Convex Feasibility Problem with Constraints Given by a Nonlinear Separation Oracle

Faranak Sharifi Mokhtarian

Department of Mathematics and Statistics
Concordia University, Montreal

email: faranak@alcor.concordia.ca

Novel Approaches to Hard Discrete Optimization

April 26-28, 2001

Problem Statement

(FP) Find y in a convex set Γ

where

- Γ contains a ball of radius $0 < \epsilon < 1$
- $\Gamma \subset \Omega^0$ where $\Omega^0 = \{y \in \mathbb{R}^m \ : \ \|y\| \leq 1\}$
- Γ is implicitly defined by a separation oracle
- For every $\hat{y} \in \Omega^0$, the oracle checks for feasibility
- If $\hat{y} \in \Gamma$, the oracle answers "yes"
- Otherwise, the oracle answers "no" and generates

Problem Statement (cont'd)

- either a separating hyperplane determined by vector a (with $\|a\| = 1$) so that

$$\Gamma \subset \left\{ y \in \mathbb{R}^m : a^T y \leq a^T \hat{y} \right\}$$

- or, a strongly convex quadratic cut with

$$\Gamma \subset \left\{ y \in \mathbb{R}^m : \langle a, y \rangle + \frac{1}{2}\langle y, Qy \rangle \leq \langle a, \hat{y} \rangle + \frac{1}{2}\langle \hat{y}, Q\hat{y} \rangle \right\}$$

- $Q \succeq \alpha I$, for some $\alpha > 0$
- It is assumed (without loss of generality) Q is symmetric and

$$\max\{\|a\|, \|Q\|\} \leq 1$$

Special Cases of the Problem

(1) The problem of finding a feasible point in

$$\Gamma = \{y \in \mathbb{R}^m : f_t(y) \leq 0 \;\; t \in \tau\}$$

(τ finite or infinite) where f_t is convex, can be cast as FP

- If, e.g., f_t is strongly convex with modulus $\alpha_t > \alpha$ then

$$f_t(y) \geq f_t(\hat{y}) + \langle s, y - \hat{y} \rangle + \frac{\alpha_t}{2}\|y - \hat{y}\|^2$$

for all $s \in \partial f_t(\hat{y})$

- In this case the oracle selects (for some $s \in \partial f_t(\hat{y})$)

$$a = \frac{s + \alpha_t \hat{y}}{l} \qquad Q = \frac{\alpha_t I}{l}$$

where $l = \max\{\|a\|, \alpha_t\}$

- The quadratic cut is also valid when f_t is not strongly convex, yet satisfies the above inequality at $\hat{y}$ for some $s \in \partial f_t(\hat{y})$.

- If f_t does not satisfy the above inequality, then a linear cut determined by $a = \frac{s}{\|s\|}$ is generated

Special Cases of the Problem

(2) The following problem can also be cast as FP:

- Find a point in Γ where

$$\begin{aligned}\Gamma \;=\; \{&\langle a_i, y\rangle \le c_i \quad i \in \tau_1,\\ &\langle a_j, y\rangle + \frac{1}{2}\langle y, Q_j y\rangle \le c_j \quad j \in \tau_2\}\end{aligned}$$

- where τ_1 and τ_2 finite or infinite
- $Q_j \succeq \alpha I$

Special Case of Variational Inequalities

(3) The following problem can also be cast as FP:

$$VI(G,T): \quad \text{Find } y^* \in G : \langle T(y^*), y - y^* \rangle \geq 0 \; \forall y \in G$$

- $T : \Omega^0 = \{y : \|y\| \leq 1\} \to R^m$ is strongly monotone, i.e.

$$\langle T(y) - T(x), y - x \rangle \geq \alpha \|y - x\|^2 , \quad \forall x, y \in \Omega^0$$

