda) \leq \tau_t \Delta_P\left(\xi, z, z^+\right) - \frac{\tau_t}{2}\left\|z^+ - z\right\|_Q^2 - \frac{\sigma}{2}\left\|\xi - z^+\right\|^2 - \frac{\delta\rho_t}{2}\left\|\mathcal{A}z^+ - b\right\|^2$$

where $\rho_t = \rho$ and $\tau_t = 1$ (when $\sigma = 0$) or $\rho_t = \rho t$ and $\tau_t = t$ (when $\sigma > 0$).

Notations.

- For any matrix $P \succeq 0$ and any three vectors $u, v, w \in \mathbb{R}^n$:

$$\Delta_P(u, v, w) := \frac{1}{2}\|u - v\|_P^2 - \frac{1}{2}\|u - w\|_P^2.$$

- When $P \equiv I_n$, the identity matrix, we simply write $\Delta_P(u, v, w) \equiv \Delta(u, v, w)$.

Examples of Nice Primal Algorithmic Maps

Lemma (Proximal augmented Lagrangian is nice)

Let $M \succeq 0$, then the algorithmic map defined by

$$z^{+} = \mathrm{Prim}_t\left(z, \lambda\right) \equiv \operatorname{argmin}_{\xi}\left\{\Psi\left(\xi\right) + \langle \lambda, \mathcal{A}\xi - b \rangle + \frac{\rho_t}{2}\left\|\mathcal{A}\xi - b\right\|^2 + \frac{\tau_t}{2}\left\|\xi - z\right\|_M^2\right\},$$

is nice with $\delta = 1$ and $P = Q = M$.

Examples of Nice Primal Algorithmic Maps

Lemma (Proximal augmented Lagrangian is nice)

Let $M \succeq 0$, then the algorithmic map defined by

$$z^+ = \text{Prim}_t\left(z, \lambda\right) \equiv \operatorname{argmin}_\xi \left\{ \Psi\left(\xi\right) + \langle \lambda, \mathcal{A}\xi - b \rangle + \frac{\rho_t}{2} \left\| \mathcal{A}\xi - b \right\|^2 + \frac{\tau_t}{2} \left\| \xi - z \right\|_M^2 \right\},$$

is nice with $\delta = 1$ and $P = Q = M$.

Lemma (Proximal linearized AL is nice)

Let $M \succeq \rho \mathcal{A}^T \mathcal{A}$, then the algorithmic map defined by

$$z^+ = \text{Prim}_t\left(z, \lambda\right) \equiv \operatorname{argmin}_\xi \left\{ \Psi\left(\xi\right) + \langle \lambda, \mathcal{A}\xi - b \rangle + \rho_t \langle \mathcal{A}z - b, \mathcal{A}\xi \rangle + \frac{\tau_t}{2} \left\| \xi - z \right\|_M^2 \right\},$$

is nice with $\delta = 1$ and $P = Q = M - \rho \mathcal{A}^T \mathcal{A} \succeq 0$.

Examples of Nice Primal Algorithmic Maps

Lemma (Proximal augmented Lagrangian is nice)

Let $M \succeq 0$, then the algorithmic map defined by

$$z^+ = \mathrm{Prim}_t(z, \lambda) \equiv \operatorname{argmin}_\xi \left\{ \Psi(\xi) + \langle \lambda, \mathcal{A}\xi - b \rangle + \frac{\rho_t}{2} \|\mathcal{A}\xi - b\|^2 + \frac{\tau_t}{2} \|\xi - z\|_M^2 \right\},$$

is nice with $\delta = 1$ and $P = Q = M$.

Lemma (Proximal linearized AL is nice)

Let $M \succeq \rho \mathcal{A}^T \mathcal{A}$, then the algorithmic map defined by

$$z^+ = \mathrm{Prim}_t(z, \lambda) \equiv \operatorname{argmin}_\xi \left\{ \Psi(\xi) + \langle \lambda, \mathcal{A}\xi - b \rangle + \rho_t \langle \mathcal{A}z - b, \mathcal{A}\xi \rangle + \frac{\tau_t}{2} \|\xi - z\|_M^2 \right\},$$

is nice with $\delta = 1$ and $P = Q = M - \rho \mathcal{A}^T \mathcal{A} \succeq 0$.

All well-known Lagrangian-based methods admit a nice primal algorithmic maps!

Examples of Nice Primal Algorithmic Maps

Lemma (Proximal augmented Lagrangian is nice)

Let $M \succeq 0$, then the algorithmic map defined by

$$z^{+} = \mathrm{Prim}_t\left(z, \lambda\right) \equiv \operatorname{argmin}_{\xi}\left\{\Psi\left(\xi\right) + \langle \lambda, \mathcal{A}\xi - b\rangle + \frac{\rho_t}{2}\left\|\mathcal{A}\xi - b\right\|^2 + \frac{\tau_t}{2}\left\|\xi - z\right\|_M^2\right\},$$

is nice with $\delta = 1$ and $P = Q = M$.

