Inexact Bundle Methods For Convex Optimization with Applications to Flows on Traffic and Telecommunication Networks*

Krzysztof C. Kiwiel
Systems Research Institute, Warsaw
Polish Academy of Sciences
kiwiel@ibspan.waw.pl

*Mathematics for Key Technologies and Innovation, Warsaw, Feb. 21–22, 2008.

Talk Outline

- Basic Problem Model
  - General Motivations
  - Personal Motivations
  - Main Algorithmic Contributions
- The Alternating Linearization Bundle Method
- Convergence
- Modifications
- Application to Multicommodity Network Flows
- Numerical Experience

Problem statement

The structured convex minimization problem

$$\theta_* := \inf\{\, \theta(\cdot) := \sigma(\cdot) + \pi(\cdot) \,\}, \tag{1}$$

where

- $\sigma : \mathbb{R}^m \to (-\infty, \infty]$ is closed proper convex
- $\pi : C \to \mathbb{R}$ is closed proper convex
- $C := \operatorname{dom} \sigma := \{u : \sigma(u) < \infty\}$ is the effective domain of σ.

Main assumptions:

- σ is "simple": minimizing σ plus a separable convex quadratic function is "easy"
- π is only known via an oracle, which at any $u \in C$ delivers an affine minorant of π

Motivation: Lagrangian relaxation of the problems

$$f_* := \inf\{ f(Ax) : x \in X \} \quad (2)$$
$$= \inf\{ f(y) : y = Ax, x \in X \}, \quad (3)$$

where $X \subset \mathbb{R}^n$ and A is an $m \times n$ matrix.

Minimizing over $(x, y) \in X \times \mathbb{R}^m$ the Lagrangian

$$L(x, y; u) := f(y) + \langle u, Ax - y \rangle$$

yields a dual problem of the form (1) with

$$\sigma(u) := f^*(u) := \sup_y \{ \langle u, y \rangle - f(y) \}, \quad (4)$$
$$\pi(u) := \sup\{ \langle -A^T u, x \rangle : x \in X \}. \quad (5)$$

The oracle may deliver an affine minorant $\langle -Ax, \cdot \rangle$ of π for a possibly inexact maximizer x in (5).

Subjective motivation for this research

- Recent terrific results of Babonneau & Vial on solving nonlinear multicommodity network flow problems with their analytic center cutting plane method (ACCPM)
- ACCPM is based on modern interior-point Newton (with optimized BLAS)
- Standard bundle with active-set QP much slower than recent ACCPM

Challenge: Can bundle be improved to compete with ACCPM?

Basic algorithmic ideas

Our method approximates the proximal point algorithm

$$\hat{u}^{k+1} = \arg\min \sigma(\cdot) + \pi(\cdot) + \tfrac{1}{2t_k}|\cdot - \hat{u}^k|^2 \quad \text{for } k = 1, 2, \ldots, \tag{6}$$

where $|\cdot|$ is the Euclidean norm and $t_k > 0$ are stepsizes.

Bundling: replace π in (6) by its polyhedral model $\check{\pi}_k \leq \pi$ derived from the past oracle answers.

Alternating linearization: a subproblem involving the sum of two functions σ and $\check{\pi}_k$ is replaced by two subproblems in which the functions are alternately represented by linear models:

$$\check{u}^{k+1} := \arg\min \bar{\sigma}_{k-1}(\cdot) + \check{\pi}_k(\cdot) + \tfrac{1}{2t_k}|\cdot - \hat{u}^k|^2, \tag{7}$$

$$u^{k+1} := \arg\min \sigma(\cdot) + \bar{\pi}_k(\cdot) + \tfrac{1}{2t_k}|\cdot - \hat{u}^k|^2. \tag{8}$$

Uses a linearization $\bar{\pi}_k \leq \check{\pi}_k$ which may a posteriori replace $\check{\pi}_k$ in (7) without changing its optimal value and solution.

The alternating linearization bundle method

Generates trial points $\{u^k\}_{k=1}^{\infty} \subset C$ at which the oracle is called.

