Primal-dual proximal bundle and conditional gradient methods for convex problems

Jiaming Liang

Department of Computer Science & Goergen Institute for Data Science
University of Rochester

October 26, 2025

INFORMS Annual Meeting 2025, Atlanta, GA

Overview

Objectives: to develop primal-dual algorithmic frameworks, i.e., the inexact proximal point framework, that systematically guides the design and analysis of algorithms for convex problems such as convex optimization and convex-concave saddle-point problems.

- Primal-dual convergence of the proximal bundle (PB) method for solving convex and nonsmooth optimization;
- Primal-dual interpretation of conditional gradient and the cutting-plane scheme used within PB;
- Optimal PB-type method for convex-concave nonsmooth saddle-point problems.

Review of Proximal Bundle Methods

The problem of interest is

$$\phi_* := \min_{x \in \mathbb{R}^n} \{\phi(x) := f(x) + h(x)\},$$

where f is convex and Lipschitz continuous, and h is a convex with a simple proximal mapping.

Approximately solve the proximal problem

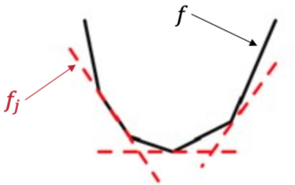
$$\hat{x} := \operatorname{argmin} \left\{ f(x) + h(x) + \frac{1}{2\lambda} \|x - x^c\|^2 \right\}$$

by an iterative process

$$x_j \leftarrow \min \left\{ f_j(x) + h(x) + \frac{1}{2\lambda} \|x - x^c\|^2 \right\}.$$

Recursively build up a cutting-plane model

$$f_j(x) = \max_{0 \le i \le j-1} \{\ell_f(x; x_i) := f(x_i) + \langle f'(x_i), x - x_i \rangle\}$$



Generic Bundle Management

- First complexity result $\mathcal{O}(\bar{\varepsilon}^{-3})$ (Kiwiel, 2000)
- Complexity in the strongly convex case $\mathcal{O}((\mu\bar{\varepsilon})^{-1})$ (Du & Ruszczyński, 2017)
- Optimal complexity $\mathcal{O}(\bar{\varepsilon}^{-2})$ (Liang & Monteiro, 2021, Díaz & Grimmer, 2023)
- A unified analysis for various bundle management (Liang & Monteiro, 2024)

Algorithm Generic Bundle Management, GBM$(\lambda, \tau_j, x_0, x_j, \Gamma_j)$

Initialize: $(\lambda, \tau_j) \in \mathbb{R}_{++} \times [0,1]$, $(x_0, x_j) \in \mathbb{R}^n \times \mathbb{R}^n$, and $\Gamma_j \in \overline{\text{Conv}}\,(\mathbb{R}^n)$ satisfying $\Gamma_j \le f$

- find a bundle model $\Gamma_{j+1} \in \overline{\text{Conv}}\,(\mathbb{R}^n)$ satisfying $\Gamma_{j+1} \le f$ and

$$\Gamma_{j+1}(\cdot) \ge \tau_j \bar{\Gamma}_j(\cdot) + (1-\tau_j)\ell_f(\cdot; x_j),$$

where $\bar{\Gamma}_j \in \overline{\text{Conv}}\,(\mathbb{R}^n)$ satisfies $\bar{\Gamma}_j \le f$ and

$$\bar{\Gamma}_j(x_j) = \Gamma_j(x_j), \quad x_j = \underset{u \in \mathbb{R}^n}{\operatorname{argmin}} \left\{ \bar{\Gamma}_j(u) + h(u) + \frac{1}{2\lambda}\|u - x_0\|^2 \right\}.$$

Output: Γ_{j+1}.

Three examples

(E1) **single cut update:** $\Gamma^+ = \Gamma^+_\tau := \tau\Gamma + (1-\tau)\ell_f(\cdot; x)$.

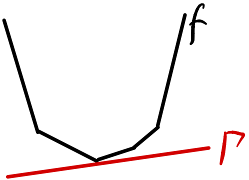
(E2) **two cuts update:** assume $\Gamma = \max\{A_f, \ell_f(\cdot; x^-)\}$ where A_f is an affine function satisfying $A_f \le f$, set

$$\Gamma^+ = \max\{A_f^+, \ell_f(\cdot; x)\}$$

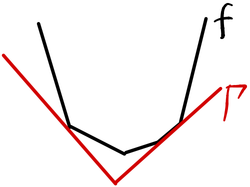
where $A_f^+ = \theta A_f + (1-\theta)\ell_f(\cdot; x^-)$.

