Chebyshev Center Based Column Generation

Lee, Chungmok

Systems Optimization Lab, KAIST

March 11, 2009

Accelerating Column Generation Procedure

- Classical column generation often shows desperately slow convergence
  - Slow convergence(the *tailing-off effect*)
  - Poor columns in initial stage(the *head-in effect*)
  - Optimal value of the restricted master problem remains same during many iterations(the *plateau effect*)
  - Dual solution is jumping one extreme point to another(the *bang-bang effect*)
  - Intermediate Lagrangian dual bounds do not converge monotonically(the *yo-yo effect*)
- Recently, several accelerating scheme for column generation are proposed
- Basic idea : Stabilizing movement of *dual solution*

Primal-Dual Relation of Column Generation

General form of the column generation master problem:

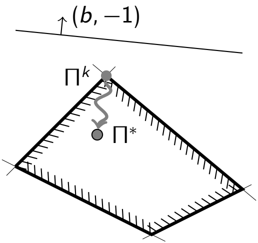
$$\begin{aligned}
\min \quad & \sum_{i \in K} c x^i \lambda_i \\
\text{subject to} \quad & \sum_{i \in K} A x^i \lambda_i \geq b, \\
& \sum_{i \in K} \lambda_i = 1, \\
& \lambda_i \geq 0, \qquad \forall i \in K.
\end{aligned}$$

And its dual:

$$\begin{aligned}
\max \quad & f_D(\pi, \pi_0) = b\pi - \pi_0 \\
\text{subject to} \quad & A x^i \pi - \pi_0 \leq c x^i, \qquad \forall i \in K, \\
& \pi \geq 0.
\end{aligned}$$

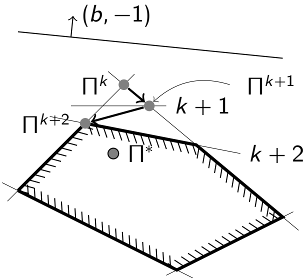
Column of primal problem $\Leftrightarrow$ Inequality of dual problem

Dual View of Column Generation



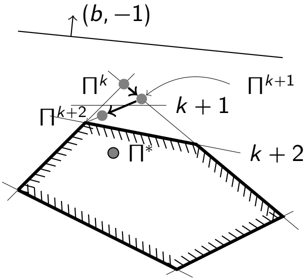
- Column generation in primal space ⇔ Cutting plane method in dual space
- In general convex optimization terms :
  - Solving column generation sub problem ⇔ obtaining gradient direction
  - Solving restricted master problem ⇔ determining next point for dual multipliers

Classical Column Generation Method



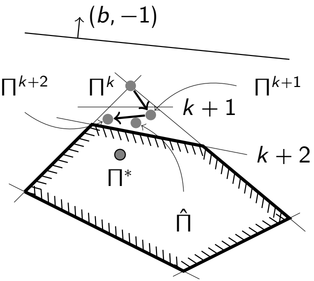
- Kelley's method ⇔ classical column generation method based on simplex algorithm
- Dual solution is likely to move among extreme points(see figure)
- Criticized for poor performance

Wentges's Weighted Dantzig-Wolfe Decomposition



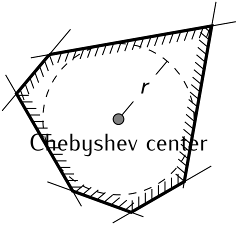
- Idea : finding better dual solutions in the neighborhood of the best solution
- Next Lagrangian multiplier(dual solution) is obtained by the convex combination of the current and the best solution so far
- Can be interpreted as a directional search starting always from the best dual solution to the current dual solution

Stabilized Column Generation



- Idea : mimicking *bundle method* of general convex optimization
- Recently, shown in many studies
- Seems promising approach
- Need to carefully manage *stabilization center* and *penalty function*

Chebyshev Center of Polyhedron



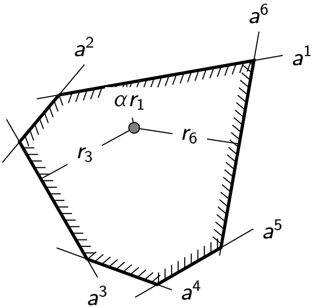
- Deepest point inside a convex set
- Can be obtained easily via *linear program*

$$\begin{aligned} \max \quad & r \\ \text{subject to} \quad & a_i x + r \|a_i\|_2 \leq b_i, \qquad \forall i \in \{1, \ldots, m\}, \\ & r \geq 0, \end{aligned}$$

