Contents

1 Lagrangian relaxation

Consider the following newsvendor-like problem

$$
\begin{aligned}
&\min f(\delta, \epsilon) \\
\textbf{s.t.} \quad & \\
& y + \delta - \epsilon = b \\
& y \in \Omega_y \subseteq \mathbb{R}^n, \delta \in \mathbb{R}^n_+, \epsilon \in \mathbb{R}^n_+
\end{aligned}
$$

where f is a convex function of δ, ϵ. The right-hand-side on the binding constraints is in the positive orthant: $b \in \mathbb{R}_+$. This problem widely appears in applications of device maintenance, inventory management, and so on. In the basic settings, let y be the ordering quantity quantities in a multi-item newsvendor problem, one minimizes the total expected cost:

$$\min_{y \in \mathbb{R}_+} \mathbf{E}\left(h \cdot e^\mathsf{T} \max\{y - b, 0\} + p \cdot e^\mathsf{T} \max\{b - y, 0\}\right)$$

It is easy to verify its equivalence to the problem above.

Let $\lambda \in \mathbb{R}^n$ be the Lagrangian multiplier, the dual function is:

$$
\begin{aligned}
&\phi(\lambda) = \min_{\delta,\epsilon} f(\delta, \epsilon) + \lambda^\mathsf{T}\delta - \lambda^\mathsf{T}\epsilon + \min_{y} \lambda^\mathsf{T} y - \lambda^\mathsf{T} b \\
\textbf{s.t.} \quad & \\
& y \in \Omega_y \\
& \delta \in \mathbb{R}^n_+, \epsilon \in \mathbb{R}^n_+
\end{aligned}
$$

We assume the resulting two subproblems for δ, ϵ and y are easy.

1.1 Affine case

The case for repair problem

Let $f = p^\intercal \delta + h^\intercal \epsilon$, we have

$$\phi(\lambda) = \min_{\delta,\epsilon} (p+\lambda)^\intercal \delta + (h-\lambda)^\intercal \epsilon + \min_{y} \lambda^\intercal y - \lambda^\intercal b$$

Then ϕ is unbounded unless $\lambda \in \Lambda$ where $\Lambda = \{\lambda : \lambda \in [-p, h]\}$, in which case

$$\phi(\lambda) = \min_{y \in \Omega_y} \lambda^\intercal y - \lambda^\intercal b,\ \lambda \in \Lambda$$

and $\delta^\star, \epsilon^\star = 0$ are corresponding optimizers for any $\lambda \in \Lambda$

1.2 Conditions for strong duality

It's well known that strong duality does not hold in general. We review some of the cases here. The Lagrangian duality theory can be found in any standard text.

(a) if Ω_y is convex then the strong duality holds ..., i.e. $\phi^\star = f^\star$

... add justifications here (slater, ...)

A more interesting result is devoted to mixed integer problems. **(Review Here)**.

(b) if $\Omega_y = \{y \in \mathbb{R}^n : y \in \Omega, y \in \mathbb{Z}^n\}$. Then we have the following relation for dual function,

$$\phi^\star = \min_{\delta,\epsilon} f(\delta, \epsilon) \quad \textbf{s.t. } y + \delta - \epsilon = b,\ y \in \mathrm{conv}(\Omega_y)$$

We conclude the strong duality holds since $Y = \{(y, \delta, \epsilon) : y + \delta - \epsilon = b,\ y \in \mathsf{conv}(\Omega_y)\}$ is already *a perfect formulation* in the sense that $Y = \mathsf{conv}(Y)$

add a proposition to show this or add more conditions to justify

2 Subgradient method

To solve the reduced problem, we consider a variant class of subgradient methods:

$$\lambda_{k+1} = \mathcal{P}(\lambda_k + s_k d_k)$$

where $\mathcal{P}$ is the projection onto dual space Λ. d_k is the update direction for current iteration and s_k is the step size using target-based or so-called Polyak's rule Polyak (1967):

$$s_k = \gamma_k \frac{\phi^\star - \phi(\lambda_k)}{||d_k||^2}$$

Note the direction d_k computed by

$$d_k = \bar{y}_k - b$$

where $\bar{y}_k$ is the convex combination of previous iterations $\{y_i\}_{i=1,\dots k}$ and each y_i solves $\phi_i = \phi(\lambda_i)$:

$$\bar{y}_k = \sum_k^i \alpha_k^i y_i, \quad \sum_k^i \alpha_k^i = 1, \alpha_k^i \geq 0$$