Lemma (Proximal linearized AL is nice)

Let $M \succeq \rho\mathcal{A}^T\mathcal{A}$, then the algorithmic map defined by

$$z^{+} = \mathrm{Prim}_t\left(z, \lambda\right) \equiv \operatorname{argmin}_{\xi}\left\{\Psi\left(\xi\right) + \langle \lambda, \mathcal{A}\xi - b\rangle + \rho_t\left\langle \mathcal{A}z - b, \mathcal{A}\xi\right\rangle + \frac{\tau_t}{2}\left\|\xi - z\right\|_M^2\right\},$$

is nice with $\delta = 1$ and $P = Q = M - \rho\mathcal{A}^T\mathcal{A} \succeq 0$.

All well-known Lagrangian-based methods admit a nice primal algorithmic maps!

More examples later.

A Unified Framework

FLAG – Faster LAGrangian based method

1. **Input: Problem data** $[\Psi, \mathcal{A}, b, \sigma]$, and a **nice primal algorithmic map** $\mathrm{Prim}_t(\cdot)$.
2. **Initialization:** Set $t_0 = 1$, $\mu \in (0, \delta]$ and $\rho > 0$. Start with any $\left(x^0, z^0, y^0\right)$.
3. **Iterations:** Generate $\left\{\left(x^k, z^k, y^k\right)\right\}_{k \in \mathbb{N}}$ and $\{t_k\}_{k \in \mathbb{N}}$ via

A Unified Framework

FLAG – Faster LAGrangian based method

1. **Input: Problem data** $[\Psi, \mathcal{A}, b, \sigma]$, and a **nice primal algorithmic map** $\text{Prim}_t(\cdot)$.
2. **Initialization:** Set $t_0 = 1$, $\mu \in (0, \delta]$ and $\rho > 0$. Start with any $\left(x^0, z^0, y^0\right)$.
3. **Iterations:** Generate $\left\{\left(x^k, z^k, y^k\right)\right\}_{k\in\mathbb{N}}$ and $\{t_k\}_{k\in\mathbb{N}}$ via

   3.1. If $\sigma = 0$, let $\rho_k = \rho$, or, if $\sigma > 0$, let $\rho_k = \rho t_k$. Compute

$$\lambda^k = y^k + \rho_k (t_k - 1)\left(\mathcal{A}x^k - b\right).$$

A Unified Framework

FLAG – Faster LAGrangian based method

1. **Input: Problem data** $[\Psi, \mathcal{A}, b, \sigma]$, and a **nice primal algorithmic map** $\mathrm{Prim}_t(\cdot)$.
2. **Initialization:** Set $t_0 = 1$, $\mu \in (0, \delta]$ and $\rho > 0$. Start with any $\left(x^0, z^0, y^0\right)$.
3. **Iterations:** Generate $\left\{\left(x^k, z^k, y^k\right)\right\}_{k \in \mathbb{N}}$ and $\{t_k\}_{k \in \mathbb{N}}$ via

   3.1. If $\sigma = 0$, let $\rho_k = \rho$, or, if $\sigma > 0$, let $\rho_k = \rho t_k$. Compute

   $$\lambda^k = y^k + \rho_k (t_k - 1)\left(\mathcal{A}x^k - b\right).$$

   3.2. Update the sequence $\left\{\left(x^k, z^k, y^k\right)\right\}_{k \in \mathbb{N}}$ by

   $$z^{k+1} \in \mathrm{Prim}_k\left(z^k, \lambda^k\right),$$

   $$y^{k+1} = y^k + \mu\rho_k\left(\mathcal{A}z^{k+1} - b\right),$$

   $$x^{k+1} = \left(1 - t_k^{-1}\right)x^k + t_k^{-1}z^{k+1}.$$

A Unified Framework

FLAG – Faster LAGrangian based method

1. **Input: Problem data** $[\Psi, \mathcal{A}, b, \sigma]$, and a **nice primal algorithmic map** $\text{Prim}_t(\cdot)$.
2. **Initialization:** Set $t_0 = 1$, $\mu \in (0, \delta]$ and $\rho > 0$. Start with any $\left(x^0, z^0, y^0\right)$.
3. **Iterations:** Generate $\left\{\left(x^k, z^k, y^k\right)\right\}_{k \in \mathbb{N}}$ and $\{t_k\}_{k \in \mathbb{N}}$ via
   3.1. If $\sigma = 0$, let $\rho_k = \rho$, or, if $\sigma > 0$, let $\rho_k = \rho t_k$. Compute

   $$\lambda^k = y^k + \rho_k (t_k - 1)\left(\mathcal{A}x^k - b\right).$$

   3.2. Update the sequence $\left\{\left(x^k, z^k, y^k\right)\right\}_{k \in \mathbb{N}}$ by

   $$z^{k+1} \in \text{Prim}_k\left(z^k, \lambda^k\right),$$

   $$y^{k+1} = y^k + \mu \rho_k \left(\mathcal{A}z^{k+1} - b\right),$$

   $$x^{k+1} = \left(1 - t_k^{-1}\right) x^k + t_k^{-1} z^{k+1}.$$ **The acceleration step!**