- Chebyshev center cutting plane method : Elzinga and Moore 1975

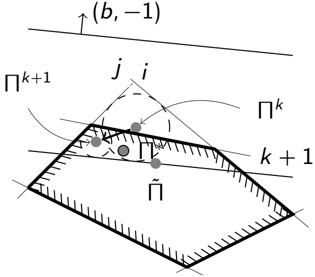
Proximity Adjusted Chebyshev Center

Motivation : Some inequalities may be more *important* than others



Proximity adjusted Chebyshev center of a polyhedron over Euclidian norm. The polyhedron is defined by six inequalities: $a^1x + \alpha \|a^1\| r \leq b^1, a^ix + \left\|a^i\right\| r \leq b^i, i = 2, \ldots, 6$. Note that, at the proximity adjusted Chebyshev center, $\alpha r_1 = r_3 = r_6$ is hold. The center point is getting nearer a^1 when $\alpha < 1$.

New Column Generation Procedure using Chebyshev Center



- Solve restricted master problem to obtain next *Chebyshev center*(contrary to extreme point)
- We can stick to linear programming model(contrary to nonlinear model in ACCPM)
- *No* need to maintain magical parameters(stabilization center, penalty functions etc)
- Quite general

Revised Column Generation Master(Dual)

Chebyshev dual master problem:

$$
\begin{aligned}
\max \quad & r \\
\text{subject to } \quad & Ax^i\pi - \pi_0 + \left\|(x^i, -1)\right\|_* r \le cx^i, && \forall i \in K, \\
& -b\pi + \pi_0 + \alpha \left\|(b, -1)\right\|_* r \le -b\tilde{\pi} + \tilde{\pi}_0, \\
& -\pi_j + r \le 0, && \forall j = 1, \ldots, m, \\
& \pi \ge 0, \\
& r \ge 0,
\end{aligned}
$$

- The above problem is to find the Chebyshev center point defined by restricted set of inequalities K, and
- a dual feasible solution $\tilde{\Pi}$.
- Note that the *proximity parameter* α

Revised Column Generation Master(Primal)

Chebyshev primal master problem:

$$
\begin{aligned}
\min \quad & \sum_{i \in K} c x^i \lambda_i - (b\tilde{\pi} - \tilde{\pi}_0) y \\
\text{subject to} \quad & \sum_{i \in K} A x^i \lambda_i - by - z \geq 0, \\
& \sum_{i \in K} \lambda_i - y = 0, \\
& \sum_{i \in K} \left\| (x^i, -1) \right\|_2 \lambda_i + \alpha \left\| (b, -1) \right\|_2 y + \sum_{j=1,\ldots,m} z_j \geq 1, \\
& \lambda_i \geq 0, && \forall i \in K, \\
& y \geq 0, z_j \geq 0, && \forall j = 1, \ldots, m.
\end{aligned}
$$

- Column generation is terminated when $r = 0$
- Possibilities of enhancement
  - Dynamically adjusting the proximity parameter α
  - Hybrid approach : Only in root node(classical approach in deeper node)

Algorithm

```
procedure ChebyshevCenterColGen
    k ← 0
    Π̃ ← (0, 0)                                          ▷ Or any dual feasible solution
    repeat
        Solve the restricted Chebyshev master problem
        (π, π_0) ← the optimal dual solution
        if the optimal value is greater than 0 then                     ▷ r > 0
            Solve the column generation sub problem
            if new column is generated then                ▷ Reduced cost is negative
                x^k ← new column
                Add x^k to the problem
                k ← k + 1
            else
                Π̃ ← (π, π_0)                          ▷ Update the best dual solution
            end if
        end if
    until r > 0
end procedure
```

Convergency of Proximity Adjusted Chebyshev Center

Proposition

The optimal value of the Chebyshev dual master problem, equivalently primal master problem, converges to zero, unless the sequence $\{\alpha_k\}$ converges to zero.

Proof.