Alternatively, one can express the convexity in a recursive manner:

$$\bar{y}_k = (1 - \alpha_k) \cdot \bar{y}_{k-1} + \alpha_k \cdot y_k$$

For we simplicity take $g_k = y_k - b$, then g_k is a subgradient of ϕ at λ_k:

$$g_k \in \partial \phi_k$$

The direction can be rewritten as the combination of the subgradient and previous directions:

$$d_k = (1 - \alpha_k) \cdot d_{k-1} + \alpha_k \cdot g_k$$

(**Primal recovery**)

It is obvious to see the solution tuple to dual problem $(y^\star, \epsilon^\star, \delta^\star) = (y_k, 0, 0)$ at each iteration is feasible if and only if we can find $y_k = b$, **which in general will not hold**. This motivates the following heuristic based on linear programming theory.

$$\epsilon_k = \max\{y_k - b, 0\}$$
$$\delta_k = \max\{b - y_k, 0\}$$

also produce

$$\bar{\epsilon}_k = \max\{\bar{y}_k - b, 0\}$$
$$\bar{\delta}_k = \max\{b - \bar{y}_k, 0\}$$

and record the corresponding primal objective value as $z_k = f(\delta_k, \epsilon_k)$. To simplify our presentation, let z be a function of y such that $z_k = z(y_k)$, then z is also convex in y since both function f and $\max\{\cdot, 0\}$ are convex. It's also worth to notice that $\bar{\epsilon}_k$ should not be calculated as running averages: $\bar{\epsilon}_k \neq \sum_k^i \alpha_k^i \epsilon_i$. For such an "averaged" solution, we let $\bar{z}_k = z(\bar{y}_k) = f(\bar{\delta}_k, \bar{\epsilon}_k)$.

(**Review**)

We first review several features for the subgradient method regarding parameters γ_k, α_k and search direction d_k.

From the dual viewpoint our method iterates on convex combination of previous direction and current subgradient with Polyak's stepsize rules. This method is similar to Polyak's **heavy ball** method, while the difference lies in the usage of *convex combination.* Bertsekas (2016) gives a detailed convergence analysis for method of this kind, especially on different choices of stepsize, including diminishing, contant, and so on. Brännlund (1995) showed that if using convex combinations on an update scheme then the optimal step size is identical to the Camerini-Fratta-Maffioli (CFM) modification Camerini et al. (1975).

From the primal perspective, our method can be seen as a *primal averaging method.* Nedić and Ozdaglar (2009) gives a line of analysis on convergence and quality of the primal approximation by averaging over all previous solutions with constant stepsize. We also refer to Kiwiel et al. (2007) for target based stepsizes. The volume algorithm proposed by Barahona and Anbil (2000) is close to the case in Brännlund (1995) in a dual viewpoint while adopting $\hat{\lambda}_k$ instead of λ_k from the best dual bound $\hat{\phi}_k = \max_{i=1,\dots,k} \phi(\lambda_i)$:

$$\lambda_{k+1} = \mathcal{P}(\hat{\lambda}_k + s_k d_k)$$

There is no existing proof of convergence for the volume algorithm, and our experiments show that the algorithm converges to non-optimal solutions occasionally.

(Remark / Difference for our method)

since the solution is strictly feasible by implementation of the recovery heuristic, i.e., there is no need to bound for feasibility gap as has been done in most of literature covering the **primal recovery**. Instead, we have analyze the quality of the heuristic, i.e.

$$|\bar{z}_k - \hat{\phi}_k|$$

or

$$|\bar{z}_k - z^\star|$$

Nedić and Ozdaglar (2009) uses a simple averaging scheme that can be rephrased into a recursive equation with $\alpha_k = 1/k$ such that:

$$\bar{y}_k = \frac{k-1}{k} \cdot \bar{y}_{k-1} + \frac{1}{k} \cdot y_k$$

Then it gives bounds for the averaged solution ... $\leq \bar{z}_k \leq$ ... that involve the primal violation, norm of the subgradient, and ...