(E3) **multi cuts update:** $\Gamma^+ = \max\{\Gamma, \ell_f(\cdot; x)\}$.

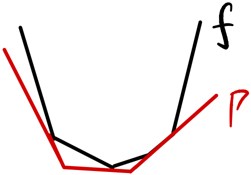
Bundle of past information $\{(x_i, f(x_i), f'(x_i))\}$



Single cut



Two cuts



Multiple cuts

Outline

1. Primal-dual Convergence
2. Duality between Cutting-plane and Conditional Gradient Methods
3. Saddle-point Problem
4. Numerical Experiments

Outline

Primal-dual Cutting-plane

The idea of the proximal bundle method is to recursively solve

$$\min_{u\in\mathbb{R}^n} \left\{ \phi^\lambda(u) := \phi(u) + \frac{1}{2\lambda}\|u - \hat{x}_{k-1}\|^2 \right\}$$

via the following recursive subproblem solver.

Algorithm Primal-Dual Cutting-Plane, PDCP$(x_0, \lambda, \varepsilon)$

---

1. compute

$$x_j = \underset{u\in\mathbb{R}^n}{\operatorname{argmin}} \left\{ \Gamma_j(u) + h(u) + \frac{1}{2\lambda}\|u - x_0\|^2 \right\}, \quad s_j \in \partial\Gamma_j(x_j) \cap (-\partial h^\lambda(x_j)),$$

choose $\tilde{x}_{j+1}$ such that

$$\phi^\lambda(\tilde{x}_{j+1}) \le \tau_j \phi^\lambda(\tilde{x}_j) + (1-\tau_j)\phi^\lambda(x_{j+1})$$

and set

$$m_j = \min_{u\in\mathbb{R}^n} \left\{ \Gamma_j(u) + h(u) + \frac{1}{2\lambda}\|u - x_0\|^2 \right\}, \quad t_j = \phi^\lambda\left(\tilde{x}_j\right) - m_j.$$

2. select $\tau_j \in [0,1]$ and update Γ_{j+1} by GBM$(\lambda, \tau_j, x_0, x_j, \Gamma_j)$ and $j \leftarrow j+1$;

Output: $(x_j, \tilde{x}_j, s_j)$ when $t_j \le \varepsilon$.

Primal-dual convergence

We can show that t_j upper bounds the primal-dual gap of the proximal subproblem and converges at rate of $\mathcal{O}(1/j)$

$$\phi^\lambda\left(\tilde{x}_j\right) + f^*(s_j) + (h^\lambda)^*(-s_j) \le t_j \le \frac{2t_1}{j(j+1)} + \frac{16\lambda M^2}{j+1}.$$

Algorithm Primal-Dual Proximal Bundle, PDPB$(\hat{x}_0, \lambda, \bar{\varepsilon})$

---

call oracle $(\hat{x}_k, \tilde{x}_k, s_k) =$ PDCP$(\hat{x}_{k-1}, \lambda, \bar{\varepsilon})$ and compute

$$\bar{x}_k = \frac{1}{k}\sum_{i=1}^{k} \tilde{x}_i, \quad \bar{s}_k = \frac{1}{k}\sum_{i=1}^{k} s_i. \tag{1}$$

To find $(\bar{x}_k, \bar{s}_k)$ such that the primal-dual gap is bounded by $\bar{\varepsilon}$

$$\hat{\phi}(\bar{x}_k) + f^*(\bar{s}_k) + \hat{h}^*(-\bar{s}_k) \le \bar{\varepsilon},$$

it takes $\mathcal{O}(M^2 d_0^2/\bar{\varepsilon}^2)$ iterations, where $d_0 := \min_{x_* \in X_*} \{\|x_* - \hat{x}_0\|\}$.