For a fixed accuracy tolerance $\epsilon_\pi \geq 0$, at each $u^k \in C$ the oracle delivers an approximate value π_u^k and an approximate subgradient g_π^k of π that produce the approximate linearization

$$\pi_k(\cdot) := \pi_u^k + \langle g_\pi^k, \cdot - u^k \rangle \leq \pi(\cdot) \quad \text{with} \quad \pi_k(u^k) = \pi_u^k \geq \pi(u^k) - \epsilon_\pi. \quad (9)$$

Thus $\pi_u^k \in [\pi(u^k) - \epsilon_\pi, \pi(u^k)]$, whereas g_π^k lies in the ϵ_π-subdifferential of π at u^k

$$\partial_{\epsilon_\pi} \pi(u^k) := \{\, g_\pi : \pi(\cdot) \geq \pi(u^k) - \epsilon_\pi + \langle g_\pi, \cdot - u^k \rangle \,\}.$$

$\theta_u^k := \sigma_u^k + \pi_u^k$ is the approximate value of θ at u^k with $\sigma_u^k := \sigma(u^k)$.

At iteration $k \geq 1$, the current stability center $\hat{u}^k := u^{k(l)} \in C$ for some $k(l) \leq k$ has the value

$$\theta^k_{\hat{u}} := \theta^{k(l)}_u \in [\theta(\hat{u}^k) - \epsilon_\pi, \theta(\hat{u}^k)]. \tag{10}$$

If $\pi^k_{\hat{u}} < \bar{\pi}_k(\hat{u}^k)$ due to oracle errors, the predicted descent

$$v_k := \theta^k_{\hat{u}} - [\sigma(u^{k+1}) + \bar{\pi}_k(u^{k+1})] \tag{11}$$

may be nonpositive; hence, if necessary, t_k is increased and (7)–(8) are solved again until $v_k \geq |u^{k+1} - \hat{u}^k|^2/2t_k$.

A descent step to $\hat{u}^{k+1} := u^{k+1}$ is taken if for a fixed $\kappa \in (0, 1)$

$$\theta^{k+1}_u \leq \theta^k_{\hat{u}} - \kappa v_k. \tag{12}$$

Else a null step $\hat{u}^{k+1} := \hat{u}^k$ occurs: $\bar{\pi}_k$ and the new linearization π_{k+1} are used to produce a better model $\check{\pi}_{k+1} \geq \max\{\bar{\pi}_k, \pi_{k+1}\}$.

Convergence

As usual in bundle methods, assume the oracle's subgradients are locally bounded:

$$\{g_\pi^k\} \text{ is bounded if } \{u^k\} \text{ is bounded.} \tag{13}$$

Consider the asymptotic objective value $\theta_{\hat{u}}^\infty := \lim_k \theta_{\hat{u}}^k$.

Theorem 1

1. *We have* $\theta_{\hat{u}}^k \downarrow \theta_{\hat{u}}^\infty \leq \theta_*$, *and* $\underline{\lim}_k V_k = 0$ *if* $\theta_* > -\infty$.
2. $\theta_* \leq \underline{\lim}_k \theta(\hat{u}^k) \leq \overline{\lim}_k \theta(\hat{u}^k) \leq \theta_{\hat{u}}^\infty + \epsilon_\pi$.

The (easily computable) optimality measure $V_k \geq 0$ satisfies

$$\theta_{\hat{u}}^k \leq \theta(u) + V_k(1 + |u|) \quad \text{for all } u. \tag{14}$$

Modifications

Looping between subproblems: If for a fixed $\check{\kappa} \in (0, 1)$,

$$\sigma(u^{k+1}) + \check{\pi}_k(u^{k+1}) > \theta^k_{\hat{u}} - \check{\kappa} v_k, \tag{15}$$

set $\bar{\sigma}_{k-1}(\cdot) := \bar{\sigma}_k(\cdot)$, and solve the two subproblems again.

Solving the σ-subproblem approximately: Solve the Fenchel dual (usually nice!)

