Decentralized optimization for saddle point problems with local and global variables

Moscow Conference on Combinatorics and Applications

Alexander Rogozin

Joint work with Alexander Beznosikov, Darina Dvinskikh, Dmitry Kovalev, Pavel Dvurechenskiy and Alexander Gasnikov

Moscow Institute of Physics and Technology

June 2, 2021

Outline

- Motivation and problem statement.

Outline

- Motivation and problem statement.
- Lagrange-based reformulation.

Outline

- Motivation and problem statement.
- Lagrange-based reformulation.
- Result for a general proximal setup.

Outline

- Motivation and problem statement.
- Lagrange-based reformulation.
- Result for a general proximal setup.
- Eucledian setup, convex-concave and strongly convex-concave cases.

Problem of interest

We study a saddle-point problem of the form

$$\min_{p,\{x_i\}_{i=1}^m} \max_{r,\{y_i\}_{i=1}^m} \frac{1}{m}\sum_{i=1}^m f_i(x_i, p, y_i, r).$$

Problem of interest

We study a saddle-point problem of the form

$$\min_{p,\{x_i\}_{i=1}^m} \max_{r,\{y_i\}_{i=1}^m} \frac{1}{m}\sum_{i=1}^m f_i(x_i, p, y_i, r).$$

- Each f_i is stored at a separate computational node.

Problem of interest

We study a saddle-point problem of the form

$$\min_{p, \{x_i\}_{i=1}^m} \max_{r, \{y_i\}_{i=1}^m} \frac{1}{m} \sum_{i=1}^m f_i(x_i, p, y_i, r).$$

- Each f_i is stored at a separate computational node.
- Nodes are connected by a decentralized communication network.

Problem of interest

We study a saddle-point problem of the form

$$\min_{p, \{x_i\}_{i=1}^m} \max_{r, \{y_i\}_{i=1}^m} \frac{1}{m} \sum_{i=1}^m f_i(x_i, p, y_i, r).$$

- Each f_i is stored at a separate computational node.
- Nodes are connected by a decentralized communication network.
- Variables $\{x_i\}_{i=1}^m$ and $\{y_i\}_{i=1}^m$ are individual for each agent.

Problem of interest

We study a saddle-point problem of the form

$$\min_{p,\{x_i\}_{i=1}^m} \max_{r,\{y_i\}_{i=1}^m} \frac{1}{m}\sum_{i=1}^m f_i(x_i, p, y_i, r).$$

- Each f_i is stored at a separate computational node.
- Nodes are connected by a decentralized communication network.
- Variables $\{x_i\}_{i=1}^m$ and $\{y_i\}_{i=1}^m$ are individual for each agent.
- Variables p and r are common for all the nodes, and agreement constraints on them are imposed.

Example 1: optimization problems with separable constraints

We begin with an example of a network utility optimization problem [Nedic and Ozdaglar, 2010].

Example 1: optimization problems with separable constraints

We begin with an example of a network utility optimization problem [Nedic and Ozdaglar, 2010].

- Sources: $\mathcal{S} = \{1, \ldots, S\}$.

Example 1: optimization problems with separable constraints

We begin with an example of a network utility optimization problem [Nedic and Ozdaglar, 2010].

- Sources: $\mathcal{S} = \{1, \ldots, S\}$.
- Links: $\mathcal{L} = \{1, \ldots, L\}$, each $\ell \in \mathcal{L}$ has capacity c_ℓ.

Example 1: optimization problems with separable constraints

We begin with an example of a network utility optimization problem [Nedic and Ozdaglar, 2010].

- Sources: $\mathcal{S} = \{1, \ldots, S\}$.
- Links: $\mathcal{L} = \{1, \ldots, L\}$, each $\ell \in \mathcal{L}$ has capacity c_ℓ.
- Denote $\mathcal{S}(\ell)$ the set of sources that use link ℓ.

Example 1: optimization problems with separable constraints

We begin with an example of a network utility optimization problem [Nedic and Ozdaglar, 2010].

- Sources: $\mathcal{S} = \{1, \ldots, S\}$.
- Links: $\mathcal{L} = \{1, \ldots, L\}$, each $\ell \in \mathcal{L}$ has capacity c_ℓ.
- Denote $\mathcal{S}(\ell)$ the set of sources that use link ℓ.
- When the i-th source transmits data at rate x_i, its utility is characterized as a concave function $u_i(x_i) : \mathbb{R}_+ \to \mathbb{R}_+$.

Example 1: optimization problems with separable constraints

We begin with an example of a network utility optimization problem [Nedic and Ozdaglar, 2010].

- Sources: $\mathcal{S} = \{1, \ldots, S\}$.
- Links: $\mathcal{L} = \{1, \ldots, L\}$, each $\ell \in \mathcal{L}$ has capacity c_ℓ.
- Denote $\mathcal{S}(\ell)$ the set of sources that use link ℓ.
- When the i-th source transmits data at rate x_i, its utility is characterized as a concave function $u_i(x_i) : \mathbb{R}_+ \to \mathbb{R}_+$.

The problem writes as

$$\max \sum_{i \in \mathcal{S}} u_i(x_i)$$
$$\text{s.t.} \sum_{i \in \mathcal{S}(\ell)} x_i \le c_\ell, \ \ x_i \ge 0.$$

Example 1: optimization problems with separable constraints

In a more general form, consider objectives $\{f_i(x_i, p)\}_{i=1}^m$ and constraint functions $g_i(x_i, p)$ with each f_i and g_i being convex in (x_i, p) [Mateos-Núnez and Cortés, 2015].

$$\min_{p, \{x_i\}_{i=1}^m} \sum_{i=1}^m f_i(x_i, p)$$
$$\text{s.t. } g_1(x_1, p) + \ldots + g_m(x_m, p) \leq 0.$$

Example 1: optimization problems with separable constraints

In a more general form, consider objectives $\{f_i(x_i, p)\}_{i=1}^m$ and constraint functions $g_i(x_i, p)$ with each f_i and g_i being convex in (x_i, p) [Mateos-Núnez and Cortés, 2015].

$$\min_{p, \{x_i\}_{i=1}^m} \sum_{i=1}^m f_i(x_i, p)$$
$$\text{s.t. } g_1(x_1, p) + \ldots + g_m(x_m, p) \leq 0.$$

Introducing Lagrange multipliers z yields an equivalent saddle-point problem

$$\min_{\{x_i\}, p} \max_z \sum_{i=1}^m f_i(x_i, p) + z^\top \sum_{i=1}^m g_i(x_i, p).$$

Example 1: optimization problems with separable constraints

In a more general form, consider objectives $\{f_i(x_i, p)\}_{i=1}^m$ and constraint functions $g_i(x_i, p)$ with each f_i and g_i being convex in (x_i, p) [Mateos-Núnez and Cortés, 2015].

$$\min_{p, \{x_i\}_{i=1}^m} \sum_{i=1}^m f_i(x_i, p)$$
$$\text{s.t. } g_1(x_1, p) + \ldots + g_m(x_m, p) \leq 0.$$

Introducing Lagrange multipliers z yields an equivalent saddle-point problem

$$\min_{\{x_i\}, p} \max_z \sum_{i=1}^m f_i(x_i, p) + z^\top \sum_{i=1}^m g_i(x_i, p).$$

In this formulation, p and z are global variables, while $\{x_i\}_{i=1}^m$ are local.