Outline

1. Primal-dual Convergence
2. Duality between Cutting-plane and Conditional Gradient Methods
3. Saddle-point Problem
4. Numerical Experiments

Dual viewpoint

Primal and dual proximal subproblems: f is convex and Lipschitz continuous, and h is convex

$$\min_{x\in\mathbb{R}^n} \left\{ f(x) + h(x) + \frac{1}{2\lambda}\|x - x_0\|^2 = f(x) + (h^\lambda)(x) \right\}, \quad \text{(P)}$$

$$\min_{z\in\mathbb{R}^n} \left\{ (h^\lambda)^*(-z) + f^*(z) \right\}. \quad \text{(D)}$$

Algorithm Regularized Cutting Plane for (P)

$$x_j = \operatorname*{argmin}_{x\in\mathbb{R}^n} \left\{ \Gamma_j(x) + (h^\lambda)(x) \right\}, \quad s_j \in \partial\Gamma_j(x_j) \cap (-\partial h^\lambda(x_j)),$$

$$\Gamma_{j+1}(\cdot) = \tau_j \Gamma_j(\cdot) + (1-\tau_j)\ell_f(\cdot; x_j), \quad \tau_j = \frac{j}{j+2}.$$

Algorithm Conditional Gradient for (D)

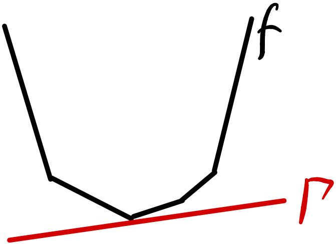
$$\bar{z}_j = \operatorname*{argmin}_{z\in\mathbb{R}^n} \left\{ \langle -\nabla(h^\lambda)^*(-z_j), z\rangle + f^*(z) \right\},$$

$$z_{j+1} = \tau_j z_j + (1-\tau_j)\bar{z}_j.$$

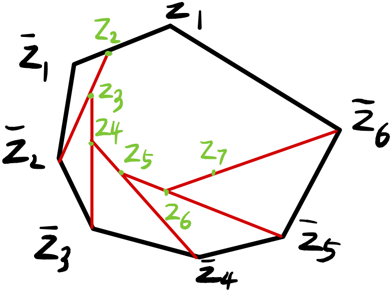
Equivalence $s_j = z_j$, $x_j = \nabla(h^\lambda)^*(-z_j)$, and $f'(x_j) = \bar{z}_j$.

Algorithmic duality visualized

Similar duality relationships have been discovered for CG and mirror descent (Bach, 2015), CG and dual averaging (Grigas, 2016), stochastic CG and randomized coordinate MD (Lu & Freund, 2021), and CG and PB (Fersztand & Sun, 2024).



Single cut



Standard CG

Equivalence $s_j = z_j$, $x_j = \nabla(h^\lambda)^*(-z_j)$, and $f'(x_j) = \bar{z}_j$.

GBM implementations inspired by CG

Algorithm Generic Bundle Management

Initialize: $(\lambda, \tau_j) \in \mathbb{R}_{++} \times [0,1]$, $(x_0, x_j) \in \mathbb{R}^n \times \mathbb{R}^n$, and $\Gamma_j \in \overline{\mathrm{Conv}}\,(\mathbb{R}^n)$ satisfying $\Gamma_j \le f$

- find a bundle model $\Gamma_{j+1} \in \overline{\mathrm{Conv}}\,(\mathbb{R}^n)$ satisfying $\Gamma_{j+1} \le f$ and

$$\Gamma_{j+1}(\cdot) \ge \tau_j \bar{\Gamma}_j(\cdot) + (1-\tau_j)\ell_f(\cdot; x_j),$$

where $\bar{\Gamma}_j \in \overline{\mathrm{Conv}}\,(\mathbb{R}^n)$ satisfies $\bar{\Gamma}_j \le f$ and

$$\bar{\Gamma}_j(x_j) = \Gamma_j(x_j), \quad x_j = \underset{u \in \mathbb{R}^n}{\mathrm{argmin}} \left\{ \bar{\Gamma}_j(u) + (h^\lambda)(u) \right\}.$$

Output: Γ_{j+1}.

Previously choice $\tau_j = \frac{j}{j+2}$, other options inspired by CG

$$\alpha_j = \max\left\{0, 1 - \frac{S(z_j)}{\lambda \|z_j - \bar{z}_j\|^2}\right\},$$

$$\beta_j = \mathrm{argmin}\left\{\psi(\beta z_j + (1-\beta)\bar{z}_j) : \beta \in [0,1]\right\}.$$

where $\psi(z) = (h^\lambda)^*(-z) + f^*(z)$ is the dual function and $S(z_j) = \phi^\lambda(x_j) + \psi(z_j)$ is the Wolfe gap.