- $\max_{y \in \Omega^0} \|T(y)\| \leq M < \infty$
- $G \subset \Omega^0$ is a closed convex set in R^m
- G contains a ball of radius ϵ
- G is defined by a **separation oracle** which (can assume for simplicity) returns only linear cuts
- The cuts are either the strongly convex quadratic cuts

$$\langle T(y_j), y - y_j \rangle + \alpha \|y - y_j\|^2 \leq 0$$

  or linear cuts coming from the feasible region

Literature Review

- Goffin, Haurie and Vial (1993) (ACCPM)
- Goffin, Luo and Ye (1996) (convergence of ACCPM)
- Luo and Sun (1998) (quadratic cuts, finite case and shallow cuts)
- Goffin and vial (2000) (multiple linear cuts)
- Lüthi and Büeler (2000) (ACQCM for strongly monotone VI over an explicitly defined convex set)
- Mokhtarian and Goffin (2000) (infinitely many strongly quadratic/ self-concordant central cuts)
- Oskoorouchi and Goffin (2000) (semidefinite version of AC-CPM)

Analytic Center

- Let Ω be a bounded set in $\mathbb{R}^m$ defined by n linear or quadratic inequalities as follows:

$$\Omega = \{y \in \mathbb{R}^m \ : \ f_j(y) \leq 0, \ \ j = 1, \ldots, n\}$$

- For $j = 1, \ldots, n$, the potential function of Ω at y is

$$\phi(\Omega, y) = -\sum_{j=1}^{n} \ln(-f_j(y))$$

- The min potential of Ω is

$$P(\Omega) = \min_{y \in \Omega_{int}} \phi(\Omega, y)$$

- The analytic center of Ω is the unique point at which the above minimum is attained

Approximate Analytic Center

- The gradient and Hessian of $\phi(\Omega, y)$ are given by

$$g(\Omega, y) = \phi'(\Omega, y) = \sum_{j=1}^{n} \frac{f_j'(y)}{-f_j(y)}$$

and

$$H(\Omega, y) = \phi''(\Omega, y) = \sum_{j=1}^{n} \left(\frac{f_j'(y)(f_j'(y))^T}{(f_j(y))^2} + \frac{f_j''(y)}{-f_j(y)} \right)$$

- The point y is called a θ-approximate analytic center, or θ-center of Ω for short, if

$$\|g(\Omega, y)\|_{H(\Omega,y)}^* \leq \theta.$$

- Notation:

$$\|u\|_{H(\Omega,y)}^* = \langle u, H^{-1}(\Omega, y)u \rangle^{1/2}$$

Algorithm

0. (Initialization) Let $\Omega^0 = \{y \in \mathbb{R}^m : \|y\|^2 \leq 1\}$.
Set $y^0 = 0$.

1. (k-th iteration) Calculate y^k, a θ-center of Ω^k (the current outer approximation of Γ).

2. Call the oracle to check if $y^k \in \Gamma$. If so stop; else the oracle returns either a linear or strongly convex quadratic cut. Add this cut to Ω^k through y^k.

3. Set $k = k + 1$ and return to Step 1.

Addition of a Quadratic Cut

- A special updating direction is used to move from the current approximate center of Ω, y^a, to an interior point of the updated Ω (as the one given by Mokhtarian and Goffin)
- The motivation is to move towards an approximate analytic center of the set of intersection of the new quadratic cut and the Dikin ellipsoid of Ω centered at y^a
- The norm of this updating direction measured in a metric, that is defined by a combination of $H(\Omega, y^a)$ and Q_{n+1} plays an important role in our analysis
- To make the algorithm more practical, an approximation to the norm of this updating direction is used ($\bar{r}$)
- Moving towards an approximate center of the updated Ω after the updating step can be done in $O(1)$

Updating Step (Motivation)

- $\min -\ln s_d - \ln s_q$
  s_d: Dikin's slack
  s_q: Quadratic cut slack
- First order optimality conditions:

$$y^u = y^a - \left(H + \frac{s_D}{s_Q}Q\right)^{-1} \frac{s_D}{s_Q}a^*$$

- Let $\bar{r}^{-1} = \frac{s_D}{s_Q}$

Addition of a Quadratic Cut (cont'd)