Considering the dual polyhedron, it is easily seen that the sequence of radius of Chebyshev hypersphere, $\{r_k\}$, is bounded and monotonously decreasing. Let S be the dual feasible set to unrelaxed dual master problem. For Chebyshev center Π^k at iteration k, if $\Pi^k \notin S$ then the column generation sub problem eventually find a cut which separates Π^k. Since there is finitely many columns, there exist n_0 such that $\Pi^l \in S$ for any $l > n_0$. Assume, for contradiction, $\lim r_k = \hat{r} > 0$. For any $l > n_0$, $(b, -1)^T \Pi^{l+1} \geq (b, -1)^T \Pi^l + \alpha_{l+1} \left\|(b, -1)^T\right\| r_{l+1}$ is hold, owe to (??). By the Cauchy-Schwarz inequality, we have $\left\|(b, -1)^T\right\| \left\|\Pi^{l+1} - \Pi^l\right\| \geq (b, -1)^T(\Pi^{l+1} - (b, -1)^T \Pi^l) \geq \alpha_{l+1} \left\|(b, -1)^T\right\| r_{l+1}$, which implies $0 < \hat{r} \leq r_{l+1} \leq \frac{\left\|\Pi^{l+1} - \Pi^l\right\|}{\alpha_{l+1}}$. If sequence $\{\alpha_k\}$ converges to $\hat{\alpha}$ such that $\hat{\alpha} > 0$, $\lim \left\|\Pi^{l+1} - \Pi^l\right\| \geq \hat{r}\hat{\alpha} > 0$, which derives contradiction, since S is bounded and closed. If $\{\alpha_k\}$ diverges, clearly, r_l should converge to zero, which contradicts to assumption. The sequence $\{r_k\}$, therefore, converges to zero. □

Binpacking

The binpacking problem is to minimize the number of bins of width L, which are needed to pack all items $i = 1, \dots, I$ of widths $w_1, \dots, w_I$.

$$
\text{(Chebyshev Primal Master)}: \quad \begin{aligned}
\min \; & \sum_{p \in P'} x_p - \sum_{i \in I} \tilde{\pi}_i z \\
\text{s.t.} \; & \sum_{p \in P'} a_{ip} x_p - y_i - z \geq 0, \forall i \in I, \\
& \sum_{p \in P'} \|a_p\| \, x_p + \sum_{i \in I} y_i + \alpha \left\| \vec{1} \right\| z \geq 1, \\
& x_p \geq 0, \forall p \in P', y_i \geq 0, \forall i \in I, z \geq 0,
\end{aligned}
$$

where $\vec{1}$ is a $|I|$ dimensional vector of all 1s and $a_p := (a_{ip})_{i \in I}$.
The column generation subproblem corresponds to the knapsack problem, which finds profitable packing pattern:

$$
\text{(Binpacking Oracle)}: \; \max \sum_{i \in I} a_i \pi_i, \; \text{s.t.} \; \sum_{i \in I} w_i a_i \leq L, a_i \in \{0, 1\}.
$$

Binpacking: Test results for different algorithm parameters

prob	α	L_1				L_2			
		#iter	time	master	oracle	#iter	time	master	oracle
u120	1	355.6	0.38	0.26	0.12	365.4	0.44	0.27	0.16
	10	294.0	0.33	0.24	0.08	355.8	0.42	0.28	0.13
	100	186.3	0.17	0.12	0.04	301.8	0.33	0.24	0.08
u250	1	710.3	2.11	1.46	0.65	727.9	2.42	1.53	0.88
	10	585.2	1.65	1.30	0.35	716.7	2.26	1.54	0.71
	100	393.4	0.90	0.68	0.22	620.1	1.78	1.39	0.39
u500	1	1344.2	15.20	9.81	5.38	1359.9	22.89	9.77	13.10
	10	1121.3	10.86	8.15	2.71	1360.0	23.27	10.08	13.18
	100	778.2	6.15	3.75	2.40	1236.3	15.08	8.48	6.59
t60	1	238.3	0.24	0.08	0.16	253.3	0.28	0.09	0.19
	10	133.2	0.07	0.05	0.02	238.1	0.26	0.08	0.17
	100	83.8	0.04	0.03	0.01	129.5	0.06	0.04	0.02
t120	1	442.3	3.52	0.34	3.18	443.2	3.28	0.38	2.90
	10	255.2	0.32	0.21	0.11	454.0	3.72	0.38	3.34
	100	161.5	0.15	0.11	0.04	264.1	0.31	0.20	0.10
t249	1	839.0	94.38	2.00	92.37	759.3	17.38	1.94	15.43
	10	514.4	1.76	1.25	0.51	726.2	16.31	1.83	14.48
	100	328.8	0.86	0.65	0.21	484.3	1.83	1.32	0.51
t501	1	1047.6	1691.32	7.14	1684.16	1460.6	22.00	13.98	8.00
	10	1012.0	10.40	8.17	2.22	1419.5	20.36	13.45	6.90
	100	664.6	4.72	3.46	1.25	994.2	11.81	9.11	2.69