Example 2: Wasserstein Barycenters (WB)

- Probability simplex $\Delta_n = \left\{ x \in \mathbb{R}^n : \ x_i \geq 0, \ \sum_{i=1}^{n} x_i = 1 \right\}$.

Example 2: Wasserstein Barycenters (WB)

- Probability simplex $\Delta_n = \left\{ x \in \mathbb{R}^n : \; x_i \geq 0, \; \sum_{i=1}^{n} x_i = 1 \right\}$.
- Ground cost matrix $C \in \mathbb{R}_+^{n \times n}$ characterizes transportation costs.

Example 2: Wasserstein Barycenters (WB)

- Probability simplex $\Delta_n = \left\{ x \in \mathbb{R}^n : \ x_i \geq 0, \ \sum_{i=1}^n x_i = 1 \right\}.$
- Ground cost matrix $C \in \mathbb{R}_+^{n \times n}$ characterizes transportation costs.
- For histograms $\tilde{p}, \tilde{q} \in \Delta_n$ define Wasserstein distance

$$\mathcal{W}(\tilde{p}, \tilde{q}) = \min_{X \in \mathbb{R}_+^{n \times n}} \langle C, X \rangle \text{ s.t. } X\mathbf{1} = \tilde{p}, \ X^\top \mathbf{1} = \tilde{q}.$$

Example 2: Wasserstein Barycenters (WB)

- Probability simplex $\Delta_n = \left\{ x \in \mathbb{R}^n : \; x_i \geq 0, \; \sum_{i=1}^{n} x_i = 1 \right\}$.
- Ground cost matrix $C \in \mathbb{R}_+^{n \times n}$ characterizes transportation costs.
- For histograms $\tilde{p}, \tilde{q} \in \Delta_n$ define Wasserstein distance

$$\mathcal{W}(\tilde{p}, \tilde{q}) = \min_{X \in \mathbb{R}_+^{n \times n}} \langle C, X \rangle \text{ s.t. } X\mathbf{1} = \tilde{p}, \; X^\top \mathbf{1} = \tilde{q}.$$

- For given vectors $q_1, q_2 \ldots, q_m$ from the probability simplex Δ_n, their WB is a solution of the following optimization problem:

$$p^* = \arg\min_{p \in \Delta_n} \frac{1}{m} \sum_{i=1}^{m} \mathcal{W}(p, q_i). \tag{1}$$

Example 2: Wasserstein Barycenters (WB)

Following the papers [Dvinskikh and Tiapkin, 2020] and [Jambulapati et al., 2019], we reformulate the WB problem (1) as a saddle point problem.

Example 2: Wasserstein Barycenters (WB)

Following the papers [Dvinskikh and Tiapkin, 2020] and [Jambulapati et al., 2019], we reformulate the WB problem (1) as a saddle point problem. Introduce

- stacked column vector $b_i = (p^\top, q_i^\top)^\top$;

Example 2: Wasserstein Barycenters (WB)

Following the papers [Dvinskikh and Tiapkin, 2020] and [Jambulapati et al., 2019], we reformulate the WB problem (1) as a saddle point problem. Introduce

- stacked column vector $b_i = (p^\top, q_i^\top)^\top$;
- vectorized cost matrix d of C;

Example 2: Wasserstein Barycenters (WB)

Following the papers [Dvinskikh and Tiapkin, 2020] and [Jambulapati et al., 2019], we reformulate the WB problem (1) as a saddle point problem. Introduce

- stacked column vector $b_i = (p^\top, q_i^\top)^\top$;
- vectorized cost matrix d of C;
- vectorized transport plan $x \in \Delta_{n^2}$ of X;

Example 2: Wasserstein Barycenters (WB)

Following the papers [Dvinskikh and Tiapkin, 2020] and [Jambulapati et al., 2019], we reformulate the WB problem (1) as a saddle point problem. Introduce

- stacked column vector $b_i = (p^\top, q_i^\top)^\top$;
- vectorized cost matrix d of C;
- vectorized transport plan $x \in \Delta_{n^2}$ of X;
- incidence matrix $A = \{0, 1\}^{2n \times n^2}$;

Example 2: Wasserstein Barycenters (WB)

Following the papers [Dvinskikh and Tiapkin, 2020] and [Jambulapati et al., 2019], we reformulate the WB problem (1) as a saddle point problem. Introduce

- stacked column vector $b_i = (p^\top, q_i^\top)^\top$;
- vectorized cost matrix d of C;
- vectorized transport plan $x \in \Delta_{n^2}$ of X;
- incidence matrix $A = \{0, 1\}^{2n \times n^2}$;
- vectors $y_i \in [-1, 1]^{2n}$, $i = 1, \ldots, m$.

Example 2: Wasserstein Barycenters (WB)

Following the papers [Dvinskikh and Tiapkin, 2020] and [Jambulapati et al., 2019], we reformulate the WB problem (1) as a saddle point problem. Introduce

- stacked column vector $b_i = (p^\top, q_i^\top)^\top$;
- vectorized cost matrix d of C;
- vectorized transport plan $x \in \Delta_{n^2}$ of X;
- incidence matrix $A = \{0,1\}^{2n \times n^2}$;
- vectors $y_i \in [-1,1]^{2n}$, $i = 1, \ldots, m$.

Then (1) can be equivalently rewritten as

$$\min_{p \in \Delta_n} \frac{1}{m} \sum_{i=1}^{m} \min_{x_i \in \Delta_{n^2}} \max_{y_i \in [-1,1]^{2n}} \left\{ d^\top x_i + 2\|d\|_\infty \left(y_i^\top A x_i - b_i^\top y_i \right) \right\}.$$

Problem statement

We study a saddle-point problem of the form

$$\min_{\substack{p\in\bar{\mathcal{P}}\\ \mathbf{x}\in\mathcal{X}}} \max_{\substack{r\in\bar{\mathcal{R}}\\ \mathbf{y}\in\mathcal{Y}}} \; f(\mathbf{x}, p, \mathbf{y}, r) = \frac{1}{m}\sum_{i=1}^{m} f_i(x_i, p, y_i, r), \tag{2}$$

Problem statement

We study a saddle-point problem of the form

$$\min_{\substack{p \in \bar{\mathcal{P}} \\ \mathbf{x} \in \mathcal{X}}} \max_{\substack{r \in \bar{\mathcal{R}} \\ \mathbf{y} \in \mathcal{Y}}} f(\mathbf{x}, p, \mathbf{y}, r) = \frac{1}{m} \sum_{i=1}^{m} f_i(x_i, p, y_i, r), \tag{2}$$

where $\mathbf{x} = (x_1^\top \ldots x_m^\top)^\top$, $\mathbf{y} = (y_1^\top \ldots y_m^\top)^\top$ and $\mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_m$, $\mathcal{Y} = \mathcal{Y}_1 \times \ldots \times \mathcal{Y}_m$.

