Decentralized optimization for saddle point problems with local and global variables

Optimization without borders

Alexander Rogozin

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

July 17, 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.
- Optimal algorithm for decentralized convex-concave saddle-point problems in Eucledian setup.

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: 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: 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: 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: 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: 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: 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 $y_i \in [-1, 1]^{2n}$, $i = 1, \dots, m$.

Example: 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 $y_i \in [-1, 1]^{2n}$, $i = 1, \dots, m$.

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

Problem statement

We study a saddle-point problem of the form

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

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 $x_i \in \mathcal{X}_i$, $p \in \mathcal{P}$.

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, $\text{Ker}\,\tilde{W} = \text{span}\,\{\mathbf{1}\}$.

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, $\mathrm{Ker}\, \tilde{W} = \mathrm{span}\, \{\mathbf{1}\}$.

An example of matrix satisfying this Assumption is the graph *Laplacian matrix* $\tilde{W} \in \mathbb{R}^{m \times m}$ 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, $\mathrm{Ker}\, \tilde{W} = \mathrm{span}\,\{\mathbf{1}\}$.

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

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

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 [Rogozin et al., 2021]

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

where γ_r and γ_p are arbitrary positive scalars.

Result for general proximal setup

We assume that a vector field associated with saddle-point problem (3) is L_ζ-Lipschitz w.r.t. corresponding norm.

Theorem

Let $(\hat{\mathbf{x}}^N, \hat{\mathbf{p}}^N, \hat{\mathbf{y}}^N, \hat{\mathbf{r}}^N, \hat{\mathbf{u}}^N, \hat{\mathbf{z}}^N)$ *be the output of Mirror-Prox 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 Mirror-Prox with stepsize* $\alpha = 1/L_\zeta$ *we have*

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

.

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

- *Convex-concave case: Mirror-Prox achieves accuracy ε after $O\left((LR^2\sqrt{\chi})/\varepsilon\right)$ communication and $O(\left(LR^2\right)/\varepsilon)$ computation steps.*
- *μ-strongly-convex-strongly-concave case: Mirror-Prox requires $N = O(\max(L/\mu, (LR^2)^2\sqrt{\chi}/\varepsilon)\log(R^2/(m\varepsilon)))$ communication and computation steps to achieve ε-accuracy (with a correct problem regularization).*

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

- *Convex-concave case: Mirror-Prox achieves accuracy ε after $O\left((LR^2\sqrt{\chi})/\varepsilon\right)$ communication and $O(\left(LR^2\right)/\varepsilon)$ computation steps.*
- *μ-strongly-convex-strongly-concave case: Mirror-Prox requires $N = O(\max(L/\mu, (LR^2)^2\sqrt{\chi}/\varepsilon)\log(R^2/(m\varepsilon)))$ communication and computation steps to achieve ε-accuracy (with a correct problem regularization).*

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

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

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

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

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

Algorithm 3 Sliding

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

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

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

3: 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 $\zeta \in Q$):

$$\left\langle \eta B(\hat{\theta}^k) + \hat{\theta}^k - \nu^k, \zeta - \hat{\theta}^k \right\rangle \geq 0. \qquad (4)$$

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

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

6: **end for**

Euclidean setup, strongly convex-concave case

Theorem

For achieving ε-accuracy, Algorithm 1 requires
$N_{comp} = O\left((L/\mu)\log(R_\zeta^2/m\varepsilon)\right)$, *computation and*
$N_{comm} = O\left(((LR^2)^2/\varepsilon)\sqrt{\chi}\log(1/\delta)\log(R_\zeta^2/m\varepsilon)\right)$ *communication steps.*

Euclidean setup, strongly convex-concave case

Theorem

For achieving ε-accuracy, Algorithm 1 requires
$N_{comp} = O\left((L/\mu)\log(R_\zeta^2/m\varepsilon)\right)$, *computation and*
$N_{comm} = O\left(((LR^2)^2/\varepsilon)\sqrt{\chi}\log(1/\delta)\log(R_\zeta^2/m\varepsilon)\right)$ *communication steps.*

- Computation and communication complexities separated.

Euclidean setup, strongly convex-concave case

Theorem

For achieving ε-accuracy, Algorithm 1 requires
$N_{comp} = O\left((L/\mu)\log(R_\zeta^2/m\varepsilon)\right)$, *computation and*
$N_{comm} = O\left(((LR^2)^2/\varepsilon)\sqrt{\chi}\log(1/\delta)\log(R_\zeta^2/m\varepsilon)\right)$ *communication steps.*

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

Euclidean setup, strongly convex-concave case

Theorem

For achieving ε-accuracy, Algorithm 1 requires
$N_{comp} = O\left((L/\mu)\log(R_\zeta^2/m\varepsilon)\right)$, *computation and*
$N_{comm} = O\left(((LR^2)^2/\varepsilon)\sqrt{\chi}\log(1/\delta)\log(R_\zeta^2/m\varepsilon)\right)$ *communication steps.*

- 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 shows a more stable performance.

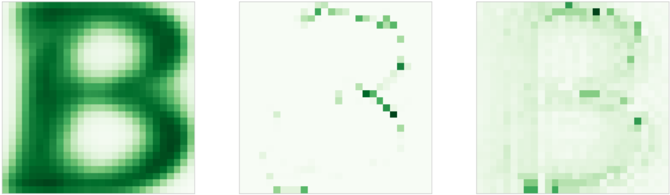

Figure: WB of letter 'B' found by DMP (left), IBP with $\gamma = 10^{-4}$ (middle) and $\gamma = 10^{-5}$ (right).

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.
- Simple optimal 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.
- Simple optimal 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.

Rogozin, A., Beznosikov, A., Dvinskikh, D., Kovalev, D., Dvurechensky, P., and Gasnikov, A. (2021).
Decentralized distributed optimization for saddle point problems.
arXiv preprint arXiv:2102.07758.