Binpacking: Comparisons with Other Algorithms

prob	Chebyshev($L_1, \alpha = 100$)				Stabilized				Kelley's			
	#iter	time	master	oracle	#iter	time	master	oracle	#iter	time	master	oracle
u120	186.3	0.17	0.12	0.04	329.5	0.38	0.25	0.12	403.7	0.31	0.2	0.1
u250	393.4	0.9	0.68	0.22	643.4	2.38	1.62	0.75	829	1.81	1.24	0.57
u500	778.2	6.15	3.75	2.4	1162.3	16.17	9.93	6.22	1574.9	10.1	8.38	1.71
u1000												
t60	83.8	0.04	0.03	0.01	102.2	0.09	0.05	0.04	214.6	0.11	0.06	0.05
t120	161.5	0.15	0.11	0.04	227.6	0.67	0.32	0.35	404.1	0.55	0.26	0.28
t249	328.8	0.86	0.65	0.21	484.3	10.47	1.76	8.71	812.6	3.21	1.58	1.62
t501	664.6	4.72	3.46	1.25	994.2	368.56	8.68	359.87	1593.9	19.83	11.31	8.5

VRP with Time Windows

Vehicle routing problem with time windows(VRPTW) is to cover given set of customers, I, exactly once with respecting customer's time window, while minimizing total traveling distance. Let R denote a set of all routes, which satisfy customers' time windows and vehicle capacity, i.e., feasible routes. Any route of vehicle should start from and end at depot.

(Chebyshev Primal Master) :

$$\begin{aligned} \min \; & \sum_{r \in R'} c_r x_r - (\sum_{i \in I} \tilde{\pi}_i + m\tilde{\pi}_0) z \\ \text{s.t.} \; & \sum_{r \in R'} \delta_{ir} x_r - y_i - z \geq 0, \forall i \in I, \\ & \sum_{r \in R'} \|(\delta_r, 1)\| \, x_r + \sum_{i \in I} y_i + \alpha \left\| (\vec{1}, m) \right\| z \geq 1, \\ & x_r \geq 0, \forall r \in R', y_i \geq 0, \forall i \in I, z \geq 0, \end{aligned}$$

where $\vec{1}$ is a $|I|$ dimensional vector of all 1s and $a_p := (a_{ip})_{i \in I}$.
The column generation subproblem corresponds to the shortest path problem with resource constraints(SPPRC), which finds profitable route such that $\sum_{i \in I} \delta_{ir} \pi_i + \pi_0 > c_r$.

VRPTW: Test results for different algorithm parameters

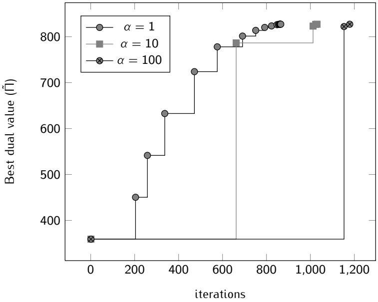
prob	α	L_1				L_2			
		#iter	time	master	oracle	#iter	time	master	oracle
C108	1	570	307.77	0.5	307.27	534	352.67	0.42	352.25
	10	647	497.85	0.76	497.09	577	479.43	0.53	478.9
	100	746	704.15	1.42	702.72	613	563.9	0.68	563.22
C109	1	555	1414.42	0.48	1413.93	505	1329.12	0.39	1328.72
	10	573	1395.75	0.64	1395.11	520	1293.66	0.47	1293.18
	100	634	2015.53	1.13	2014.39	574	1678.05	0.68	1677.37
R101	1	446	4.83	0.35	4.48	488	5.52	0.31	5.21
	10	430	4.9	0.36	4.53	418	4.81	0.3	4.5
	100	484	6.41	0.74	5.67	437	5.59	0.42	5.17
R110	1	531	202.58	0.61	201.97	560	231.47	0.54	230.92
	10	551	267.62	0.86	266.76	521	223.11	0.57	222.53
	100	581	355.34	1.67	353.66	534	264.95	0.82	264.13
RC102	1	612	70.12	0.58	69.54	624	88.96	0.53	88.43
	10	510	77.15	0.59	76.55	514	71.01	0.48	70.52
	100	567	90.37	1.3	89.07	528	81.12	0.7	80.42
RC107	1	542	463.94	0.52	463.41	549	470.74	0.46	470.27
	10	512	462.71	0.72	461.98	474	388.54	0.45	388.09
	100	526	534.05	1.11	532.93	507	482.26	0.62	481.64