A Unified Framework

FLAG – **F**aster **LAG**rangian based method

1. **Input: Problem data** $[\Psi, \mathcal{A}, b, \sigma]$, and a **nice primal algorithmic map** $\text{Prim}_t(\cdot)$.
2. **Initialization:** Set $t_0 = 1$, $\mu \in (0, \delta]$ and $\rho > 0$. Start with any $\left(x^0, z^0, y^0\right)$.
3. **Iterations:** Generate $\left\{\left(x^k, z^k, y^k\right)\right\}_{k \in \mathbb{N}}$ and $\{t_k\}_{k \in \mathbb{N}}$ via
   3.1. If $\sigma = 0$, let $\rho_k = \rho$, or, if $\sigma > 0$, let $\rho_k = \rho t_k$. Compute
   $$\lambda^k = y^k + \rho_k (t_k - 1)\left(\mathcal{A}x^k - b\right).$$
   3.2. Update the sequence $\left\{\left(x^k, z^k, y^k\right)\right\}_{k \in \mathbb{N}}$ by
   $$z^{k+1} \in \text{Prim}_k\left(z^k, \lambda^k\right),$$
   $$y^{k+1} = y^k + \mu \rho_k \left(\mathcal{A}z^{k+1} - b\right),$$
   $$x^{k+1} = \left(1 - t_k^{-1}\right) x^k + t_k^{-1} z^{k+1}.$$ **The acceleration step!**
   3.3. Update the sequence $\{t_k\}_{k \in \mathbb{N}}$ by solving the equation $t_{k+1}^p - t_k^p = t_{k+1}^{p-1}$, *i.e.*,
   $$t_{k+1} = \begin{cases} t_k + 1, & p = 1 \text{ (convex case)}, \\ \left(1 + \sqrt{1 + 4t_k^2}\right)/2, & p = 2 \text{ (strongly convex case)}. \end{cases}$$

FLAG - A Few Comments

- Setting $t_k \equiv 1$ in FLAG, implies $\rho_k \equiv \rho$, $\lambda^k \equiv y^k$, and $x^k \equiv z^k$, thus **recovering the classical basic Lagrangian-based methods**.

FLAG - A Few Comments

- Setting $t_k \equiv 1$ in FLAG, implies $\rho_k \equiv \rho$, $\lambda^k \equiv y^k$, and $x^k \equiv z^k$, thus **recovering the classical basic Lagrangian-based methods**.
- A main new feature of FLAG is the **auxiliary variable λ^k** defined by:

$$\lambda^k = y^k + \rho_k (t_k - 1) \left(\mathcal{A}x^k - b \right),$$

which enable us to derive the new faster **non-ergodic** rate of convergence results!

FLAG - A Few Comments

- Setting $t_k \equiv 1$ in FLAG, implies $\rho_k \equiv \rho$, $\lambda^k \equiv y^k$, and $x^k \equiv z^k$, thus **recovering the classical basic Lagrangian-based methods**.
- A main new feature of FLAG is the **auxiliary variable** λ^k defined by:

$$\lambda^k = y^k + \rho_k (t_k - 1) \left(\mathcal{A}x^k - b \right),$$

  which enable us to derive the new faster **non-ergodic** rate of convergence results!

- As we shall see, when λ^k **coincides with** y^k, only **ergodic type** rates (classical and fast) can be obtained.

FLAG - A Few Comments

- Setting $t_k \equiv 1$ in FLAG, implies $\rho_k \equiv \rho$, $\lambda^k \equiv y^k$, and $x^k \equiv z^k$, thus **recovering the classical basic Lagrangian-based methods**.

- A main new feature of FLAG is the **auxiliary variable λ^k** defined by:

$$\lambda^k = y^k + \rho_k (t_k - 1) \left(\mathcal{A} x^k - b \right),$$

  which enable us to derive the new faster **non-ergodic** rate of convergence results!

- As we shall see, when λ^k **coincides with y^k**, only **ergodic type** rates (classical and fast) can be obtained.

- The augmented parameter ρ_k and the prox parameter τ_k defined via t_k:

$$\begin{cases} t_{k+1} = t_k + 1 & \rho_k = \rho \quad \tau_k = 1, \quad (\textbf{convex case}), \\ t_{k+1} = \frac{1}{2}\left(1 + \sqrt{1 + 4t_k^2}\right) & \rho_k = \rho t_k \quad \tau_k = t_k, \quad (\textbf{strongly convex case}). \end{cases}$$

FLAG - A Few Comments

- Setting $t_k \equiv 1$ in FLAG, implies $\rho_k \equiv \rho$, $\lambda^k \equiv y^k$, and $x^k \equiv z^k$, thus **recovering the classical basic Lagrangian-based methods**.
- A main new feature of FLAG is the **auxiliary variable** λ^k defined by:

$$\lambda^k = y^k + \rho_k (t_k - 1) \left(\mathcal{A}x^k - b \right),$$

which enable us to derive the new faster **non-ergodic** rate of convergence results!

- As we shall see, when λ^k **coincides with** y^k, only **ergodic type** rates (classical and fast) can be obtained.
- The augmented parameter ρ_k and the prox parameter τ_k defined via t_k:

$$\begin{cases} t_{k+1} = t_k + 1 & \rho_k = \rho \quad \tau_k = 1, \quad (\textbf{convex case}), \\ t_{k+1} = \frac{1}{2}\left(1 + \sqrt{1 + 4t_k^2}\right) & \rho_k = \rho t_k \quad \tau_k = t_k, \quad (\textbf{strongly convex case}). \end{cases}$$

- Prim_t is assumed to be **nice primal algorithmic map** and this is **all we need** to guarantee rate of convergence results (classical and fast)!