- Suppose Ω is changed to Ω^+ as follows:

$$\begin{aligned}\Omega^+ &= \{f_j(y) \le 0, \quad j = 1, \ldots, n \\ &\qquad h_{n+1}(y) \le h_{n+1}(y^a)\}\end{aligned}$$

  where y^a is a θ-center of Ω with $0 < \theta < 1$ and

$$h_{n+1}(y) = -c_{n+1} + \langle a_{n+1}, y\rangle + \frac{1}{2}\langle y, Q_{n+1}y\rangle$$

- (Updating step)

$$y^u = y^a - \frac{\beta}{\bar{r}}\bar{H}^{-1}a^*$$

  where $0 < \beta < 1$ and

$$\begin{aligned}a^* &= a_{n+1} + Q_{n+1}y^a \\ \bar{H} &= H(\Omega, y^a) + \frac{Q_{n+1}}{\bar{r}} \\ \chi(\bar{r}) &= \langle a^*, \bar{H}^{-1}a^*\rangle^{\frac{1}{2}}\end{aligned}$$

  and $0 < \bar{r} < 1$ satisfies

$$\frac{1}{\sqrt{2}}\bar{r} \le \chi(\bar{r}) \le \bar{r}$$

Potential Changes (Addition of a Quadratic Cut)

- $y^u \in \Omega^+_{int}$
- $P(\Omega^+) \leq \phi(\Omega^+, y^u) \leq \phi(\Omega, y^a) - \ln \bar{r} + \alpha_1(\beta, \theta)$

  $$\alpha_1(\beta, \theta) = -\ln(1-\beta) - \ln\left[\beta\left(1 - \frac{\beta}{2}\right)\right] - (1-\theta)\beta$$

- $P(\Omega^+) \geq \phi(\Omega, y^a) - \ln \bar{r} + \alpha_2(\theta)$

  $$\alpha_2(\theta) = \min_{v>0}\{(1-\theta)v - \ln(1+v) - \ln(1+\sqrt{1+v^2})\}$$

- Furthermore

  $$\phi(\Omega^+, y^u) - P(\Omega^+) \leq \alpha_1(\beta, \theta) - \alpha_2(\theta)$$

  this is used in finding an upper bound for the number of centering steps

Computation of $\bar{r}$

- $\chi(\bar{r})$ satisfies
$$(\chi(\bar{r}))^2 = \sum_{i=1}^{m} \frac{\tilde{a}_i^2 \bar{r}}{\bar{r} + \lambda_i}$$
- $\tilde{a}_i$ is the i-th component of $\tilde{a} = L^{-1}a^*$
- L is a lower triangular matrix such that $H = LL^T$ and
- λ_i is the i-th eigenvalue of $L^{-1}QL^{-T}$ with $i = 1, 2, \ldots, m$
- We want $0 < \bar{r} < 1$, to satisfy
$$\frac{1}{\sqrt{2}}\bar{r} \leq \chi(\bar{r}) \leq \bar{r}$$
This implies that
$$-\frac{\bar{r}}{4} \leq \left[\sum_{i=1}^{m} \frac{\tilde{a}_i^2}{\bar{r} + \lambda_i} - \frac{3\bar{r}}{4}\right] \leq \frac{\bar{r}}{4}$$
- So, finding $\bar{r}$ is like finding a root of
$$\psi(\bar{r}) = \sum_{i=1}^{m} \frac{\tilde{a}_i^2}{\bar{r} + \lambda_i} - \frac{3\bar{r}}{4}$$
in $(0, 1)$ with accuracy $\frac{\bar{r}}{4}$

Computation of $\bar{r}$ (cont'd)

- The hybrid algorithm of Ye can be used here to find a suitable $\bar{r}$
- This hybrid algorithm combines Newton's method and a binary search and can compute a root of $\chi(\bar{r})$ inside $(0, 1)$ with error $\delta = \frac{\bar{r}}{4}$ in $O(\log_2(\log_2(\frac{1}{\delta})))$ time
- A lower bound on $\bar{r}$ is

$$\bar{r} \geq \frac{\epsilon^4(1-3\theta)^2}{4[(3k+1-3\theta)^2 + 4(1-3\theta)^2]}$$