Problem statement

We study a saddle-point problem of the form

$$\min_{\substack{p\in\bar{\mathcal{P}}\\ \mathbf{x}\in\mathcal{X}}} \max_{\substack{r\in\bar{\mathcal{R}}\\ \mathbf{y}\in\mathcal{Y}}} f(\mathbf{x}, p, \mathbf{y}, r) = \frac{1}{m}\sum_{i=1}^{m} f_i(x_i, p, y_i, r), \tag{2}$$

where $\mathbf{x} = (x_1^\top \ldots x_m^\top)^\top$, $\mathbf{y} = (y_1^\top \ldots y_m^\top)^\top$ and $\mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_m$, $\mathcal{Y} = \mathcal{Y}_1 \times \ldots \times \mathcal{Y}_m$.

Assumption

- Sets $\mathcal{X}_i, \mathcal{Y}_i$, $i = 1, \ldots, m$, $\bar{\mathcal{P}}$, $\bar{\mathcal{R}}$ are convex compacts.
- Each $f_i(\cdot, \cdot, y_i, r)$ is convex on $\mathcal{X}_i \times \bar{\mathcal{P}}$ for every fixed $y_i \in \mathcal{Y}_i$, $r \in \bar{\mathcal{R}}$.
- Each $f_i(x_i, p, \cdot, \cdot)$ is concave on $\mathcal{Y}_i \times \bar{\mathcal{R}}$ for every fixed $x_i \in \mathcal{X}_i$, $p \in \mathcal{P}$.

Problem statement

We study a saddle-point problem of the form

$$\min_{\substack{p \in \bar{\mathcal{P}} \\ \mathbf{x} \in \mathcal{X}}} \max_{\substack{r \in \bar{\mathcal{R}} \\ \mathbf{y} \in \mathcal{Y}}} f(\mathbf{x}, p, \mathbf{y}, r) = \frac{1}{m} \sum_{i=1}^{m} f_i(x_i, p, y_i, r), \tag{2}$$

where $\mathbf{x} = (x_1^\top \ldots x_m^\top)^\top$, $\mathbf{y} = (y_1^\top \ldots y_m^\top)^\top$ and $\mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_m$, $\mathcal{Y} = \mathcal{Y}_1 \times \ldots \times \mathcal{Y}_m$.

Assumption

- Sets $\mathcal{X}_i, \mathcal{Y}_i$, $i = 1, \ldots, m$, $\bar{\mathcal{P}}$, $\bar{\mathcal{R}}$ are convex compacts.
- Each $f_i(\cdot, \cdot, y_i, r)$ is convex on $\mathcal{X}_i \times \bar{\mathcal{P}}$ for every fixed $y_i \in \mathcal{Y}_i$, $r \in \bar{\mathcal{R}}$.
- Each $f_i(x_i, p, \cdot, \cdot)$ is concave on $\mathcal{Y}_i \times \bar{\mathcal{R}}$ for every fixed $x_i \in \mathcal{X}_i$, $p \in \mathcal{P}$.

Variables x_i, p, y_i, r have dimensions d_x, d_p, d_y, d_r, respectively.

Communication constraints

- Each f_i is stored at a separate computational agent.

Communication constraints

- Each f_i is stored at a separate computational agent.
- The agents interact via a connected undirected network represented by a fixed graph $\mathcal{G} = (V, E)$. Every pair of agents (i, j) can communicate iff $(i, j) \in E$.

Communication constraints

- Each f_i is stored at a separate computational agent.
- The agents interact via a connected undirected network represented by a fixed graph $\mathcal{G} = (V, E)$. Every pair of agents (i, j) can communicate iff $(i, j) \in E$.
- Each agent i stores a local copy p_i, r_i of the global variables p and r, and *consensus constraints* $p_1 = \ldots = p_m$, $r_1 = \ldots = r_m$ are imposed.

Communication constraints

We introduce a matrix $\tilde{W}$ associated with the network and satisfying the following assumption.

Communication constraints

We introduce a matrix $\tilde{W}$ associated with the network and satisfying the following assumption.

Assumption

- $\tilde{W}$ is symmetric positive semi-definite.
- (Network compatibility) $[\tilde{W}]_{ij} = 0$ if $(i, j) \notin E$ and $i \neq j$.
- (Kernel property) For any $v = [v_1, \ldots, v_m]^\top \in \mathbb{R}^m$, $\tilde{W} v = 0$ if and only if $v_1 = \ldots = v_m$. In other words, $\operatorname{Ker} \tilde{W} = \operatorname{span}\{\mathbf{1}\}$.

Communication constraints

We introduce a matrix $\tilde{W}$ associated with the network and satisfying the following assumption.

Assumption

- $\tilde{W}$ is symmetric positive semi-definite.
- (Network compatibility) $[\tilde{W}]_{ij} = 0$ if $(i, j) \notin E$ and $i \neq j$.
- (Kernel property) For any $v = [v_1, \ldots, v_m]^\top \in \mathbb{R}^m$, $\tilde{W} v = 0$ if and only if $v_1 = \ldots = v_m$. In other words, $\mathrm{Ker}\, \tilde{W} = \mathrm{span}\, \{\mathbf{1}\}$.

An example of matrix satisfying this Assumption is the graph *Laplacian matrix* $\tilde{W} \in \mathbb{R}^{m \times m}$

Communication constraints

We introduce a matrix $\tilde{W}$ associated with the network and satisfying the following assumption.

Assumption

- $\tilde{W}$ is symmetric positive semi-definite.
- (Network compatibility) $[\tilde{W}]_{ij} = 0$ if $(i, j) \notin E$ and $i \neq j$.
- (Kernel property) For any $v = [v_1, \ldots, v_m]^\top \in \mathbb{R}^m$, $\tilde{W} v = 0$ if and only if $v_1 = \ldots = v_m$. In other words, $\text{Ker } \tilde{W} = \text{span}\{\mathbf{1}\}$.