New variants of CG inspired by GBM implementations

General CG: $z_j \in \text{Conv}\{z_1, \bar{z}_1, \ldots, \bar{z}_{j-1}\} = \text{Conv}\{f'(x_0), f'(x_1), \ldots, f'(x_{j-1})\}$

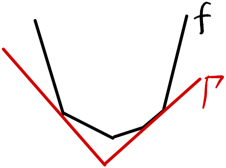
Two-cuts scheme for (P)

$$\Gamma_j(\cdot) = \max\{\bar{\Gamma}_{j-1}(\cdot), \ell_f(\cdot; x_{j-1})\}$$
$$\bar{\Gamma}_j(\cdot) = \theta_{j-1}\bar{\Gamma}_{j-1}(\cdot) + (1-\theta_{j-1})\ell_f(\cdot; x_{j-1})$$
$$\min_{(u,r)\in\mathbb{R}^n\times\mathbb{R}} \left\{ r + h^\lambda(u) : \bar{\Gamma}_{j-1}(u) \le r,\ \ell_f(u, x_{j-1}) \le r \right\}$$

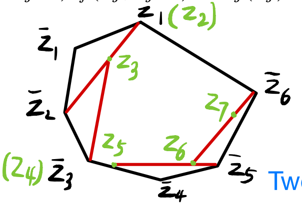
Algorithm Two-cuts scheme-inspied CG for (D)

$$z_j = \theta_{j-1} z_{j-1} + (1-\theta_{j-1})\bar{z}_{j-1} = \nabla\bar{\Gamma}_j,$$
$$\bar{z}_j = \underset{z\in\mathbb{R}^n}{\text{argmin}} \left\{ \langle -\nabla(h^\lambda)^*(-z_j), z\rangle + f^*(z) \right\} = f'(x_j).$$

Solve: $x_j = \nabla(h^\lambda)^*(-z_j)$ and $\theta_{j-1}\bar{\Gamma}_{j-1}(x_j) + (1-\theta_{j-1})\ell_f(x_j; x_{j-1}) = \Gamma_j(x_j)$



Two cuts



Two-cuts CG

New variants of CG inspired by GBM implementations

Multiple-cuts scheme for (P)

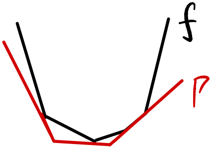
$$\Gamma_j(\cdot) = \max\left\{\ell_f(\cdot; x_i) : i \in I_j\right\}$$

$$I_{j+1} = \bar{I}_{j+1} \cup \{j\}, \quad \bar{I}_{j+1} = \{i \in I_j : \theta_j^i > 0\}$$

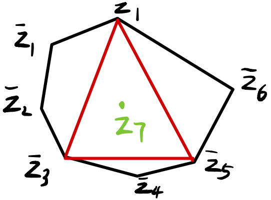
$$\min_{(u,r)\in\mathbb{R}^n\times\mathbb{R}} \left\{r + h^\lambda(u) : \ell_f(u; x_i) \le r,\ \forall i \in I_j\right\}$$

Algorithm Multi-cuts scheme-inspied CG for (D)

$$z_j = \sum_{i\in I_j} \theta_j^i \bar{z}_i, \quad \bar{z}_j = f'(x_j)$$



Multiple cuts



Multi-cuts CG

Outline

1. Primal-dual Convergence
2. Duality between Cutting-plane and Conditional Gradient Methods
3. Saddle-point Problem
4. Numerical Experiments

Saddle-point problem

Convex-concave nonsmooth saddle-point problem

$$\min_{x\in\mathbb{R}^n}\max_{y\in\mathbb{R}^m}\{\phi(x,y):=f(x,y)+h_1(x)-h_2(y)\}$$

$\|f'_x(u,v)\|\le M$, $\|f'_y(u,v)\|\le M$, proximal mappings of h_1 and h_2 are simple, $\mathrm{dom}\, h_1\times\mathrm{dom}\, h_2$ is bounded with finite diameter $D>0$

A pair (x,y) is called a $\bar\varepsilon$-saddle-point if

$$\varphi(x)-\psi(y)\le\bar\varepsilon,$$

where

$$\varphi(u)=\max_{v\in\mathbb{R}^m}\phi(u,v),\quad \psi(v)=\min_{u\in\mathbb{R}^n}\phi(u,v).$$