Application to multicommodity network flows

Let $(\mathcal{N}, \mathcal{A})$ be a directed graph with $N := |\mathcal{N}|$ nodes and $m := |\mathcal{A}|$ arcs. Let $E \in \mathbb{R}^{N \times m}$ be its node-arc incidence matrix. There are n commodities to be routed through the network. For each commodity i there is a required flow $r_i > 0$ from its source node o_i to its sink node d_i. Let s_i be the supply N-vector of commodity i, having components $s_{io_i} = r_i$, $s_{id_i} = -r_i$, $s_{il} = 0$ if $l \neq o_i, d_i$. Our nonlinear multicommodity flow problem (NMFP for short) is:

$$\min \quad f(y) := \sum_{j=1}^{m} f_j(y_j) \tag{16a}$$

$$\text{s.t.} \quad y = \sum_{i=1}^{n} x_i, \tag{16b}$$

$$x_i \in X_i := \{\, x_i : Ex_i = s_i, 0 \leq x_i \leq \bar{x}_i \,\}, \quad i = 1{:}n, \tag{16c}$$

where each arc cost function f_j is closed proper convex, y is the total flow vector, x_i is the flow vector of commodity i, and $\bar{x}_i$ is a fixed positive vector of flow bounds for each i.

We may treat problem (16) as (3) with

$$Ax = \sum_{i=1}^{n} x_i, \quad X = \prod_{i=1}^{n} X_i.$$

At each u^k, the oracle solves shortest path problems to evaluate

$$\pi(u^k) = -\sum_{i=1}^{n} \min\{\langle u^k, x_i \rangle : x_i \in X_i\}.$$

Specific arc costs

Kleinrock's average delays with arc capacities $c_j > 0$:

$$f_j(y_j) := \begin{cases} \infty & \text{if } y_j \geq c_j, \\ y_j/(c_j - y_j) & \text{if } y_j \in [0, c_j), \\ y_j/c_j & \text{if } y_j < 0, \end{cases} \tag{17a}$$

$$f_j^*(u_j) := \begin{cases} \left(\sqrt{c_j u_j} - 1\right)^2 & \text{if } u_j \geq 1/c_j, \\ \infty & \text{if } u_j < 1/c_j, \end{cases} \tag{17b}$$

The BPR delays with $\alpha_j \geq 0$, $\beta_j > 0$, $\gamma_j \geq 2$:

$$f_j(y_j) := \begin{cases} \alpha_j y_j + \beta_j y_j^{\gamma_j} & \text{if } y_j \geq 0, \\ \alpha_j y_j & \text{if } y_j < 0, \end{cases} \tag{18a}$$

$$f_j^*(u_j) := \begin{cases} \frac{\gamma_j - 1}{\gamma_j}(u_j - \alpha_j)^{\gamma_j/(\gamma_j-1)}/(\beta_j \gamma_j)^{1/(\gamma_j-1)} & \text{if } u_j \geq \alpha_j, \\ \infty & \text{if } u_j < \alpha_j. \end{cases} \tag{18b}$$

Implementation issues

- Subproblem solution:
  - $\check{\pi}_k$-subproblem: QP with hot restarts
  - σ-subproblem: Newton with Armijo's backtracks
  - Looping: at most 30 loops at any iteration
- Shortest-path oracle: subroutine L2QUE of Gallo & Pallotino
- Termination criterion: stop when the relative optimality gap is small enough:

$$\gamma_{\text{rel}}^k := (f_{\text{up}}^k - f_{\text{low}}^k)/\max\{f_{\text{low}}^k, 1\} \leq \epsilon_{\text{opt}} = 10^{-5}, \qquad (19)$$

where f_{up}^k and f_{low}^k are the best upper and lower bounds on f_* obtained so far.

Test problems

We used the test problems of Babonneau & Vial.

Notation:

- N is the number of nodes
- m is the number of arcs
- n is the number of commodities
- S is the number of common sources
- $f_*^{\text{Kleinrock}}$ is the optimal value for the Kleinrock costs
- f_*^{BPR} is the optimal value for the BPR costs

Table 1: Test problems: Planar

Problem	N	m	n	S	$f_*^{\text{Kleinrock}}$	f_*^{BPR}
planar30	30	150	92	29	40.5668	4.44549×10^7
planar50	50	250	267	50	109.478	1.21236×10^8
planar80	80	440	543	80	232.321	1.81906×10^8
planar100	100	532	1085	100	226.299	2.29114×10^8
planar150	150	850	2239	150	715.309	5.27985×10^8
planar300	300	1680	3584	300	329.120	6.90748×10^8
planar500	500	2842	3525	500	196.394	4.83309×10^9
planar800	800	4388	12756	800	354.008	1.16952×10^9
planar1000	1000	5200	20026	1000	1250.92	3.41859×10^9
planar2500	2500	12990	81430	2500	3289.05	1.23827×10^{10}