VRPTW: Comparisons with Other Algorithms

prob	Chebyshev($L_1, \alpha = 1$)				Stabilized				Kelley's			
	#iter	time	master	oracle	#iter	time	master	oracle	#iter	time	master	oracle
C101	554	17.41	0.43	16.97	407	18.48	0.51	17.96	1008	43.08	0.98	42.08
C102	866	228.44	0.76	227.68	644	308.59	0.97	307.62	1262	493.15	1.46	491.67
C105	599	22.26	0.46	21.8	473	33.35	0.66	32.68	1033	79.11	1.12	77.98
C106	598	77.6	0.48	77.11	506	83.46	0.64	82.81	1041	241.07	1.13	239.92
C107	810	42.88	0.68	42.2	522	70.84	0.97	69.87	1153	164.78	1.4	163.36
C108	570	307.77	0.5	307.27	628	570.99	0.68	570.3	714	723.18	0.62	722.55
C109	555	1414.42	0.48	1413.93	616	1956.4	0.76	1955.64	635	2651.24	0.55	2650.68
R101	446	4.83	0.35	4.48	458	6.16	0.54	5.62	509	6.55	0.41	6.14
R102	626	31.34	0.56	30.77	528	31.47	0.64	30.82	574	34.66	0.48	34.17
R103	908	120.64	1.02	119.62	596	124.14	1	123.14	656	129.9	0.68	129.22
R105	547	20.32	0.63	19.68	537	31.01	1.07	29.93	587	36.02	0.77	35.24
R106	849	104.87	1.12	103.74	638	117.28	1.07	116.21	653	122.1	0.75	121.34
R107	762	268.87	1.06	267.8	652	415.09	1.37	413.72	692	409.2	0.86	408.33
R109	478	61.71	0.5	61.2	527	106.63	0.85	105.78	567	123.69	0.6	123.09
R110	531	202.58	0.61	201.97	605	389.07	1.06	388.01	609	386.45	0.66	385.78
R111	657	173.2	0.75	172.44	600	267.89	1.1	266.78	674	307.44	0.82	306.61
RC101	484	12.17	0.42	11.74	481	14.91	0.62	14.29	504	15.33	0.43	14.9
RC102	612	70.12	0.58	69.54	526	98	0.69	97.31	559	120.74	0.52	120.21
RC103	662	270.67	0.73	269.93	577	390.68	0.87	389.8	597	381.78	0.6	381.18
RC105	511	32.59	0.45	32.13	495	40.33	0.62	39.7	549	46.06	0.49	45.56
RC106	486	72.6	0.44	72.16	502	105.25	0.62	104.62	533	107.13	0.47	106.65
RC107	542	463.94	0.52	463.41	510	529.12	0.64	528.48	546	607.01	0.5	606.5

Dependency on α

Evolutions of best dual values for different proximity parameter α for problem C102 with L_1



It is not easy to determine best proximity parameter α in advance, since it clearly depends on specific properties of problem to be solved. We may, however, propose some general guidelines as follows: Use small α if initial dual feasible solution is poor. This is based on simple observation, which says that it is better to find good dual feasible solution at early stage of iterations by forcing the next Chebyshev point to be near to the current best dual solution. As illustrated in figure, best dual values are getting better rapidly when α is small, eventually smaller number of iterations is needed to finish the algorithm.

Oracle

Problem C102: Detailed statistics of oracles

Algorithm	#iter	avg oracle	max oracle	#iter:oracle $\geq$ 1sec.(%)
Chebyshev	866	0.26	1.39	1.04%
Stabilized	644	0.48	22.47	7.61%
Kelley's	1262	0.39	4.64	4.68%

It is interesting that sometimes number of iterations and computational time seem not to be correlated. Looking times spent for solving oracle problem during column generation algorithm, they fluctuate greatly at some peculiar iterations. For example, for VRPTW problem C102 and stabilized method, it has took 22.47 seconds to solve the oracle at iteration 415, while average of times spent in oracles is only 0.48 second, see table. Roughly speaking, the overall performance seems to be governed by oftenness of these expensive oracles.

Some Remarks

- Expensive oracles originate from many nonzero values in dual solutions. In Chebyshev method, the values of nonzeros, i.e., relative distance to the nonnegative hyperplane, can be controlled by adjusting the proximity parameter, which is not achievable in the stabilized method or Kelley's.
- The size of Chebyshev primal master problem is $(m+1) \times (n+m+1)$, where m, n are the number of rows and columns of the standard Dantzig-Wolfe decomposition formulation. And for the stabilized method, the size is $(3m) \times (n+2m)$. Typically, empirical performance of simplex method used to solve the master problem depends on the number of rows. When the master problem is hard to solve, the largeness in size of the master problem of stabilized method may become problematic.
- Contrast to the proximity parameter, the choice of norm did not make significant differences. For the sake of numerical stability, nevertheless, L_1 norm seems to be a fair choice. It is noteworthy that we have tested L_∞ norm also, but failed to obtain better results.
- An natural extension of this work may be directed towards combining of two acceleration methods, i.e., stabilized Chebyshev center method.

Reference

C. Lee and S. Park, *Chebyshev Center based Column Generation*, Working paper.