Composite subgradient method for SPP (CS-SPP)

$$x_k=\operatorname*{argmin}_{u\in\mathbb{R}^n}\left\{\ell_{f(\cdot,y_{k-1})}(u;x_{k-1})+h_1(u)+\frac{1}{2\lambda}\|u-x_{k-1}\|^2\right\},$$

$$y_k=\operatorname*{argmin}_{v\in\mathbb{R}^m}\left\{-\ell_{f(x_{k-1},\cdot)}(v;y_{k-1})+h_2(v)+\frac{1}{2\lambda}\|v-y_{k-1}\|^2\right\}.$$

Letting $\lambda=\bar\varepsilon/(32M^2)$, then the iteration-complexity for CS-SPP to generate a $\bar\varepsilon$-saddle point is $\mathcal{O}(M^2D^2/\bar\varepsilon^2)$.

Proximal bundle method

Proximal point formulation ($\lambda_k = D/(4M\sqrt{k})$)

$$(x_k, y_k) = \underset{x\in\mathbb{R}^n}{\operatorname{argmin}}\ \underset{y\in\mathbb{R}^m}{\operatorname{argmax}} \left\{ \phi(x,y) + \frac{1}{2\lambda_k}\|x - x_{k-1}\|^2 - \frac{1}{2\lambda_k}\|y - y_{k-1}\|^2 \right\}$$

Apply the cutting plane method (PDCP) to

$$\min_{x\in\mathbb{R}^n} \left\{ f(x, y_{k-1}) + h_1(x) + \frac{1}{2\lambda_k}\|u - x_{k-1}\|^2 \right\},$$
$$\min_{y\in\mathbb{R}^m} \left\{ -f(x_{k-1}, y) + h_2(y) + \frac{1}{2\lambda_k}\|v - y_{k-1}\|^2 \right\}.$$

Iteration-complexity $\mathcal{O}((MD/\bar{\varepsilon})^{2.5})$ to find an $\bar{\varepsilon}$-saddle-point, for GBM.

Improved complexity $\mathcal{O}((MD/\bar{\varepsilon})^2)$, for two-cuts and multi-cuts schemes.

Outline

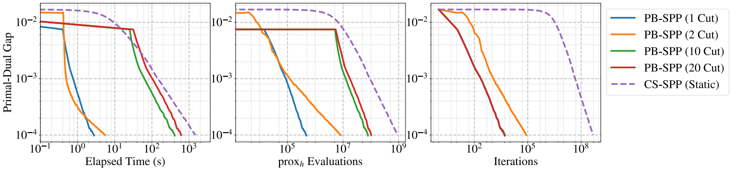
Regularized matrix game

Consider a matrix game

$$\min_{x\in\Delta_n}\max_{y\in\Delta_m}\{y^\top Ax+\gamma_x\|x\|_\infty-\gamma_y\|y\|_\infty\}$$

where $A\in\mathbb{R}^{m\times n}$ is the payoff matrix, x and y are mixed strategies on unit simplices. The ℓ_∞ regularization terms with parameters $\gamma_x\geq 0$ and $\gamma_y\geq 0$ discourage overly concentrated strategies by penalizing large coordinates.

We compare five methods: CS-SPP with a static stepsize of $\bar{\varepsilon}/(32M^2)$, and PB-SPP with a dynamic stepsize of $1/(2M\sqrt{k})$ under one-cut, two-cuts, 10-cuts, and 20-cuts schemes.



References

- F. Bach. Duality between subgradient and conditional gradient methods. SIOPT, 2015.
- M. Díaz and B. Grimmer. Optimal convergence rates for the proximal bundle method. SIOPT, 2023.
- Y. Du and A. Ruszczyński. Rate of convergence of the bundle method. JOTA, 2017.
- D. Fersztand and X. A. Sun. The proximal bundle algorithm under a Frank-Wolfe perspective: an improved complexity analysis. Preprint, 2024.
- P. Grigas, Methods for convex optimization and statistical learning. Phd thesis, MIT, 2016.
- K. C. Kiwiel. Efficiency of proximal bundle methods. JOTA, 2000.
- J. Liang and R. D. C. Monteiro. A proximal bundle variant with optimal iteration-complexity for a large range of prox stepsizes. SIOPT, 2021.
- J. Liang and R. D. C. Monteiro. A unified analysis of a class of proximal bundle methods for solving hybrid convex composite optimization problems. MOR, 2024.
- H. Lu and R. M. Freund. Generalized stochastic Frank-Wolfe algorithm with stochastic “substitute" gradient for structured convex optimization. MP, 2021.

Thank you!