Table 2: Test problems: Grid

Problem	N	m	n	S	$f_*^{\text{Kleinrock}}$	f_*^{BPR}
Grid problems						
grid1	25	80	50	23	66.4002	8.33599×10^5
grid2	25	80	100	25	194.512	1.72689×10^6
grid3	100	360	50	40	84.5618	1.53241×10^6
grid4	100	360	100	63	171.331	3.05543×10^6
grid5	225	840	100	83	236.699	5.07921×10^6
grid6	225	840	200	135	652.877	1.05075×10^7
grid7	400	1520	400	247	776.566	2.60669×10^7
grid8	625	2400	500	343	1542.15	4.21240×10^7
grid9	625	2400	1000	495	2199.83	8.36394×10^7
grid10	625	2400	2000	593	2212.89	1.66084×10^8
grid11	625	2400	4000	625	1502.75	3.32475×10^8
grid12	900	3480	6000	899	1478.93	5.81488×10^8
grid13	900	3480	12000	900	1760.53	1.16933×10^9
grid14	1225	4760	16000	1225	1414.39	1.81297×10^9
grid15	1225	4760	32000	1225	1544.15	3.61568×10^9

Table 3: Test problems: Telecommunication & Transportation

Problem	N	m	n	S	$f_*^{\text{Kleinrock}}$	f_*^{BPR}
Telecommunication problems						
ndo22	14	22	23	5	103.412	1.86767×10^3
ndo148	61	148	122	61	151.926	1.40233×10^5
904	106	904	11130	106	33.4931	1.29197×10^7
Transportation problems						
Sioux-Falls	24	76	528	24	600.679	4.23133×10^6
Winnipeg	1067	2836	4344	135	1527.41	8.25673×10^5
Barcelona	1020	2522	7922	97	845.872	1.22856×10^6
Chicago-sketch	933	2950	93135	386	614.726	1.67484×10^7
Chicago-region	12982	39018	2296227	1771	3290.49	2.58457×10^7
Philadelphia	13389	40003	1149795	1489	2557.42	2.24926×10^8

Numerical results

- k and l are the numbers of iterations and descent steps
- Sigma is the average number of subproblems solved per iteration
- Newton is the average number of Newton's iterations for the one-dimensional subproblems
- CPU is the total CPU time in seconds
- %Si is the percentage of CPU time on the σ-subproblems
- %Or is the percentage of CPU time on the oracle's shortest path subproblems
- AC/AL is the ratio of the CPU times of ACCPM and our AL

Our SPEC marks (and hence CPU times) are comparable with Babonneau & Vial's.

Table 4: Peformance of AL for Kleinrock costs

Problem	k	l	Sigma	Newton	CPU	%Si	%Or	AC/AL
planar30	125	62	4.7	1.9	0.1	60	0	11.0
planar50	214	73	3.2	2.2	0.2	31	10	11.0
planar80	308	80	3.0	2.2	0.6	28	28	10.8
planar100	312	75	3.9	2.4	0.8	24	28	7.5
planar150	979	95	1.7	2.1	12.2	3	17	10.8
planar300	303	84	6.4	2.7	4.7	27	46	4.7
planar500	253	77	8.3	2.6	9.7	23	55	2.5
planar800	341	82	7.7	2.7	28.1	16	69	2.7
planar1000	648	104	4.1	3.0	74.8	8	73	4.1
planar2500	1530	103	2.5	2.6	1092.1	2	86	2.2

Table 5: Peformance of AL for Kleinrock costs

Problem	k	l	Sigma	Newton	CPU	%Si	%Or	AC/AL
grid1	92	65	8.2	2.3	0.1	20	20	5.0
grid2	185	62	2.9	2.4	0.0	0	0	8.0
grid3	222	74	6.7	2.2	0.4	43	13	5.7
grid4	247	79	5.3	2.7	0.4	43	9	7.7
grid5	290	82	5.5	2.3	1.2	40	19	10.0
grid6	453	89	2.9	2.5	2.3	17	26	10.6
grid7	646	98	3.0	2.4	8.3	12	32	11.0
grid8	940	102	2.1	2.3	21.0	8	42	18.3
grid9	900	99	2.2	2.4	24.3	7	49	12.6
grid10	730	100	2.8	2.7	22.0	9	54	9.1
grid11	424	85	5.6	3.3	14.0	19	51	6.9
grid12	458	96	5.8	3.4	26.9	16	59	4.0
grid13	423	94	6.4	3.7	26.0	20	58	4.8
grid14	470	106	7.1	3.9	49.2	18	62	3.4
grid15	451	102	7.7	4.1	49.4	19	62	3.3