We wish to derive a similar bound. Furthermore, it uses constant stepsize $s_k = s, s \geq 0$ and the search direction defined solely by the subgradient. So we want to verify if the the **case for target-based** rules.

from the dual viewpoint we are close to Brännlund (1995) since we are using the convex combinations so as to generate a fastest convergent speed. we can use the results here to verify our choice of parameters (γ, α, d)

2.1 Convergence

2.1.1 Analysis outline

- we've already showed zero duality gap $\phi^\star = f^\star = z^\star$
- we show λ_k converges to $\lambda^\star \in \Lambda^\star$ for our choices of γ_k, α_k
- we show primal solution $\bar{z}_k$ converges to $z^\star$

Lemma 1 ϵ-subgradient.

$$g_k^\intercal(\lambda_k - \lambda) \leq \phi_k - \phi(\lambda)$$

$$d_k^\intercal(\lambda_k - \lambda) \leq \phi_k - \phi(\lambda) + \epsilon_k$$

where

$$\epsilon_k = \sum_k^i \alpha_k^i \cdot [g_i^\mathsf{T}(\lambda_k - \lambda_i) + \phi_i - \phi_k]$$

Notice ϵ_k can be further simplified by the definition of ϕ:

$$\epsilon_k = \sum_k^i \alpha_k^i \cdot (g_i^\mathsf{T}\lambda_k - \phi_k)$$

Lemma 2 Dual convergence, Brännlund (1995). The subgradient method is convergent if ϵ_k satisfies:

$$\frac{1}{2}(2 - \gamma_k)(\phi_k - \phi^\star) + \epsilon_k \leq 0$$

The proof can be done by showing the monotonic decrease of $\|\lambda_k - \lambda^\star\|$ via the iterative equations.

$$\|\lambda_{k+1} - \lambda^\star\|^2 \leq ||\lambda_k - \lambda^\star||^2 + 2 \cdot \gamma_k \frac{(\phi^\star - \phi_k)}{\|d_k\|^2} d_k^\mathsf{T}(\lambda_k - \lambda^\star) + (\gamma_k)^2 \frac{(\phi^\star - \phi_k)^2}{\|d_k\|^2}$$

Notice:

$$\begin{aligned}
&2 \cdot d_k^\mathsf{T}(\lambda_k - \lambda^\star) + \gamma_k(\phi^\star - \phi_k) \\
\leq &2(\phi_k - \phi^\star + \epsilon_k) + \gamma_k(\phi^\star - \phi_k) \\
= &(2 - \gamma_k)(\phi_k - \phi^\star) + 2\epsilon_k \leq 0
\end{aligned}$$

and we have the convergence by plugging in Lemma 1.

The next proposition states several convergence-guaranteed choices on parameters for convexity α_k and stepsize γ_k. Part (a) devotes to the results originally appeared in Brännlund (1995). Besides, we also consider a slower scheme that is widely used and simple to implement.

Proposition 1

(a) The choice of stepsize and direction in the subgradient method defined by

$$\alpha_k = \gamma_k = \begin{cases} \|d_{k-1}\|^2/(\|d_{k-1}\|^2 - g_k^\mathsf{T} d_{k-1}), & \text{if } g_k^\mathsf{T} d_{k-1} < 0 \\ 1, & \text{otherwise} \end{cases}$$

generates the fastest convergence speed with respect to

$$\|\lambda_{k+1} - \lambda^\star\|^2 \leqslant \|\lambda_k - \lambda^\star\|^2 - F(\gamma_k, \alpha_k)(\phi_k - \phi^\star)^2$$

where

$$F(\gamma_k, \alpha_k) = \begin{cases} \frac{\|d_k\|^2}{\|d_k\|^2\|g_k\|^2 - (g_k^\intercal d_k)^2}, & \text{if } g_k^\intercal d_k < 0 \\ 1/\|g_k\|^2, & \text{otherwise} \end{cases}$$

(b) to show the following is also convergent.

$$\alpha_k = \frac{1}{k}, \gamma_k = \gamma \in [1, 2]$$

Proposition 2

(a) For fixed $y = y_k$, (ϵ_k, δ_k) is the optimal solution for the restricted primal problem.

$$f(\epsilon_k, \delta_k) \leq f(\epsilon, \delta), \quad \forall \delta \geq 0, \epsilon \geq 0, y = y_k$$

(b)

$$\bar{z}_k \leq \sum_k^i \alpha_k^i z^i$$

PF. By convexity.

