Accelerated Primal-Dual Methods for Convex-Strongly-Concave Saddle Point Problems

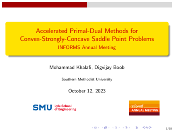
INFORMS Annual Meeting

Mohammad Khalafi, Digvijay Boob

Southern Methodist University

October 17, 2023

SMU Lyle School of Engineering



Outline

1. Problem Definition
2. Assumptions
3. Motivation of Study
4. Linearized Primal-Dual method
5. Accelerated LPD (ALPD)
6. Inexact ALPD
7. Numerical Experiments
8. Conclusion

Problem Definition and Important Measures

- Saddle Point Problem:

$$\mathcal{L}(x,y) := \min_{x\in X}\max_{y\in Y} f(x) + \phi(x,y) - g(y). \tag{1}$$

- Convergence Criterion $\bar{z} = (\bar{x}, \bar{y})$:

$$Gap(\bar{z}) = \max_{z\in X\times Y}\{Q(\bar{z},z) := \mathcal{L}(\bar{x},y) - \mathcal{L}(x,\bar{y})\}.$$

  - $Q(\bar{z}, z)$ is the Primal-Dual gap function.
  - $\bar{z}$ is an ϵ-approximate solution if $Gap(\bar{z}) \leq \epsilon$.

Assumptions

$\phi(\cdot, y)$ is L_{xx}-smooth, $\phi(x, \cdot)$ is L_{yy}-smooth and ϕ is L_{xy}-smooth, if the followings hold for all $x, x' \in X,\ y, y' \in Y$ respectively:

$$\begin{aligned}
\|\nabla_x \phi(x', y) - \nabla_x \phi(x, y)\| &\leq L_{xx}\|x' - x\|, \\
\|\nabla_y \phi(x, y') - \nabla_y \phi(x, y)\| &\leq L_{yy}\|y' - y\|, \\
\|\nabla_y \phi(x', y) - \nabla_y \phi(x, y)\| &\leq L_{xy}\|x' - x\|.
\end{aligned}$$

Motivation of Study

In many problems, the following function is a nonsmooth function which is hard to optimize.

$$P(x) : f(x) + \max_{y \in Y} \phi(x, y). \tag{2}$$

1. One way to smoothen this function is to use Nesterov's smoothing technique. This technique involves subtracting a strongly convex regularizing function, resulting in a convex-strongly-concave SPP.
2. We assume that $f(x)$ is an easy function to evaluate. This might not be true in many cases. Hence, linearization of f might be a good approach to handle this problem.
3. A popular approach is using a linearized primal-dual method (LPD).

	Coupling	Linearizing f	Gradient Complexity	
			$\mu_f > 0$	$\mu_g > 0$
[Chambolle and Pock(2011)]	bilinear	No	$\mathcal{O}(\frac{1}{\sqrt{\epsilon}})$	NA
[Chambolle and Pock(2016)]	bilinear	Yes	$\mathcal{O}(\frac{1}{\sqrt{\epsilon}})$	NA
[Hamedani and Aybat(2021)]	semi-linear	No	$\mathcal{O}(\frac{1}{\sqrt{\epsilon}})$	NA
LPD (Algorithm 1)	bilinear	Yes	$\mathcal{O}(\frac{1}{\sqrt{\epsilon}})$	$\mathcal{O}(\frac{L_f}{\epsilon} + \frac{\\|A\\|}{\sqrt{\mu_g \epsilon}})$
ALPD (Algorithm 2)	semi-linear	Yes	NA	$\mathcal{O}(\sqrt{\frac{L_f+L_{yy}}{\epsilon}} + \frac{L_{xy}}{\sqrt{\mu_g \epsilon}})$
	general			$\mathcal{O}(\sqrt{\frac{L_f+L_{yy}}{\epsilon}} + \frac{L_{xy}}{\sqrt{\mu_g \epsilon}} + \frac{L_{xx}}{\epsilon})$
Inexact ALPD (Algorithm 3)	general	Yes	NA	For $\nabla f, \nabla_y \phi$: $\mathcal{O}(\sqrt{\frac{L_f+L_{yy}}{\epsilon}})$ For $\nabla_x \phi$: $\mathcal{O}(\frac{\sqrt{L_{xx}}}{\epsilon^{3/4}} \log(\frac{1}{\epsilon}))$

Table: Comparison of our work. Gradient complexity is for obtaining an ϵ error in gap function.