An example of matrix satisfying this Assumption is the graph *Laplacian matrix* $\tilde{W} \in \mathbb{R}^{m \times m}$ such that a) $[\tilde{W}]_{ij} = -1$ if $(i, j) \in E$,

Communication constraints

We introduce a matrix $\tilde{W}$ associated with the network and satisfying the following assumption.

Assumption

- $\tilde{W}$ is symmetric positive semi-definite.
- (Network compatibility) $[\tilde{W}]_{ij} = 0$ if $(i, j) \notin E$ and $i \neq j$.
- (Kernel property) For any $v = [v_1, \ldots, v_m]^\top \in \mathbb{R}^m$, $\tilde{W} v = 0$ if and only if $v_1 = \ldots = v_m$. In other words, $\operatorname{Ker} \tilde{W} = \operatorname{span}\{\mathbf{1}\}$.

An example of matrix satisfying this Assumption is the graph *Laplacian matrix* $\tilde{W} \in \mathbb{R}^{m \times m}$ such that a) $[\tilde{W}]_{ij} = -1$ if $(i, j) \in E$, b) $[\tilde{W}]_{ij} = \deg(i)$ if $i = j$,

Communication constraints

We introduce a matrix $\tilde{W}$ associated with the network and satisfying the following assumption.

Assumption

- $\tilde{W}$ is symmetric positive semi-definite.
- (Network compatibility) $[\tilde{W}]_{ij} = 0$ if $(i, j) \notin E$ and $i \neq j$.
- (Kernel property) For any $v = [v_1, \ldots, v_m]^\top \in \mathbb{R}^m$, $\tilde{W} v = 0$ if and only if $v_1 = \ldots = v_m$. In other words, $\operatorname{Ker} \tilde{W} = \operatorname{span}\{\mathbf{1}\}$.

An example of matrix satisfying this Assumption is the graph *Laplacian matrix* $\tilde{W} \in \mathbb{R}^{m\times m}$ such that a) $[\tilde{W}]_{ij} = -1$ if $(i, j) \in E$, b) $[\tilde{W}]_{ij} = \deg(i)$ if $i = j$, c) $[\tilde{W}]_{ij} = 0$ otherwise. Here $\deg(i)$ is the degree of the node i.

Communication constraints

The *communication matrix* is then defined as $\tilde{\mathbf{W}} \triangleq \tilde{W} \otimes I_d$, where $\otimes$ denotes the Kronecker product and d is the dimension of variables on which affine constraints are imposed.

Communication constraints

The *communication matrix* is then defined as $\tilde{\mathbf{W}} \triangleq \tilde{W} \otimes I_d$, where $\otimes$ denotes the Kronecker product and d is the dimension of variables on which affine constraints are imposed.

- Let $\mathbf{v} = (v_1^\top \dots v_m^\top)^\top$, $\mathbf{w} = (w_1^\top \dots w_m^\top)^\top$. Multiplication $\mathbf{w} = \tilde{\mathbf{W}}\mathbf{v}$ corresponds to one communication round: $w_i = \sum_{(i,j)\in E} [\tilde{W}]_{ij} v_j$.

Communication constraints

The *communication matrix* is then defined as $\tilde{\mathbf{W}} \triangleq \tilde{W} \otimes I_d$, where $\otimes$ denotes the Kronecker product and d is the dimension of variables on which affine constraints are imposed.

- Let $\mathbf{v} = (v_1^\top \dots v_m^\top)^\top$, $\mathbf{w} = (w_1^\top \dots w_m^\top)^\top$. Multiplication $\mathbf{w} = \tilde{\mathbf{W}}\mathbf{v}$ corresponds to one communication round: $w_i = \sum_{(i,j)\in E} [\tilde{W}]_{ij} v_j$.
- Constraints $v_1 = \ldots = v_n$ can be written as $\tilde{\mathbf{W}}\mathbf{v} = 0$.

Communication constraints

The *communication matrix* is then defined as $\tilde{\mathbf{W}} \triangleq \tilde{W} \otimes I_d$, where $\otimes$ denotes the Kronecker product and d is the dimension of variables on which affine constraints are imposed.

- Let $\mathbf{v} = (v_1^\top \dots v_m^\top)^\top$, $\mathbf{w} = (w_1^\top \dots w_m^\top)^\top$. Multiplication $\mathbf{w} = \tilde{\mathbf{W}}\mathbf{v}$ corresponds to one communication round: $w_i = \sum\limits_{(i,j)\in E} [\tilde{W}]_{ij} v_j$.
- Constraints $v_1 = \ldots = v_n$ can be written as $\tilde{\mathbf{W}}\mathbf{v} = 0$.
- Performance depends on the condition number $\chi = \frac{\lambda_{\max}(\tilde{W})}{\lambda_{\min}^+(\tilde{W})}$.

Chebyshev acceleration

Communication matrix $\tilde{\mathbf{W}}$ can be replaced with a polynomial $P_K(\tilde{\mathbf{W}})$ of degree $K = \lfloor \sqrt{\chi} \rfloor$ and $\chi(P_K(\tilde{\mathbf{W}})) = O(1)$. Due to the specific polynomial structure, $P_K(\tilde{\mathbf{W}})$ is positive semi-definite and satisfies the kernel property in Assumption 2.

Chebyshev acceleration

Communication matrix $\tilde{\mathbf{W}}$ can be replaced with a polynomial $P_K(\tilde{\mathbf{W}})$ of degree $K = \lfloor\sqrt{\chi}\rfloor$ and $\chi(P_K(\tilde{\mathbf{W}})) = O(1)$. Due to the specific polynomial structure, $P_K(\tilde{\mathbf{W}})$ is positive semi-definite and satisfies the kernel property in Assumption 2. Multiplication on $P_K(\tilde{\mathbf{W}})$ is performed by a communication subroutine described below.

Algorithm 2 Chebyshev gossip subroutine

Require: $\mathbf{x},\ c_2 = \frac{\chi+1}{\chi-1},\ a_0 = 1,\ a_1 = c_2,\ c_3 = \frac{2}{\lambda_{\max}(\tilde{W})+\lambda_{\min}^+(\tilde{W})}$.

1: $\mathbf{x}^0 = \mathbf{x},\ \mathbf{x}^1 = c_2(\mathbf{I} - c_3\mathbf{W})\mathbf{x}$.
2: **for** $k = 1, \ldots, K-1$ **do**
3: $\quad a_{k+1} = 2c_2 a_k - a_{k-1}$.
4: $\quad \mathbf{x}^{k+1} = 2c_2(\mathbf{I} - c_3\mathbf{W})\mathbf{x}^k - \mathbf{x}^{k-1}$.
5: **end for**

Ensure: $\mathbf{x}^0 - \frac{\mathbf{x}^K}{a_K}$.