Now we visit properties for primal solutions.

Proposition 3 Primal solution bounds $|\bar{z}_k - z^\star|$?

- $-\delta_k + \epsilon_k = g_k = y_k - d$ is bounded, suppose $\|g_k - g^\star\| \leq L_g$
- f, z is Lipschitz continuous with L_z
- $\phi^\star - \phi_k \leq g_k^\intercal(\lambda^\star - \lambda^k) \leq \|g_k\|\|\lambda^\star - \lambda^k\| \Rightarrow \phi^k - \phi^\star$ by boundedness of g^k
- $\epsilon_k \leq \frac{1}{2}(2 - \gamma_k)(\phi^\star - \phi_k) \to 0$
- $\epsilon_k = d_k^\intercal \lambda_k - \phi_k \to 0$ (converge from above)
- $d_k^\intercal \lambda_k = (\bar{y}_k - b)^\intercal \lambda_k \to \phi^\star$

(affine case)

we notice a strong duality pair with fixed d_k at each iteration k.

(P)

$$\begin{aligned} &\min_{\delta,\epsilon} p^\top \delta + h^\top \epsilon \\ \textbf{s.t.} \quad & d_k + \delta - \epsilon = 0 \\ & \delta \in \mathbb{R}^n_+, \epsilon \in \mathbb{R}^n_+ \end{aligned}$$

and

(D)

$$\max_{\lambda} d_k^\top \lambda$$

by ..., $(\bar{\epsilon}_k, \bar{\delta}_k)$ minimizes the primal problem. Since (P) is well-defined, $\exists\ \lambda_k^\star \in [-p, h]$ such that:

$$\begin{aligned} & d_k^\top \lambda_k^\star = \bar{z}^k = p^\top \bar{\delta}_k + h^\top \bar{\epsilon}_k \\ & d_k^\top \lambda_k^\star \geq d_k^\top \lambda_k \end{aligned}$$

Then the sequence $\{d_k^\top \lambda_k^\star\}_k$ is bounded from below and above $(z^\star)$. As $d_k^\top \lambda_k \to \phi^\star$ and by strong duality $\phi^\star = z^\star$ we conclude $\bar{z}^k \to z^\star$

2.2 Computational Results

We summarize all 60 test cases from the repair model

Table 1: Computational results from the repair model

	$\lvert I\rvert$	$\lvert T\rvert$	bench		normal					volume				
			$\hat{\phi}$	$\bar{z}$	$\hat{\phi}$	$\bar{z}$	z	ϕ_gap	$\bar{z}$_gap	$\hat{\phi}$	$\bar{z}$	z	ϕ_gap	$\bar{z}$_gap
0	10	10	36.00	36.00	36.00	36.36	66	0.00%	0.99%	35.87	39.09	63	-0.35%	8.60%
1	25	20	158.00	158.00	158.00	160.51	263	0.00%	1.59%	158.00	159.96	270	0.00%	1.24%
2	20	25	172.00	172.00	172.00	178.91	280	0.00%	4.02%	172.00	173.73	232	0.00%	1.01%
3	10	20	56.00	56.00	55.99	57.15	81	-0.01%	2.05%	55.95	60.95	108	-0.09%	8.84%
4	10	10	30.00	30.00	29.98	30.52	60	-0.07%	1.72%	30.00	31.66	73	0.00%	5.52%
5	15	10	36.00	36.00	35.97	37.40	82	-0.09%	3.90%	35.99	37.20	82	-0.02%	3.32%
6	15	20	70.00	70.00	69.93	74.27	143	-0.10%	6.10%	70.00	71.00	119	0.00%	1.43%
7	10	10	40.00	40.00	39.96	42.47	73	-0.10%	6.18%	40.00	40.40	75	0.00%	1.00%
8	20	20	116.00	116.00	115.87	118.70	234	-0.11%	2.33%	116.00	117.17	156	0.00%	1.00%
9	15	15	36.00	36.00	35.84	38.29	93	-0.45%	6.35%	36.00	36.70	101	0.00%	1.96%
10	15	10	30.95	32.00	30.78	33.72	60	-0.53%	5.38%	30.99	31.91	90	0.13%	-0.28%
11	15	10	48.00	48.00	47.17	51.31	79	-1.73%	6.90%	46.98	49.48	79	-2.12%	3.07%
12	10	20	30.97	32.00	30.08	33.33	57	-2.87%	4.16%	30.00	33.44	90	-3.11%	4.49%
13	15	15	46.00	46.00	44.20	49.56	125	-3.92%	7.73%	44.00	74.86	110	-4.35%	62.74%
14	10	15	72.00	72.00	72.00	72.84	93	0.00%	1.16%	72.00	72.83	96	0.00%	1.16%
15	15	20	118.00	118.00	118.00	119.67	154	0.00%	1.41%	118.00	119.68	187	0.00%	1.42%
16	15	25	136.00	136.00	136.00	138.02	178	0.00%	1.48%	136.00	137.65	172	0.00%	1.22%
17	10	10	28.00	28.00	28.00	28.47	58	0.00%	1.67%	28.00	35.44	58	0.00%	26.56%
18	10	15	46.00	46.00	46.00	46.81	97	0.00%	1.76%	46.00	46.86	88	0.00%	1.88%