Linearized Primal-Dual method

Algorithm Linearized PD (LPD) method

1: **Initialize** $\tilde{x}_1 = x_1 \in X,\ \ y_1 \in Y$
2: **for** $t = 1, \ldots, K$ **do**
3: $\quad y_{t+1} \leftarrow \arg\min_{y \in Y} \langle -A\tilde{x}_t, y\rangle + g(y) + \frac{1}{2\tau_t}\|y - y_t\|^2$
4: $\quad x_{t+1} \leftarrow \arg\min_{x \in X} \langle \nabla f(x_t) + A^\top y_{t+1}, x\rangle + \frac{1}{2\eta_t}\|x - x_t\|^2$
5: $\quad \tilde{x}_{t+1} \leftarrow x_{t+1} + \theta_t(x_{t+1} - x_t)$
6: **end for**
7: **return** $\bar{x}_{K+1} = \frac{\sum_{t=1}^K \gamma_{t+1} x_{t+1}}{\sum_{t=1}^K \gamma_{t+1}}, \bar{y}_{K+1} = \frac{\sum_{t=1}^K \gamma_{t+1} y_{t+1}}{\sum_{t=1}^K \gamma_{t+1}}$

Convergence analysis of LPD

Theorem

For a μ_f-strongly-convex-concave bilinear SPP, LPD has the optimal convergence rate of $\mathcal{O}(\frac{L_f+\|A\|^2}{K^2})$, and for a μ_g-strongly-concave-convex bilinear SPP, it has convergence rate of $\mathcal{O}(\frac{L_f}{K}+\frac{\|A\|^2}{K^2})$ where f is L_f-smooth.

- **Observation**: Strong concavity can not handle the errors caused by the linearization of f.

Accelerated LPD (ALPD)

Algorithm Accelerated Linearized PD (ALPD) method

Initialize $\bar{x}_1 = x_0 = x_1 \in X, \bar{y}_1 = y_0 = y_1 \in Y$
for $t = 1, \ldots, K$ **do**
  $\underline{x}_t \leftarrow (1 - \beta_t^{-1})\bar{x}_t + \beta_t^{-1} x_t$
  $v_t \leftarrow (1 + \theta_t)\nabla_y \phi(x_t, y_t) - \theta_t \nabla_y \phi(x_{t-1}, y_{t-1})$
  $y_{t+1} \leftarrow \arg\min_{y \in Y} \langle -v_t + \nabla g(y_t), y \rangle + \frac{1}{2\tau_t}\|y - y_t\|^2$
  $x_{t+1} \leftarrow \arg\min_{x \in X} \langle \nabla f(\underline{x}_t) + \nabla_x \phi(x_t, y_{t+1}), x \rangle + \frac{1}{2\eta_t}\|x - x_t\|^2$
  $\bar{x}_{t+1} = (1 - \beta_t^{-1})\bar{x}_t + \beta_t^{-1} x_{t+1}$
  $\bar{y}_{t+1} = (1 - \beta_t^{-1})\bar{y}_t + \beta_t^{-1} y_{t+1}$
end for
return $\bar{x}_{K+1}, \bar{y}_{K+1}$

Convergence rates of ALPD for semi-linear and nonlinear coupling

Theorem

- *Case 1: Semi-linear ϕ with $L_{xx} = 0$:*

$$\max_{z \in X \times Y} \{Q(\bar{z}_{K+1})\} = \mathcal{O}\left(\frac{L_f + L_{yy}}{K^2} + \frac{L_{xy}^2}{\mu_g K^2}\right)$$

- *Case 2: nonlinear ϕ with $L_{xx} > 0$:*

$$\max_{z \in X \times Y} \{Q(\bar{z}_{K+1})\} = \mathcal{O}\left(\frac{L_f + L_{yy}}{K^2} + \frac{L_{xy}^2}{\mu_g K^2} + \frac{L_{xx}}{K}\right)$$

- ALPD is not optimal at full nonlinearity.

Inexact ALPD

Algorithm Inexact ALPD Method

Initialize $\bar{x}_1 = x_0 = x_1 \in X, \bar{y}_1 = y_0 = y_1 \in Y$

for $t = 1, \ldots, K$ **do**

$\quad \underline{x}_t \leftarrow (1 - \beta_t^{-1})\bar{x}_t + \beta_t^{-1} x_t$

$\quad v_t \leftarrow (1 + \theta_t)\nabla_y \phi(x_t, y_t) - \theta_t \nabla_y \phi(x_{t-1}, y_{t-1})$

$\quad y_{t+1} \leftarrow \arg\min_{y \in Y} \langle -v_t + \nabla g(y_t), y \rangle + \frac{1}{2\tau_t}\|y - y_t\|^2$

$\quad x_{t+1}$ is a δ_t-approximate solution of the problem:

$$\min_{x \in X} \langle \nabla f(\underline{x}_t), x \rangle + \phi(x, y_{t+1}) + \frac{1}{2\eta_t}\|x - x_t\|^2$$

$\quad \bar{x}_{t+1} \leftarrow (1 - \beta_t^{-1})\bar{x}_t + \beta_t^{-1} x_{t+1}$