Problem reformulation

Introduce $F(\mathbf{x}, \mathbf{p}, \mathbf{y}, \mathbf{r}) = \sum_{i=1}^m f_i(x_i, p_i, y_i, r_i)$, two communication matrices $\mathbf{W_r}$, $\mathbf{W_p}$ and rewrite problem (2) as

$$\min_{\substack{\mathbf{W_p p}=0 \\ \mathbf{x}\in\mathcal{X}, \mathbf{p}\in\mathcal{P}}} \max_{\substack{\mathbf{W_r r}=0 \\ \mathbf{y}\in\mathcal{Y}, \mathbf{r}\in\mathcal{R}}} \frac{1}{m}\sum_{i=1}^m f_i(x_i, p_i, y_i, r_i).$$

Problem reformulation

Introduce $F(\mathbf{x}, \mathbf{p}, \mathbf{y}, \mathbf{r}) = \sum_{i=1}^m f_i(x_i, p_i, y_i, r_i)$, two communication matrices $\mathbf{W_r}$, $\mathbf{W_p}$ and rewrite problem (2) as

$$\min_{\substack{\mathbf{W_p p}=0 \\ \mathbf{x}\in\mathcal{X},\mathbf{p}\in\mathcal{P}}} \max_{\substack{\mathbf{W_r r}=0 \\ \mathbf{y}\in\mathcal{Y},\mathbf{r}\in\mathcal{R}}} \frac{1}{m}\sum_{i=1}^m f_i(x_i, p_i, y_i, r_i).$$

After that, we introduce Lagrangian multipliers and get a reformulation

$$\min_{\substack{\mathbf{x}\in\mathcal{X},\mathbf{p}\in\mathcal{P} \\ \mathbf{u}\in\mathbb{R}^{md_r}}} \max_{\substack{\mathbf{y}\in\mathcal{Y},\mathbf{r}\in\mathcal{R} \\ \mathbf{z}\in\mathbb{R}^{md_p}}} \left[F(\mathbf{x}, \mathbf{p}, \mathbf{y}, \mathbf{r}) + \gamma_r \langle \mathbf{u}, \mathbf{W_r r}\rangle + \gamma_p \langle \mathbf{z}, \mathbf{W_p p}\rangle\right], \qquad (3)$$

where γ_r and γ_p are arbitrary positive scalars.

Aggregate all the variables in two blocks: $\xi = (\mathbf{x}^\top, \mathbf{p}^\top, \mathbf{u}^\top)^\top$, $\eta = (\mathbf{y}^\top, \mathbf{r}^\top, \mathbf{z}^\top)^\top$

Aggregate all the variables in two blocks: $\xi = (\mathbf{x}^\top, \mathbf{p}^\top, \mathbf{u}^\top)^\top$, $\eta = (\mathbf{y}^\top, \mathbf{r}^\top, \mathbf{z}^\top)^\top$ and define constraint sets $Q_\xi = \mathcal{X} \times \mathcal{P} \times \mathbb{R}^{md_r}$, $Q_\eta = \mathcal{Y} \times \mathcal{R} \times \mathbb{R}^{md_p}$.

$$\min_{\xi \in Q_\xi} \max_{\eta \in Q_\eta} \; S(\xi, \eta) \triangleq F(\mathbf{x}, \mathbf{p}, \mathbf{y}, \mathbf{r}) + \gamma_r \langle \mathbf{u}, \mathbf{W_r r} \rangle + \gamma_p \langle \mathbf{z}, \mathbf{W_p p} \rangle.$$

Aggregate all the variables in two blocks: $\xi = (\mathbf{x}^\top, \mathbf{p}^\top, \mathbf{u}^\top)^\top$, $\eta = (\mathbf{y}^\top, \mathbf{r}^\top, \mathbf{z}^\top)^\top$ and define constraint sets $Q_\xi = \mathcal{X} \times \mathcal{P} \times \mathbb{R}^{md_r}$, $Q_\eta = \mathcal{Y} \times \mathcal{R} \times \mathbb{R}^{md_p}$.

$$\min_{\xi \in Q_\xi} \max_{\eta \in Q_\eta} S(\xi, \eta) \triangleq F(\mathbf{x}, \mathbf{p}, \mathbf{y}, \mathbf{r}) + \gamma_r \langle \mathbf{u}, \mathbf{W_r r} \rangle + \gamma_p \langle \mathbf{z}, \mathbf{W_p p} \rangle.$$

Introduce $\zeta = (\xi^\top, \eta^\top)^\top$, constraint set $Q_\zeta = Q_\xi \times Q_\eta$ and vector-field $g(\zeta) = (\nabla_\xi^\top S(\xi, \eta), -\nabla_\eta^\top S(\xi, \eta))^\top$.

Aggregate all the variables in two blocks: $\xi = (\mathbf{x}^\top, \mathbf{p}^\top, \mathbf{u}^\top)^\top$, $\eta = (\mathbf{y}^\top, \mathbf{r}^\top, \mathbf{z}^\top)^\top$ and define constraint sets $Q_\xi = \mathcal{X} \times \mathcal{P} \times \mathbb{R}^{md_r}$, $Q_\eta = \mathcal{Y} \times \mathcal{R} \times \mathbb{R}^{md_p}$.

$$\min_{\xi \in Q_\xi} \max_{\eta \in Q_\eta} \; S(\xi, \eta) \triangleq F(\mathbf{x}, \mathbf{p}, \mathbf{y}, \mathbf{r}) + \gamma_r \langle \mathbf{u}, \mathbf{W_r r} \rangle + \gamma_p \langle \mathbf{z}, \mathbf{W_p p} \rangle.$$

Introduce $\zeta = (\xi^\top, \eta^\top)^\top$, constraint set $Q_\zeta = Q_\xi \times Q_\eta$ and vector-field $g(\zeta) = (\nabla_\xi^\top S(\xi, \eta), -\nabla_\eta^\top S(\xi, \eta))^\top$.Saddle-point problem comes down to solving a variational inequality (VI)

$$\text{find } \zeta^* \text{ such that } \; \langle g(\zeta^*), \zeta - \zeta^* \rangle \geq 0.$$

Introduce prox-structure:

- Aggregated norm $\|\zeta\|_\zeta^2 = \|\mathbf{x}\|_\mathbf{x}^2 + \|\mathbf{p}\|_\mathbf{p}^2 + \|\mathbf{u}\|_2^2 + \|\mathbf{y}\|_\mathbf{y}^2 + \|\mathbf{r}\|_\mathbf{r}^2 + \|\mathbf{z}\|_2^2$.