Types of Rate of Convergence Results

We focus on iteration complexity using the following two classical measures:

(i) **Function values gap** in terms of $\Psi\left(x^k\right) - \Psi\left(x^*\right)$.

(ii) **Feasibility violation** of the constraints of problem (P) in terms of $\left\|\mathcal{A}x^k - b\right\|$.

Types of Rate of Convergence Results

We focus on iteration complexity using the following two classical measures:

(i) **Function values gap** in terms of $\Psi\left(x^k\right) - \Psi\left(x^*\right)$.

(ii) **Feasibility violation** of the constraints of problem (P) in terms of $\left\|\mathcal{A}x^k - b\right\|$.

Other measures in the literature: Lagrangian, $\left\|x^k - x^*\right\|^2$, $\left\|x^{k+1} - x^k\right\|^2$, etc.

Types of Rate of Convergence Results

We focus on iteration complexity using the following two classical measures:

(i) **Function values gap** in terms of $\Psi\left(x^k\right) - \Psi\left(x^*\right)$.

(ii) **Feasibility violation** of the constraints of problem (P) in terms of $\left\|\mathcal{A}x^k - b\right\|$.

Other measures in the literature: Lagrangian, $\left\|x^k - x^*\right\|^2$, $\left\|x^{k+1} - x^k\right\|^2$, etc.

When discussing these measures, we also distinguish between rates expressed in terms of the **original produced sequence** or of the **ergodic sequence**.

Types of Rate of Convergence Results

We focus on iteration complexity using the following two classical measures:

(i) **Function values gap** in terms of $\Psi\left(x^k\right) - \Psi\left(x^*\right)$.

(ii) **Feasibility violation** of the constraints of problem (P) in terms of $\left\|\mathcal{A}x^k - b\right\|$.

Other measures in the literature: Lagrangian, $\left\|x^k - x^*\right\|^2$, $\left\|x^{k+1} - x^k\right\|^2$, etc.

When discussing these measures, we also distinguish between rates expressed in terms of the **original produced sequence** or of the **ergodic sequence**.

- Ergodic $O(1/N)$ rate of convergence result for the Linearized ADMM was proven first in (He and Yuan (12)), (Chambolle and Pock (11)) and (Monteiro and Svaiter (13)).

Types of Rate of Convergence Results

We focus on iteration complexity using the following two classical measures:

(i) **Function values gap** in terms of $\Psi\left(x^k\right) - \Psi\left(x^*\right)$.

(ii) **Feasibility violation** of the constraints of problem (P) in terms of $\left\|\mathcal{A}x^k - b\right\|$.

Other measures in the literature: Lagrangian, $\left\|x^k - x^*\right\|^2$, $\left\|x^{k+1} - x^k\right\|^2$, etc.

When discussing these measures, we also distinguish between rates expressed in terms of the **original produced sequence** or of the **ergodic sequence**.

- Ergodic $O(1/N)$ rate of convergence result for the Linearized ADMM was proven first in (He and Yuan (12)), (Chambolle and Pock (11)) and (Monteiro and Svaiter (13)).
- Many more rate of convergence results in the literature!

Types of Rate of Convergence Results

We focus on iteration complexity using the following two classical measures:

(i) **Function values gap** in terms of $\Psi\left(x^k\right) - \Psi\left(x^*\right)$.

(ii) **Feasibility violation** of the constraints of problem (P) in terms of $\left\|\mathcal{A}x^k - b\right\|$.

Other measures in the literature: Lagrangian, $\left\|x^k - x^*\right\|^2$, $\left\|x^{k+1} - x^k\right\|^2$, etc.

When discussing these measures, we also distinguish between rates expressed in terms of the **original produced sequence** or of the **ergodic sequence**.

- Ergodic $O(1/N)$ rate of convergence result for the Linearized ADMM was proven first in (He and Yuan (12)), (Chambolle and Pock (11)) and (Monteiro and Svaiter (13)).
- Many more rate of convergence results in the literature! **Mostly ergodic**.

Types of Rate of Convergence Results

We focus on iteration complexity using the following two classical measures:

(i) **Function values gap** in terms of $\Psi\left(x^k\right) - \Psi\left(x^*\right)$.

(ii) **Feasibility violation** of the constraints of problem (P) in terms of $\left\|\mathcal{A}x^k - b\right\|$.

Other measures in the literature: Lagrangian, $\left\|x^k - x^*\right\|^2$, $\left\|x^{k+1} - x^k\right\|^2$, etc.

When discussing these measures, we also distinguish between rates expressed in terms of the **original produced sequence** or of the **ergodic sequence**.