- If a quadratic cut is added at the kth iteration of the algorithm then the number of arithmetic operations required to find a suitable $\bar{r}$ is at most

$$O\left(\log_2\left(\log_2\left(\frac{(3k+1-3\theta)^2 + 4(1-3\theta)^2}{\epsilon^4(1-3\theta)^2}\right)\right)\right)$$

or

$$O\left(\log_2\left(\log_2\left(\frac{k^2}{\epsilon^4}\right)\right)\right)$$

Potential Changes (Addition of a Linear Cut)

- Let y^a be a θ-center of Ω. Consider

$$\Omega^+ = \{y : f_i(y) \leq 0,\ i = 1, \ldots, n,\ (a^+)^T y \leq (a^+)^T y^a\}$$

- the following holds due to Nesterov:

$$y^+ = y^a - \frac{[H(\Omega, y^a)]^{-1} a^+}{3\|a^+\|^*_{H(\Omega,y^a)}} \in (\Omega^+)_{int}$$

$$\begin{aligned} \phi(\Omega, y^a) - \ln \|a^+\|^*_{H(\Omega,y^a)} + c(\theta) &\leq P(\Omega^+) \leq \phi(\Omega^+, y^+) \\ &\leq \phi(\Omega, y^a) - \ln \|a^+\|^*_{H(\Omega,y^a)} \\ &\quad + \bar{c}(\theta) \end{aligned}$$

- $c(\theta)$ and $\bar{c}(\theta)$ are constants (depending on θ only) and

$$\|a\|^*_{H(\Omega,y^a)} = \langle a^*, H(\Omega, y^a)^{-1} a^* \rangle^{1/2}$$

- It follows from the above that

$$P(\Omega^+) \geq P(\Omega) - \ln \|a^+\|^*_{H(\Omega,y^c)} + c(\theta)$$

- and

$$\phi(\Omega^+, y^+) - P(\Omega^+) \leq \bar{c}(\theta) - c(\theta)$$

Centering

- The total number of (damped) Newton steps (not including the arithmetic complexity of computation of $\bar{r}$) to update to a new θ-center after addition of a quadratic cut is at most $O(1)$, with $O(1)$ depending on θ and β only
- The total number of (damped) Newton steps to update to a new θ-center after addition of a linear cut, is at most $O(1)$ (depending on θ only)

Convergence of Algorithm

- The following holds for the algorithm:

$$(k+1)\ln\left(\frac{2}{\epsilon^2}\right) \;\geq\; P(\Omega^0) - \frac{1}{2}\left(\sum_{j\in J_q}\ln(\bar{r}_j^2) + \sum_{j\in J_l}\ln(r_j^2)\right) + (k+1)\zeta(\theta)$$

- $\zeta(\theta) = \min\{\alpha_2(\theta), c(\theta)\}$
- $P(\Omega^0) = 0$
- This implies

$$\epsilon^4 \leq \left[\frac{48(1+\ln(k_q+1))}{\alpha(k_q+1)}\right]\left[\frac{8(1+\ln(k_l+1))}{(k_l+1)}\right]e^{-\zeta(\theta)}$$

- k is the total number of iterations
- $k_q \leq k$ is the total number of quadratic cuts added
- $k_l \leq k$ is the total number of linear cuts added
- Note that for large k (either large k_l or large k_q) the above inequality is violated

Conclusions

- A convex feasibility problem defined by a separation oracle was considered
- The oracle can return linear or strongly convex quadratic cuts that will be added through the approximate analytic center
- In case a quadratic cut is added the arithmetic complexity of updating from one approximate center to another is at most $O(1) + O(\log_2 \log_2(\frac{k^2}{\epsilon^4}))$ with the second term corresponding to $\bar{r}$
- In case a linear cut is added the arithmetic complexity of updating from one approximate center to another is at most $O(1)$
- The total number of cuts, $k = k_q + k_l$, added satisfies

$$\epsilon \leq O\left(\left(\frac{\ln(k_q+1)}{(\alpha(k_q+1)}\right)^{1/4}\left(\frac{\ln(k_l+1)}{k_l+1}\right)^{1/4}\right)$$