Introduce prox-structure:

- Aggregated norm $\|\zeta\|_\zeta^2 = \|\mathbf{x}\|_\mathbf{x}^2 + \|\mathbf{p}\|_\mathbf{p}^2 + \|\mathbf{u}\|_2^2 + \|\mathbf{y}\|_\mathbf{y}^2 + \|\mathbf{r}\|_\mathbf{r}^2 + \|\mathbf{z}\|_2^2$.
- Prox-function $d_\zeta(\zeta) = \sum_{i=1}^m (d_{x;i}(x_i) + d_{p;i}(p_i) + d_{u;i}(u_i) + d_{y;i}(y_i) + d_{r;i}(r_i) + d_{z;i}(z_i))$, which induces Bregman divergence $B_\zeta(\zeta, \breve{\zeta})$.

Introduce prox-structure:

- Aggregated norm $\|\zeta\|_\zeta^2 = \|\mathbf{x}\|_\mathbf{x}^2 + \|\mathbf{p}\|_\mathbf{p}^2 + \|\mathbf{u}\|_2^2 + \|\mathbf{y}\|_\mathbf{y}^2 + \|\mathbf{r}\|_\mathbf{r}^2 + \|\mathbf{z}\|_2^2$.
- Prox-function $d_\zeta(\zeta) = \sum_{i=1}^m (d_{x;i}(x_i) + d_{p;i}(p_i) + d_{u;i}(u_i) + d_{y;i}(y_i) + d_{r;i}(r_i) + d_{z;i}(z_i))$, which induces Bregman divergence $B_\zeta(\zeta, \breve{\zeta})$.

Algorithm 5 Mirror-Prox

Require: Initial guess ζ^0

1: **for** $k = 0, 1, \ldots, N-1$ **do**

2: $\quad \zeta^{k+\frac{1}{2}} = \operatorname*{argmin}_{\zeta \in Q_\zeta} \left\{ \left\langle g(\zeta^k), \zeta - \zeta^k \right\rangle + B_\zeta(\zeta, \zeta^k) \right\}$

3: $\quad \zeta^{k+1} = \operatorname*{argmin}_{\zeta \in Q_\zeta} \left\{ \left\langle g(\zeta^{k+\frac{1}{2}}), \zeta - \zeta^k \right\rangle + B_\zeta(\zeta, \zeta^k) \right\}$

4: **end for**

Ensure: $\hat{\xi}^N = \frac{1}{N} \sum_{k=0}^{N-1} \zeta^{k+\frac{1}{2}}$.

Result for general proximal setup

Standard analysis of Mirror-Prox requires a smoothness assumption.

Result for general proximal setup

Standard analysis of Mirror-Prox requires a smoothness assumption.

Assumption

Vector-field $g(\zeta)$ is L_ζ-Lipschitz w.r.t. $\|\cdot\|_\zeta$.

Result for general proximal setup

Standard analysis of Mirror-Prox requires a smoothness assumption.

Assumption

Vector-field $g(\zeta)$ is L_ζ-Lipschitz w.r.t. $\|\cdot\|_\zeta$.

Moreover, we need to localize the solution on a bounded set.

Lemma

There exist positive scalars M_p, M_r s.t. for all $i = 1, \ldots, m$ and for any $x_i \in \mathcal{X}_i$, $y_i \in \mathcal{Y}_i$, $p_i \in \bar{\mathcal{P}}$, $r_i \in \bar{\mathcal{R}}$ it holds $\|\nabla_p f_i(x_i, p_i, y_i, r_i)\|_2 \leq M_p$, $\|\nabla_r f_i(x_i, p_i, y_i, r_i)\|_2 \leq M_r$. Introduce $R_z^2 = 2mM_p^2(\gamma_p \lambda_{\min}^+(\mathbf{W_p}))^{-1}$, $R_u^2 = 2mM_r^2(\gamma_r \lambda_{\min}^+(\mathbf{W_r}))^{-1}$, where $\lambda_{\min}^+(\cdot)$ denotes the minimal non-zero eigenvalue of matrix. Then there exists a saddle point $(\mathbf{x}^*, \mathbf{p}^*, \mathbf{y}^*, \mathbf{r}^*, \mathbf{u}^*, \mathbf{z}^*)$ of problem (3) such that $\|\mathbf{u}^*\|_2 \leq R_u$, $\|\mathbf{z}^*\|_2 \leq R_z$.

Result for general proximal setup

Theorem

Let $(\hat{\mathbf{x}}^N, \hat{\mathbf{p}}^N, \hat{\mathbf{y}}^N, \hat{\mathbf{r}}^N, \hat{\mathbf{u}}^N, \hat{\mathbf{z}}^N) = \hat{\zeta}^N$ *and introduce* $\bar{p}^N = \frac{1}{m}\sum_{i=1}^m \hat{p}_i^N$, $\bar{r}^N = \frac{1}{m}\sum_{i=1}^m \hat{r}_i^N$. *Then, for a given accuracy* $\varepsilon > 0$, *after* $N = \left\lceil \frac{L_\zeta R_\zeta^2}{m\varepsilon} \right\rceil$ *steps of Algorithm 3 with stepsize* $\alpha = 1/L_\zeta$ *we have*

$$\max_{\mathbf{y}\in\mathcal{Y}, \bar{r}\in\bar{\mathcal{R}}} f(\hat{\mathbf{x}}^N, \bar{p}^N, \mathbf{y}, \bar{r}) - \min_{\mathbf{x}\in\mathcal{X}, \bar{p}\in\bar{\mathcal{P}}} f(\mathbf{x}, \bar{p}, \hat{\mathbf{y}}^N, \bar{r}^N) \le \varepsilon$$

.

Lower bounds for Euclidean setup

Introduce constraint set size $R^2 = R_{\mathcal{X}}^2 + R_{\bar{\mathcal{P}}}^2 + R_{\mathcal{Y}}^2 + R_{\bar{\mathcal{R}}}^2$.

Lower bounds for Euclidean setup

Introduce constraint set size $R^2 = R_{\mathcal{X}}^2 + R_{\bar{\mathcal{P}}}^2 + R_{\mathcal{Y}}^2 + R_{\bar{\mathcal{R}}}^2$.

- L-smooth convex-concave saddles: $\Omega\left(\frac{LR^2}{\varepsilon}\right)$ oracle calls, $\Omega\left(\frac{LR^2}{\varepsilon}\sqrt{\chi}\right)$ communications.

Lower bounds for Euclidean setup

Introduce constraint set size $R^2 = R_{\mathcal{X}}^2 + R_{\bar{\mathcal{P}}}^2 + R_{\mathcal{Y}}^2 + R_{\bar{\mathcal{R}}}^2$.

- L-smooth convex-concave saddles: $\Omega\left(\frac{LR^2}{\varepsilon}\right)$ oracle calls, $\Omega\left(\frac{LR^2}{\varepsilon}\sqrt{\chi}\right)$ communications.
- L-smooth μ-convex-concave saddles: $\Omega\left(\frac{L}{\mu}\log\frac{1}{\varepsilon}\right)$ computations, $\Omega\left(\frac{L}{\mu}\sqrt{\chi}\log\frac{1}{\varepsilon}\right)$ communications.