- Ergodic $O(1/N)$ rate of convergence result for the Linearized ADMM was proven first in (He and Yuan (12)), (Chambolle and Pock (11)) and (Monteiro and Svaiter (13)).
- Many more rate of convergence results in the literature! **Mostly ergodic**.
- Non-ergodic $O(1/N)$ result for $\left\|x^{k+1} - x^k\right\|^2$ (He and Yuan (15)).
- Non-ergodic $O(1/N^2)$ result for $\left\|x^k - x^*\right\|^2$, in the strongly convex setting (Chambolle and Pock (11)).

Types of Rate of Convergence Results

We focus on iteration complexity using the following two classical measures:

(i) **Function values gap** in terms of $\Psi\left(x^k\right) - \Psi\left(x^*\right)$.

(ii) **Feasibility violation** of the constraints of problem (P) in terms of $\left\|\mathcal{A}x^k - b\right\|$.

Other measures in the literature: Lagrangian, $\left\|x^k - x^*\right\|^2$, $\left\|x^{k+1} - x^k\right\|^2$, etc.

When discussing these measures, we also distinguish between rates expressed in terms of the **original produced sequence** or of the **ergodic sequence**.

- Ergodic $O(1/N)$ rate of convergence result for the Linearized ADMM was proven first in (He and Yuan (12)), (Chambolle and Pock (11)) and (Monteiro and Svaiter (13)).
- Many more rate of convergence results in the literature! **Mostly ergodic**.
- Non-ergodic $O(1/N)$ result for $\left\|x^{k+1} - x^k\right\|^2$ (He and Yuan (15)).
- Non-ergodic $O(1/N^2)$ result for $\left\|x^k - x^*\right\|^2$, in the strongly convex setting (Chambolle and Pock (11)).
- Non-ergodic $O(1/N)$ rate of convergence result in terms of function values and feasibility violation for the **specific Linearized ADMM** (Li and Lin (19)).

Main Results - Non-Ergodic Rate $O(1/N^2)$

In the results below, we let $c > 0$ be a constant such that $c \geq 2\|y^*\|$, where y^* is an optimal solution of the dual problem.

The strongly convex case $\sigma > 0$.

Theorem 1. (A fast non-ergodic function values and feasibility violation rates)

Let $\{(x^k, z^k, y^k)\}_{k \in \mathbb{N}}$ be a sequence generated by FLAG. **Suppose that $\sigma > 0$ and $0 \preceq P \preceq (\sigma/2) I_n$.** Then, for any optimal solution x^* of problem (P), we have

$$\Psi\left(x^N\right) - \Psi\left(x^*\right) \leq \frac{B_{\rho,c}\left(x^*\right)}{2N^2} \qquad \text{and} \qquad \left\|\mathcal{A}x^N - b\right\| \leq \frac{B_{\rho,c}\left(x^*\right)}{cN^2},$$

where $B_{\rho,c}\left(x^*\right) := 4\left(\left\|x^* - z^0\right\|_P^2 + \frac{1}{\mu\rho}\left(\left\|y^0\right\| + c\right)^2\right)$.

Main Results - Non-Ergodic Rate $O(1/N)$

The convex case $\sigma = 0$.

Theorem 2. (A non-ergodic function values and feasibility violation rates)

Let $\left\{\left(x^k, z^k, y^k\right)\right\}_{k\in\mathbb{N}}$ be a sequence generated by FLAG and **suppose that $\sigma = 0$**. Then, for any optimal solution x^* of problem (P), we have

$$\Psi\left(x^N\right) - \Psi\left(x^*\right) \leq \frac{B_{\rho,c}\left(x^*\right)}{2N} \qquad \text{and} \qquad \left\|\mathcal{A}x^N - b\right\| \leq \frac{B_{\rho,c}\left(x^*\right)}{cN},$$

where $B_{\rho,c}\left(x^*\right) := 2\left(\left\|x^* - z^0\right\|_P^2 + \frac{1}{\mu\rho}\left(\left\|y^0\right\| + c\right)^2\right)$.

Ergodic Version

FLAG

1. **Input: Problem data** $[\Psi, \mathcal{A}, b, \sigma]$, and a **nice primal algorithmic map** $\text{Prim}_t(\cdot)$.
2. **Initialization:** Set $t_0 = 1$, $\mu \in (0, 1 + \delta]$ and $\rho > 0$. Start with any (z^0, y^0).
3. **Iterations:** Generate $\{(z^k, y^k)\}_{k\in\mathbb{N}}$ and $\{t_k\}_{k\in\mathbb{N}}$ via

   3.1. Update the sequence $\{(z^k, y^k)\}_{k\in\mathbb{N}}$ by

$$z^{k+1} \in \text{Prim}_k\left(z^k, \mathbf{y^k}\right),$$

$$y^{k+1} = y^k + \mu\rho_k\left(\mathcal{A}z^{k+1} - b\right),$$

   where $\rho_k = \rho$ (if $\sigma = 0$), or $\rho_k = \rho t_k$ (if $\sigma > 0$).