Table 1: (continued)

19	15	15	72.00	72.00	72.00	73.30	102	0.00%	1.80%	72.00	73.30	102	0.00%	1.80%
20	10	20	58.00	58.00	58.00	59.04	133	0.00%	1.80%	58.00	59.10	94	0.00%	1.89%
21	20	20	108.00	108.00	108.00	110.17	177	0.00%	2.01%	108.00	110.17	183	0.00%	2.01%
22	25	25	198.00	198.00	198.00	202.10	360	0.00%	2.07%	198.00	201.57	321	0.00%	1.80%
23	15	25	92.00	92.00	92.00	93.93	206	0.00%	2.10%	92.00	93.98	158	0.00%	2.16%
24	15	10	32.00	32.00	32.00	32.77	83	0.00%	2.41%	32.00	50.78	89	0.00%	58.67%
25	15	15	48.00	48.00	48.00	49.17	132	0.00%	2.44%	48.00	49.35	138	0.00%	2.81%
26	20	20	82.00	82.00	82.00	84.07	196	0.00%	2.53%	82.00	84.07	157	0.00%	2.53%
27	15	25	70.00	70.00	70.00	71.83	211	0.00%	2.62%	70.00	71.74	193	0.00%	2.48%
28	25	25	72.00	72.00	72.00	75.07	204	0.00%	4.26%	72.00	75.11	354	0.00%	4.32%
29	10	15	46.79	48.00	47.05	49.11	64	0.56%	2.31%	47.45	48.94	70	1.41%	1.96%
30	25	25	137.14	138.00	138.00	146.71	288	0.63%	6.31%	138.00	230.87	336	0.63%	67.30%
31	25	20	198.61	200.00	200.00	203.30	281	0.70%	1.65%	200.00	203.23	296	0.70%	1.62%
32	15	25	70.36	72.00	70.89	76.53	155	0.75%	6.29%	68.00	105.19	203	-3.36%	46.10%
33	20	20	140.90	142.00	141.98	144.60	203	0.77%	1.83%	142.00	143.82	236	0.78%	1.28%
34	25	20	148.80	150.00	149.97	154.31	289	0.79%	2.88%	149.51	159.45	304	0.47%	6.30%
35	25	25	208.20	210.00	210.00	213.88	360	0.86%	1.85%	210.00	213.46	333	0.86%	1.65%
36	20	20	132.72	134.00	134.00	136.65	203	0.97%	1.98%	134.00	136.68	215	0.97%	2.00%
37	25	25	190.03	192.00	192.00	196.06	294	1.04%	2.11%	192.00	196.12	348	1.04%	2.15%
38	25	20	132.13	134.00	133.64	137.69	199	1.14%	2.76%	134.00	135.50	271	1.42%	1.12%
39	25	20	144.26	146.00	145.92	149.13	184	1.15%	2.14%	146.00	147.70	205	1.21%	1.16%
40	20	25	138.33	140.00	139.99	142.56	243	1.20%	1.83%	140.00	141.77	249	1.21%	1.27%
41	20	25	98.63	100.00	99.99	103.42	248	1.38%	3.42%	100.00	101.70	287	1.39%	1.70%
42	10	25	74.86	76.00	75.98	77.10	150	1.49%	1.45%	75.99	83.03	135	1.50%	9.25%
43	10	25	78.65	80.00	79.87	82.00	135	1.55%	2.50%	80.00	80.80	135	1.71%	1.00%
44	15	25	90.52	92.00	92.00	94.12	212	1.64%	2.30%	92.00	97.61	209	1.64%	6.09%
45	15	10	68.81	70.00	69.97	70.88	102	1.69%	1.26%	69.95	71.18	78	1.66%	1.68%
46	10	15	60.87	62.00	62.00	62.77	77	1.85%	1.24%	62.00	62.77	77	1.85%	1.24%
47	15	15	76.32	78.00	77.76	79.95	127	1.89%	2.49%	78.00	78.73	144	2.21%	0.94%
48	10	25	56.91	58.00	58.00	59.26	157	1.91%	2.17%	58.00	59.31	154	1.91%	2.26%
49	15	20	52.77	54.00	53.81	57.58	129	1.98%	6.63%	54.00	55.03	161	2.33%	1.91%
50	20	25	146.99	150.00	149.99	153.33	263	2.04%	2.22%	149.99	153.86	296	2.04%	2.57%
51	15	20	64.61	66.00	66.00	67.56	162	2.15%	2.36%	66.00	67.55	135	2.15%	2.35%
52	10	15	66.45	68.00	68.00	69.41	90	2.33%	2.08%	68.00	69.18	90	2.33%	1.74%
53	10	10	44.75	46.00	45.83	47.01	57	2.41%	2.20%	45.72	47.92	63	2.18%	4.17%
54	10	20	48.66	50.00	50.00	50.93	83	2.76%	1.85%	50.00	50.96	95	2.76%	1.93%
55	10	20	30.43	32.00	31.37	34.86	84	3.08%	8.94%	32.00	32.52	81	5.14%	1.64%
56	10	25	63.92	66.00	66.00	68.33	117	3.25%	3.52%	66.00	67.50	114	3.25%	2.27%
57	10	25	44.43	46.00	46.00	47.20	103	3.53%	2.60%	46.00	47.25	145	3.53%	2.71%
58	20	25	160.21	166.00	165.99	169.60	218	3.61%	2.17%	165.60	171.88	218	3.36%	3.54%
59	15	20	53.34	56.00	56.00	57.65	149	4.99%	2.95%	56.00	57.80	131	4.99%	3.21%