Result for Euclidean setup

Let each f_i be L-smooth w.r.t. $\|\cdot\|_2$ and $\mathbf{W_r} = \mathbf{W_p} = \mathbf{W}$.

Result for Euclidean setup

Let each f_i be L-smooth w.r.t. $\|\cdot\|_2$ and $\mathbf{W_r} = \mathbf{W_p} = \mathbf{W}$.

Corollary

Algorithm 3 achieves accuracy ε after $O\left((LR^2\chi_{1,2})/\varepsilon\right)$ communication and computation steps, where $\chi_1 = \chi$ corresponds to a single-step communication protocol and $\chi_2 = \sqrt{\chi}$ is achieved in the multi-step case (Chebyshev acceleration).

Result for Euclidean setup

Let each f_i be L-smooth w.r.t. $\|\cdot\|_2$ and $\mathbf{W_r} = \mathbf{W_p} = \mathbf{W}$.

Corollary

Algorithm 3 achieves accuracy ε after $O\left((LR^2\chi_{1,2})/\varepsilon\right)$ communication and computation steps, where $\chi_1 = \chi$ corresponds to a single-step communication protocol and $\chi_2 = \sqrt{\chi}$ is achieved in the multi-step case (Chebyshev acceleration).

Mirror-Prox is optimal is the non-strongly-convex-concave case!

Euclidean setup, strongly convex-concave case

Let each f_i be μ-strongly convex-concave. w.r.t. $\|\cdot\|_2$ and $\mathbf{W_r} = \mathbf{W_p} = \mathbf{W}$.

Euclidean setup, strongly convex-concave case

Let each f_i be μ-strongly convex-concave. w.r.t. $\|\cdot\|_2$ and $\mathbf{W_r} = \mathbf{W_p} = \mathbf{W}$.For a given accuracy ε we introduce $\alpha = \frac{\varepsilon \lambda_{\min}^{+}(\mathbf{W})}{8m(LR^2)^2}$ and consider problem

$$\min_{\substack{\mathbf{x}\in\mathcal{X},\mathbf{p}\in\mathcal{P} \\ \mathbf{u}\in\mathbb{R}^{md_r}}} \max_{\substack{\mathbf{y}\in\mathcal{Y},\mathbf{r}\in\mathcal{R} \\ \mathbf{z}\in\mathbb{R}^{md_p}}} F(\mathbf{x},\mathbf{p},\mathbf{y},\mathbf{r}) + \langle \mathbf{u}, \mathbf{W_r r}\rangle + \langle \mathbf{z}, \mathbf{W_p p}\rangle + \frac{\alpha}{2}\|\mathbf{u}\|_2^2 - \frac{\alpha}{2}\|\mathbf{z}\|_2^2, \quad (4)$$

which is strongly-convex-strongly-concave in $(\mathbf{u}, \mathbf{z})$.

Euclidean setup, strongly convex-concave case

Let each f_i be μ-strongly convex-concave. w.r.t. $\|\cdot\|_2$ and $\mathbf{W_r} = \mathbf{W_p} = \mathbf{W}$.For a given accuracy ε we introduce $\alpha = \frac{\varepsilon\lambda_{\min}^+(\mathbf{W})}{8m(LR^2)^2}$ and consider problem

$$\min_{\substack{\mathbf{x}\in\mathcal{X},\mathbf{p}\in\mathcal{P}\\ \mathbf{u}\in\mathbb{R}^{md_r}}} \max_{\substack{\mathbf{y}\in\mathcal{Y},\mathbf{r}\in\mathcal{R}\\ \mathbf{z}\in\mathbb{R}^{md_p}}} F(\mathbf{x},\mathbf{p},\mathbf{y},\mathbf{r}) + \langle \mathbf{u}, \mathbf{W_r}\mathbf{r}\rangle + \langle \mathbf{z}, \mathbf{W_p}\mathbf{p}\rangle + \frac{\alpha}{2}\|\mathbf{u}\|_2^2 - \frac{\alpha}{2}\|\mathbf{z}\|_2^2, \tag{4}$$

which is strongly-convex-strongly-concave in $(\mathbf{u},\mathbf{z})$.The choice of α guarantees that $(\varepsilon/2)$-solution of (4) is an ε-solution of non-regularized problem (3).

Euclidean setup, strongly convex-concave case

Let each f_i be μ-strongly convex-concave. w.r.t. $\|\cdot\|_2$ and $\mathbf{W_r} = \mathbf{W_p} = \mathbf{W}$.For a given accuracy ε we introduce $\alpha = \frac{\varepsilon \lambda^+_{\min}(\mathbf{W})}{8m(LR^2)^2}$ and consider problem

$$\min_{\substack{\mathbf{x}\in\mathcal{X},\mathbf{p}\in\mathcal{P}\\ \mathbf{u}\in\mathbb{R}^{md_r}}} \max_{\substack{\mathbf{y}\in\mathcal{Y},\mathbf{r}\in\mathcal{R}\\ \mathbf{z}\in\mathbb{R}^{md_p}}} F(\mathbf{x},\mathbf{p},\mathbf{y},\mathbf{r}) + \langle \mathbf{u}, \mathbf{W_r}\mathbf{r}\rangle + \langle \mathbf{z}, \mathbf{W_p}\mathbf{p}\rangle + \frac{\alpha}{2}\|\mathbf{u}\|_2^2 - \frac{\alpha}{2}\|\mathbf{z}\|_2^2, \tag{4}$$

which is strongly-convex-strongly-concave in $(\mathbf{u},\mathbf{z})$.The choice of α guarantees that $(\varepsilon/2)$-solution of (4) is an ε-solution of non-regularized problem (3).

Theorem

Let $\mathbf{W_p} = \mathbf{W_r} = \mathbf{W}$. *Mirror-Prox requires* $N = O(\max(L/\mu, (LR^2)^2\chi_{1,2}/\varepsilon)\log(R^2/(m\varepsilon)))$ *communication and computation steps to achieve* ε*-accuracy, with* $\chi_1 = \chi$ *in single-step and* $\chi_2 = \sqrt{\chi}$ *in multi-step scenarios, correspondingly.*

Euclidean setup, strongly convex-concave case

Computational and oracle complexities can be separated by a *sliding* technique. Let $g(\zeta) = A(\zeta) + B(\zeta)$,

Euclidean setup, strongly convex-concave case

Computational and oracle complexities can be separated by a *sliding* technique. Let $g(\zeta) = A(\zeta) + B(\zeta)$,where

$$A(\zeta) = F(\mathbf{x}, \mathbf{p}, \mathbf{y}, \mathbf{r}),\ \ B(\zeta) = \langle \mathbf{u}, \mathbf{W_r r} \rangle + \langle \mathbf{z}, \mathbf{W_p p} \rangle + \alpha/2 \left\| u \right\|_2^2 - \alpha/2 \left\| z \right\|_2^2.$$

Euclidean setup, strongly convex-concave case

Computational and oracle complexities can be separated by a *sliding* technique. Let $g(\zeta) = A(\zeta) + B(\zeta)$,where

$$A(\zeta) = F(\mathbf{x}, \mathbf{p}, \mathbf{y}, \mathbf{r}),\;\; B(\zeta) = \langle \mathbf{u}, \mathbf{W_r r} \rangle + \langle \mathbf{z}, \mathbf{W_p p} \rangle + \alpha/2 \left\| u \right\|_2^2 - \alpha/2 \left\| z \right\|_2^2 .$$

Algorithm 8 Sliding

Require: Initial guess $x^0 \in Q$, step-size $\eta > 0$.