   3.2. Update the sequence $\{t_k\}_{k\in\mathbb{N}}$ by

$$t_{k+1} = \begin{cases} 1, & (\textbf{convex case}), \\ \left(1 + \sqrt{1 + 4t_k^2}\right)/2, & (\textbf{strongly convex case}). \end{cases}$$

Ergodic Version

FLAG

1. **Input: Problem data** $[\Psi, \mathcal{A}, b, \sigma]$, and a **nice primal algorithmic map** $\text{Prim}_t(\cdot)$.
2. **Initialization:** Set $t_0 = 1$, $\mu \in (0, 1+\delta]$ and $\rho > 0$. Start with any (z^0, y^0).
3. **Iterations:** Generate $\{(z^k, y^k)\}_{k\in\mathbb{N}}$ and $\{t_k\}_{k\in\mathbb{N}}$ via

   3.1. Update the sequence $\{(z^k, y^k)\}_{k\in\mathbb{N}}$ by

   $$z^{k+1} \in \text{Prim}_k\left(z^k, \mathbf{y^k}\right),$$
   $$y^{k+1} = y^k + \mu\rho_k\left(\mathcal{A}z^{k+1} - b\right),$$

   where $\rho_k = \rho$ (if $\sigma = 0$), or $\rho_k = \rho t_k$ (if $\sigma > 0$).

   3.2. Update the sequence $\{t_k\}_{k\in\mathbb{N}}$ by

   $$t_{k+1} = \begin{cases} 1, & (\textbf{convex case}), \\ \left(1 + \sqrt{1 + 4t_k^2}\right)/2, & (\textbf{strongly convex case}). \end{cases}$$

In the ergodic version, the sequences $\{x^k\}_{k\in\mathbb{N}}$ and $\{\lambda^k\}_{k\in\mathbb{N}}$ are not used!

Ergodic Rate of Convergence Results

Corollary (A fast ergodic function values and feasibility violation rates)

Let $\left\{\left(z^k, y^k\right)\right\}_{k\in\mathbb{N}}$ *be a sequence generated by FLAG.* **Suppose that** $\sigma > 0$ **and** $0 \preceq P \preceq (\sigma/2)\, I_n$. *Then, for any optimal solution* z^* *of problem (P), the following holds for the ergodic sequence* $\bar{z}^N = t_{N-1}^{-2} \sum_{k=0}^{N-1} t_k z^{k+1}$

$$\Psi\left(\bar{z}^N\right) - \Psi\left(z^*\right) \leq \frac{B_{\rho,c}\left(z^*\right)}{2N^2} \qquad \text{and} \qquad \left\|\mathcal{A}\bar{z}^N - b\right\| \leq \frac{B_{\rho,c}\left(z^*\right)}{cN^2},$$

where $B_{\rho,c}\left(z^*\right) := 4\left(\left\|z^* - z^0\right\|_P^2 + \frac{1}{\mu\rho}\left(\left\|y^0\right\| + c\right)^2\right)$.

Ergodic Rate of Convergence Results

Corollary (A fast ergodic function values and feasibility violation rates)

Let $\left\{\left(z^k, y^k\right)\right\}_{k\in\mathbb{N}}$ *be a sequence generated by FLAG.* **Suppose that** $\sigma > 0$ **and** $0 \preceq P \preceq (\sigma/2)\, I_n$. *Then, for any optimal solution* z^* *of problem (P), the following holds for the ergodic sequence* $\bar{z}^N = t_{N-1}^{-2}\sum_{k=0}^{N-1} t_k z^{k+1}$

$$\Psi\left(\bar{z}^N\right) - \Psi\left(z^*\right) \leq \frac{B_{\rho,c}\left(z^*\right)}{2N^2} \quad \text{and} \quad \left\|\mathcal{A}\bar{z}^N - b\right\| \leq \frac{B_{\rho,c}\left(z^*\right)}{cN^2},$$

where $B_{\rho,c}\left(z^*\right) := 4\left(\left\|z^* - z^0\right\|_P^2 + \frac{1}{\mu\rho}\left(\left\|y^0\right\| + c\right)^2\right)$.

Corollary (An ergodic function values and feasibility violation rates)

Let $\left\{\left(z^k, y^k\right)\right\}_{k\in\mathbb{N}}$ *be a sequence generated by FLAG with* $\sigma = 0$ **and** $t_k = 1$ **for all** $k \in \mathbb{N}$. *Then, for any optimal solution* z^* *of problem (P), the following holds for the ergodic sequence* $\bar{z}^N = N^{-1}\sum_{k=0}^{N-1} z^{k+1}$

$$\Psi\left(\bar{z}^N\right) - \Psi\left(z^*\right) \leq \frac{B_{\rho,c}\left(z^*\right)}{2N} \quad \text{and} \quad \left\|\mathcal{A}\bar{z}^N - b\right\| \leq \frac{B_{\rho,c}\left(z^*\right)}{cN},$$

where $B_{\rho,c}\left(z^*\right) := 2\left(\left\|z^* - x^0\right\|_P^2 + \frac{1}{\mu\rho}\left(\left\|y^0\right\| + c\right)^2\right)$.

Nice Primal Algorithmic Maps for Block Model

The notion of nice algorithmic map is flexible and **easily adapt to the block setting**. Recalling the model

$$\min_{(u,v)\in\mathbb{R}^n} \left\{ f(u) + g(v) \ : \ Au + Bv = b \right\}.$$

In the block model, we can assume that **either f or g is possibly strongly convex**. Here, without loss of generality, the **σ-strong convexity is assumed in g**.