Table 6: Peformance of AL for Kleinrock costs

Problem	k	l	Sigma	Newton	CPU	%Si	%Or	AC/AL
ndo22	361	187	17.9	2.0	0.1	30	0	2.0
ndo148	94	53	2.3	2.0	0.0	0	0	8.0
904	240	58	7.5	3.1	1.5	53	22	5.1
Sioux-Falls	497	252	2.4	2.1	0.1	8	0	16.0
Winnipeg	1298	482	4.6	1.8	123.7	4	10	1.1
Barcelona	2611	434	1.7	1.6	127.6	2	17	0.6
Chicago-sketch	375	92	8.1	2.5	18.3	18	60	1.6
Chicago-region	303	73	7.7	2.1	901.0	4	88	9.6
Philadelphia	433	89	8.4	3.2	1431.3	5	85	9.1

Table 7: Peformance of AL for BPR costs

Problem	k	l	Sigma	Newton	CPU	%Si	%Or	AC/AL
planar30	75	69	1.3	1.1	0.0	66	33	12.0
planar50	105	64	1.4	1.3	0.0	66	33	29.0
planar80	150	59	1.1	1.3	0.2	8	73	33.5
planar100	108	44	1.4	1.3	0.2	20	54	21.5
planar150	194	52	1.1	1.5	0.9	12	67	24.7
planar300	97	31	1.3	1.2	1.4	8	86	9.0
planar500	50	23	1.7	1.0	3.3	4	92	2.6
planar800	108	33	1.9	1.2	25.4	2	94	1.3
planar1000	209	41	1.4	1.3	32.6	2	88	4.1
planar2500	264	52	1.3	1.6	411.8	0	97	4.0

Table 8: Peformance of AL for BPR costs

Problem	k	l	Sigma	Newton	CPU	%Si	%Or	AC/AL
grid1	48	29	3.6	2.2	0.0	25	0	4.0
grid2	61	27	1.7	2.2	0.0	0	0	8.0
grid3	43	23	2.5	1.3	0.0	100	0	7.0
grid4	59	26	1.8	2.2	0.1	50	50	15.0
grid5	86	28	2.1	1.7	0.3	44	35	8.0
grid6	150	33	2.0	2.0	0.6	47	35	11.3
grid7	108	31	2.1	2.3	0.9	35	56	10.2
grid8	143	36	1.6	2.3	2.4	22	58	12.8
grid9	183	37	1.7	2.4	4.0	21	60	11.6
grid10	200	34	2.3	2.5	5.4	22	57	8.4
grid11	120	32	4.2	3.2	4.1	40	48	7.3
grid12	122	31	5.8	3.4	8.8	40	47	3.9
grid13	140	30	5.5	3.6	10.1	37	48	4.4
grid14	111	28	8.0	4.0	16.1	42	45	3.3
grid15	115	26	8.0	4.3	17.1	44	45	3.5

Table 9: Peformance of AL for BPR costs

Problem	k	l	Sigma	Newton	CPU	%Si	%Or	AC/AL
ndo22	11	8	2.2	2.2	0.0	0	0	1.0
ndo148	14	11	2.4	2.1	0.0	0	100	2.0
904	116	32	1.2	2.8	0.5	27	62	12.4
Sioux-Falls	105	37	6.3	2.6	0.1	83	0	14.0
Winnipeg	127	31	8.4	1.8	4.5	53	37	2.4
Barcelona	92	24	14.3	3.0	5.5	72	18	1.4
Chicago-sketch	129	32	7.0	2.2	7.1	32	55	2.6
Chicago-region	300	51	3.6	2.6	891.0	5	89	9.2
Philadelphia	671	62	2.7	1.9	3239.7	2	94	2.6