1: **for** $k = 0, 1, 2, \dots$ **do**

2: $\quad \nu^k = \zeta^k - \eta A(\zeta^k)$

3: $\quad$ Find $\theta^k \in Q$, such that $\theta^k \approx \hat{\theta}^k$, where $\hat{\theta}^k \in Q$ is a solution to variational inequality (for all $\zeta \in Q$):

$$\left\langle \eta B(\hat{\theta}^k) + \hat{\theta}^k - \nu^k, \zeta - \hat{\theta}^k \right\rangle \geq 0. \tag{5}$$

4: $\quad \omega^k = \theta^k + \eta(A(\zeta^k) - A(\theta^k))$

5: $\quad \zeta^{k+1} = \mathrm{Proj}_Q(\omega^k)$

6: **end for**

Euclidean setup, strongly convex-concave case

Theorem

For achieving ε-accuracy, Algorithm 6 requires
$N_{comp} = O\left((L/\mu)\log(R_\zeta^2/m\varepsilon)\right)$, *computation and*
$N_{comm} = O\left(((LR^2)^2/\varepsilon)\chi_{1,2}\log(1/\delta)\log(R_\zeta^2/m\varepsilon)\right)$ *communication steps, where* $\chi_1 = \chi$ *corresponds to single-step protocol and* $\chi_2 = \sqrt{\chi}$ *to the multi-step one.*

Euclidean setup, strongly convex-concave case

Theorem

For achieving ε-accuracy, Algorithm 6 requires
$N_{comp} = O\left((L/\mu)\log(R_\zeta^2/m\varepsilon)\right)$*, computation and*
$N_{comm} = O\left(((LR^2)^2/\varepsilon)\chi_{1,2}\log(1/\delta)\log(R_\zeta^2/m\varepsilon)\right)$ *communication steps, where* $\chi_1 = \chi$ *corresponds to single-step protocol and* $\chi_2 = \sqrt{\chi}$ *to the multi-step one.*

- Computation and communication complexities separated.

Euclidean setup, strongly convex-concave case

Theorem

For achieving ε-accuracy, Algorithm 6 requires
$N_{comp} = O\left((L/\mu)\log(R_\zeta^2/m\varepsilon)\right)$*, computation and*
$N_{comm} = O\left(((LR^2)^2/\varepsilon)\chi_{1,2}\log(1/\delta)\log(R_\zeta^2/m\varepsilon)\right)$ *communication steps, where* $\chi_1 = \chi$ *corresponds to single-step protocol and* $\chi_2 = \sqrt{\chi}$ *to the multi-step one.*

- Computation and communication complexities separated.
- Optimal in the number of oracle calls.

Euclidean setup, strongly convex-concave case

Theorem

For achieving ε-accuracy, Algorithm 6 requires
$N_{comp} = O\left((L/\mu)\log(R_\zeta^2/m\varepsilon)\right)$*, computation and*
$N_{comm} = O\left(((LR^2)^2/\varepsilon)\chi_{1,2}\log(1/\delta)\log(R_\zeta^2/m\varepsilon)\right)$ *communication steps, where* $\chi_1 = \chi$ *corresponds to single-step protocol and* $\chi_2 = \sqrt{\chi}$ *to the multi-step one.*

- Computation and communication complexities separated.
- Optimal in the number of oracle calls.
- Not optimal in number of communication rounds.

Numerical tests

We compare against IBP algorithm [Benamou et al., 2015] on the decentralized WB computation problem. Mirror-Prox (Algorithm 3) shows a more stable performance.

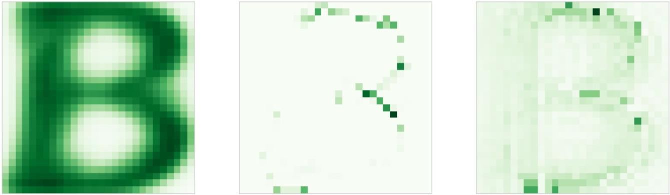

Figure: WB of letter 'B' found by DMP (left), IBP with $\gamma = 10^{-4}$ (middle) and $\gamma = 10^{-5}$ (right).

Conclusion

- Saddle-point problems with local (individual) and global (common) variables.

Conclusion

- Saddle-point problems with local (individual) and global (common) variables.
- Lagrange reformulation of the constraints allows to apply Mirror-Prox and obtain results immediately.

Conclusion

- Saddle-point problems with local (individual) and global (common) variables.
- Lagrange reformulation of the constraints allows to apply Mirror-Prox and obtain results immediately.
- Optimal but still simple algorithm for the Euclidean convex-concave case.

Conclusion

- Saddle-point problems with local (individual) and global (common) variables.
- Lagrange reformulation of the constraints allows to apply Mirror-Prox and obtain results immediately.
- Optimal but still simple algorithm for the Euclidean convex-concave case.
- Splitting oracle and communication complexities in the strongly-convex-concave setup.

References I

Benamou, J.-D., Carlier, G., Cuturi, M., Nenna, L., and Peyré, G. (2015).
Iterative bregman projections for regularized transportation problems.
SIAM Journal on Scientific Computing, 37(2):A1111–A1138.

Dvinskikh, D. and Tiapkin, D. (2020).
Improved complexity bounds in the wasserstein barycenter problem.
arXiv preprint arXiv:2010.04677.

Jambulapati, A., Sidford, A., and Tian, K. (2019).
A direct tilde {O}(1/epsilon) iteration parallel algorithm for optimal transport.
Advances in Neural Information Processing Systems, 32:11359–11370.

References II

Mateos-Núnez, D. and Cortés, J. (2015).
Distributed subgradient methods for saddle-point problems.
In *2015 54th IEEE Conference on Decision and Control (CDC)*, pages 5462–5467. IEEE.

Nedic, A. and Ozdaglar, A. (2010).
Cooperative distributed multi-agent optimization.
Convex optimization in signal processing and communications, 340.