Nice Primal Algorithmic Maps for Block Model

The notion of nice algorithmic map is flexible and **easily adapt to the block setting**. Recalling the model

$$\min_{(u,v)\in\mathbb{R}^n} \left\{ f(u) + g(v) \,:\, Au + Bv = b \right\}.$$

In the block model, we can assume that **either f or g is possibly strongly convex**. Here, without loss of generality, the **σ-strong convexity is assumed in g**.

Definition (Nice primal algorithmic map - Block version)

Given $\rho, t > 0$. A primal algorithmic map $\mathrm{Prim}_t : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ that generates $z^+ = \left(u^+, v^+\right)$ via $z^+ \in \mathrm{Prim}_t(z, \lambda)$, is called **nice**, **if there exist a parameter $\delta \in (0, 1]$ and matrices $P_1, Q_1 \in \mathbb{S}_+^p$ and $P_2, Q_2 \in \mathbb{S}_+^q$ with $P = (P_1, P_2)$ and $Q = (Q_1, Q_2)$**, such that for any $(\xi_1, \xi_2) \in \mathcal{F}$ we have

$$\begin{aligned}\mathcal{L}_{\rho_t}\left(z^+, \lambda\right) - \mathcal{L}_{\rho_t}(\xi, \lambda) \leq{} & \Delta_{P_1}\left(\xi_1, u, u^+\right) - \frac{1}{2}\left\|u^+ - u\right\|_{Q_1}^2 \\ & + \tau_t \Delta_{P_2}\left(\xi_2, v, v^+\right) - \frac{\tau_t}{2}\left\|v^+ - v\right\|_{Q_2}^2 - \frac{\sigma}{2}\left\|\xi_2 - v^+\right\|^2 \\ & - \frac{\delta \rho_t}{2}\left\|\mathcal{A}z^+ - b\right\|^2,\end{aligned}$$

where $\tau_t = 1$ and $\rho_t = \rho$ (when $\sigma = 0$) or $\tau_t = t$ and $\rho_t = \rho t$ (when $\sigma > 0$).

Iconic Lagrangian-based Methods Admit a Nice Primal Algorithmic Map!

- Augmented Lagrangian Methods (classical, proximal, and prox-linearized)
- Alternating Direction Method of Multipliers ADMM
- Proximal ADMM
- Proximal Linearized ADMM
- Chambolle-Pock Method
- Proximal Jacobi Direction Method of Multipliers
- Predictor Corrector Proximal Multipliers

For each method the explicit parameter δ and the matrices P, Q can be found!
(See details in paper.)

Iconic Lagrangian-based Methods Admit a Nice Primal Algorithmic Map!

- Augmented Lagrangian Methods (classical, proximal, and prox-linearized)
- Alternating Direction Method of Multipliers ADMM
- Proximal ADMM
- Proximal Linearized ADMM
- Chambolle-Pock Method
- Proximal Jacobi Direction Method of Multipliers
- Predictor Corrector Proximal Multipliers

For each method the explicit parameter δ and the matrices P, Q can be found! (See details in paper.)

Meaning, they all **admit Nice Primal Algorithmic Map!**

Therefore, our non-ergodic convergence rate results can be applied.

In addition, nice primal algorithmic maps, can be also be identified for problems with **composite objective**...

Model (P) with Additional Smooth Function

We consider the following model

$$\min_{x \in \mathbb{R}^n} \left\{ f(x) + h(x) \, : \, \mathcal{A}x = b \right\},$$

where $f : \mathbb{R}^n \to (-\infty, +\infty]$ and $h : \mathbb{R}^n \to \mathbb{R}$ is a proper, lower semi-continuous and **σ-strongly convex (with $\sigma \geq 0$)** while h is convex and continuously differentiable with **L-Lipschitz continuous gradient**.

Model (P) with Additional Smooth Function

We consider the following model

$$\min_{x \in \mathbb{R}^n} \{ f(x) + h(x) \, : \, \mathcal{A}x = b \},$$

where $f : \mathbb{R}^n \to (-\infty, +\infty]$ and $h : \mathbb{R}^n \to \mathbb{R}$ is a proper, lower semi-continuous and **σ-strongly convex (with $\sigma \geq 0$)** while h is convex and continuously differentiable with **L-Lipschitz continuous gradient**.

Lemma (Proximal AL is nice)

Let $M \succeq LI_n$, the primal algorithmic map $\mathrm{Prim}_t(\cdot)$ *defined by*

$$z^+ = \mathrm{argmin}_\xi \left\{ f(\xi) + \langle \nabla h(z), \xi \rangle + \langle \lambda, \mathcal{A}\xi - b \rangle + \frac{\rho_t}{2} \|\mathcal{A}\xi - b\|^2 + \frac{\tau_t}{2} \|\xi - z\|_M^2 \right\},$$

is nice with $\delta = 1$ and $P = M$ and $Q = M - LI_n$.

Model (P) with Additional Smooth Function

We consider the following model

$$\min_{x \in \mathbb{R}^n} \left\{ f(x) + h(x) \; : \; \mathcal{A}x = b \right\},$$

where $f : \mathbb{R}^n \to (-\infty, +\infty]$ and $h : \mathbb{R}^n \to \mathbb{R}$ is a proper, lower semi-continuous and **σ-strongly convex (with $\sigma \geq 0$)** while h is convex and continuously differentiable with **L-Lipschitz continuous gradient**.