in 1, $\bar{z}$ is the objective value computed from the averaged primal solution, $\hat{\phi}$ is the best lower bound at termination.

We find that using λ_k instead of $\hat{\lambda}_k$ can be better! Below is a typical case of divergence of $\bar{z}_k$ and $\hat{\phi}_k$ computed from the repair model. `normal_x` means the values are computed from subgradient method by using λ_k. `volume_x` is from the volume algorithm with $\hat{\lambda}_k = \arg\max_k \hat{\phi}_k$

(So we wish to show the convergence): $|\bar{z}_k - \hat{\phi}_k|$

Reference

10 Barahona F, Anbil R (2000) The volume algorithm: Producing primal solutions with a subgradient method. *Mathematical Programming* 87(3):385–399.

Bertsekas DP (2016) *Nonlinear programming* 3rd ed. (Athena Scientific).

Brännlund U (1995) A generalized subgradient method with relaxation step. *Mathematical Programming* 71(2):207–219.

Camerini PM, Fratta L, Maffioli F (1975) On improving relaxation methods by modified gradient techniques. *Nondifferentiable optimization.* (Springer), 26–34.

Kiwiel KC, Larsson T, Lindberg PO (2007) Lagrangian relaxation via ballstep subgradient methods. *Mathematics of Operations Research* 32(3):669–686.

Nedić A, Ozdaglar A (2009) Approximate primal solutions and rate analysis for dual subgradient methods. *SIAM Journal on Optimization* 19(4):1757–1780.

Polyak BT (1967) A general method for solving extremal problems. *Soviet Mathematics Doklady*:5.