$\quad \bar{y}_{t+1} \leftarrow (1 - \beta_t^{-1})\bar{y}_t + \beta_t^{-1} y_{t+1}$

end for

return $\bar{x}_{K+1}, \bar{y}_{K+1}$

Complexity analysis of inexact ALPD

Theorem

Inexact ALPD requires $\mathcal{O}(\sqrt{\frac{L_f+L_{yy}}{\epsilon}})$ *gradient evaluation of* ∇f *and* $\nabla_y \phi$, *and requires* $\mathcal{O}(\frac{\sqrt{L_{xx}}}{\epsilon^{3/4}}\log(\frac{1}{\epsilon})) = \tilde{\mathcal{O}}(\frac{\sqrt{L_{xx}}}{\epsilon^{3/4}})$ *gradient evaluation of* $\nabla_x \phi$. *Hence, the gradient complexity of* $\nabla_x \phi$ *improves significantly (c.f.* $\mathcal{O}(\frac{L_{xx}}{\epsilon})$ *gradient complexity in ALPD).*

Numerical Experiment: ALPD vs. LPD

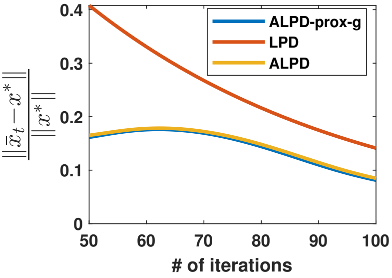
The ℓ_q-norm penalty problem with linear constraints is

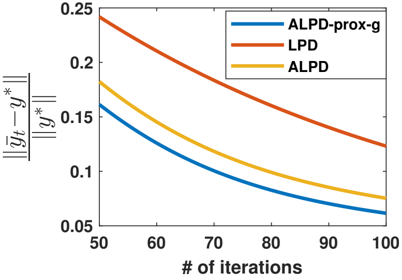
$$\min_{x\in X} f(x) + \rho\|Ax - b\|_q \equiv \min_{x\in X} \max_{\|y\|_p\leq 1} f(x) + \rho\langle y, Ax - b\rangle,$$

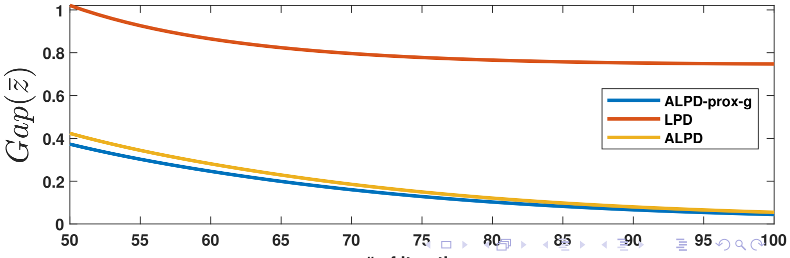
Smooth approximation of the nonsmooth penalty term using Nesterov's smoothing technique:

$$\min_{x\in X} \max_{\|y\|_p\leq 1} \{f(x) + \rho\langle y, Ax - b\rangle - \tfrac{\mu_g}{2}\|y\|^2\}, \tag{3}$$

Numerical Experiment: ALPD vs. LPD







Numerical Experiment: ALPD vs. Inexact ALPD

- Consider a penalty problem with non-linear constraints.
- The corresponding coupling function in SPP becomes nonlinear ($L_{xx} > 0$)

Numerical Experiment: ALPD vs. Inexact ALPD

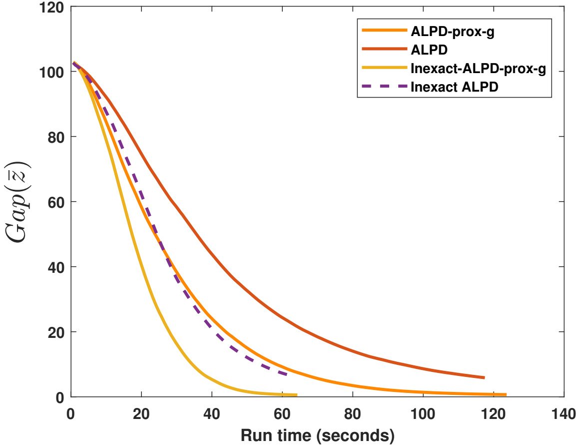


Figure: Comparison of the ALPD and inexact ALPD method and their prox-g variants using the Gap function vs run-time (seconds) plot for 10 i.i.d. instances.

- Investigate the SPPs without strong convexity assumptions in primal function.
- Obtaining optimal convergence rate for semi-linear coupling functions.
- Designing simple and easy-to-implement algorithms for general SPPs.
- Improving computational complexity for non-linear coupling functions.

References

Antonin Chambolle and Thomas Pock.
A first-order primal-dual algorithm for convex problems with applications to imaging.
Journal of mathematical imaging and vision, 40(1):120–145, 2011.

Antonin Chambolle and Thomas Pock.
On the ergodic convergence rates of a first-order primal–dual algorithm.
Mathematical Programming, 159(1):253–287, 2016.

Erfan Yazdandoost Hamedani and Necdet Serhat Aybat.
A primal-dual algorithm with line search for general convex-concave saddle point problems.
SIAM Journal on Optimization, 31(2):1299–1329, 2021.

- Thank you!
- Questions?