Lemma (Proximal AL is nice)

Let $M \succeq LI_n$, the primal algorithmic map $\mathrm{Prim}_t(\cdot)$ *defined by*

$$z^+ = \mathrm{argmin}_\xi \left\{ f(\xi) + \langle \nabla h(z), \xi \rangle + \langle \lambda, \mathcal{A}\xi - b \rangle + \frac{\rho_t}{2} \|\mathcal{A}\xi - b\|^2 + \frac{\tau_t}{2} \|\xi - z\|_M^2 \right\},$$

is nice with $\delta = 1$ and $P = M$ and $Q = M - LI_n$.

Lemma (Proximal Linearized AL is nice)

Let $M \succeq \rho\mathcal{A}^T\mathcal{A} + LI_n$, the primal algorithmic map $\mathrm{Prim}_t(\cdot)$ *defined by*

$$z^+ = \mathrm{argmin}_\xi \left\{ f(\xi) + \langle \nabla h(z), \xi \rangle + \langle \lambda, \mathcal{A}\xi - b \rangle + \rho_t \langle \mathcal{A}z - b, \mathcal{A}\xi \rangle + \frac{\tau_t}{2} \|\xi - z\|_M^2 \right\},$$

is nice with $\delta = 1$ and $P = M - \rho\mathcal{A}^T\mathcal{A}$ and $Q = M - \rho\mathcal{A}^T\mathcal{A} - LI_n$.

To summarize

A Recipe for Rate of Convergence of Lagrangian-based Methods

(i) Formulate the problem at hand via model (P), *i.e.*, **identify the relevant problem data** $[\Psi, \mathcal{A}, b, \sigma]$. **The value of σ will determine the type of rate that can be achieved (classical or fast)**.

To summarize

A Recipe for Rate of Convergence of Lagrangian-based Methods

(i) Formulate the problem at hand via model (P), *i.e.*, **identify the relevant problem data** $[\Psi, \mathcal{A}, b, \sigma]$. **The value of σ will determine the type of rate that can be achieved (classical or fast)**.

(ii) **Define the desired iterative step(s) of the primal algorithmic map** $\text{Prim}_t(\cdot)$ applied on the augmented Lagrangian $\mathcal{L}_{\rho_t}(\cdot)$ of model (P).

To summarize

A Recipe for Rate of Convergence of Lagrangian-based Methods

(i) Formulate the problem at hand via model (P), *i.e.*, **identify the relevant problem data** $[\Psi, \mathcal{A}, b, \sigma]$. **The value of σ will determine the type of rate that can be achieved (classical or fast)**.

(ii) **Define the desired iterative step(s) of the primal algorithmic map** $\text{Prim}_t(\cdot)$ applied on the augmented Lagrangian $\mathcal{L}_{\rho_t}(\cdot)$ of model (P).

(iii) **Show that the defined primal algorithmic map is nice**, *i.e.*, determine the parameter δ and the matrices P and Q.

To summarize

A Recipe for Rate of Convergence of Lagrangian-based Methods

(i) Formulate the problem at hand via model (P), *i.e.*, **identify the relevant problem data** $[\Psi, \mathcal{A}, b, \sigma]$. **The value of σ will determine the type of rate that can be achieved (classical or fast)**.

(ii) **Define the desired iterative step(s) of the primal algorithmic map** $\text{Prim}_t(\cdot)$ applied on the augmented Lagrangian $\mathcal{L}_{\rho_t}(\cdot)$ of model (P).

(iii) **Show that the defined primal algorithmic map is nice**, *i.e.*, determine the parameter δ and the matrices P and Q.

(iv) Apply Theorem 1 (if $\sigma > 0$) or Theorem 2 (if $\sigma = 0$) to **obtain a faster non-ergodic rate of convergence for the designed method**.

To summarize

A Recipe for Rate of Convergence of Lagrangian-based Methods

(i) Formulate the problem at hand via model (P), *i.e.*, **identify the relevant problem data** $[\Psi, \mathcal{A}, b, \sigma]$. **The value of σ will determine the type of rate that can be achieved (classical or fast)**.

(ii) **Define the desired iterative step(s) of the primal algorithmic map** $\text{Prim}_t(\cdot)$ applied on the augmented Lagrangian $\mathcal{L}_{\rho_t}(\cdot)$ of model (P).

(iii) **Show that the defined primal algorithmic map is nice**, *i.e.*, determine the parameter δ and the matrices P and Q.

(iv) Apply Theorem 1 (if $\sigma > 0$) or Theorem 2 (if $\sigma = 0$) to **obtain a faster non-ergodic rate of convergence for the designed method**.

Therefore, there is no need any more to enter into the machinery of the proofs!

For more information and results see

Sabach, S. and Teboulle, M.: **Faster Lagrangian-Based Methods in Convex Optimization**, SIAM Journal on Optimization (2022).

Thanks for your attention!

Email: ssabach@technion.ac.il

Website: http://ssabach.net.technion.ac.il/