Bull. Korean Math. Soc. 28(1991), No. 2, pp. 243–254

DUAL ALGORITHM FOR L_1 ISOTONIC OPTIMIZATION WITH WEIGHTS ON A PARTIALLY ORDERED SET

SEIYOUNG CHUNG

1. Introduction

For a given function $f \in \mathcal{F}$ and a set of functions $\mathcal{J} \subseteq \mathcal{F}$, the problem of isotonic optimization is to determine an element in the set nearest to f in some sense. Specifically, let X be a partially ordered finite set with a partial order $\ll$ and, let $\mathcal{F} = \mathcal{F}(X)$ be the linear space of all bounded real valued functions on X. A function $g \in \mathcal{F}$ is said to be an isotonic function if $g(x) \leq g(y)$ whenever $x, y \in X$ and $x \ll y$. Let $\mathcal{J} = \mathcal{J}(X)$ be the convex cone of isotonic functions on X. As a measure of distance, define a weighted L_p norm on $\mathcal{F}$ by

$$\|h\|_p = \sum_{x \in X} |h_x|^p \omega(x), \qquad 1 \leq p < \infty, h \in \mathcal{F},$$
$$\|h\|_p = \max_{x \in X} |h(x)| \omega(x), \qquad p = \infty, h \in \mathcal{F},$$

for a given weight function $\omega \in \mathcal{F}$, $\omega(x) \geq \sigma > 0$ for all $x \in X$.

These isotonic optimization problems are motivated mainly because of their applications to order restricted statistical analsis. The L_2 version of this problem has been thoroughly discussed. See [1,Chapters 1 and 2]. The *Minimum Lower Sets Algorithm (MLSA)*, which is used most often, is given in [2] and [3]. For the case of total order, the *Pool Adjacent-Violators Algorithm* and the *Up-and-Down Blocks Algorithm* were developed by J.B. Kruskal [4] and by Ayer et al. [5], respectively. The L_p problems, $1 < p \leq \infty$, are considered by Barlow and Ubhaya in

---

Received Decemver 21, 1990.
This paper was supported by RESEARCH FUND for JUNIOR SCHOLARS, Korea Research Foundation.

[13], and by Ubhaya in [14], [15]. The L_1 problem has been considered in [6] through [12]. The MLSA originally developed for the L_2 problem is modified in [12] so that it can be applied to the more general cases which include the L_1 problem as a special case. An algorithm, called *Dual Algorithm*, for the L_1 problem with $\omega \equiv 1$ were developed by S. Y. Chung in [16]. The linearity and hence the duality of the problem are much used in the latter but not in the former. The L_1 isotonic optimization with weights under consideration in this paper is:

$$\begin{aligned}(\mathrm{P}): \ & \text{Given } f \in \mathcal{F}, \text{ find } g^* \in \mathcal{F}, \text{ if one exists, such that} \\ & \|f - g^*\|_1 = \inf\{\|f - g\|_1 \mid g \in \mathcal{F}\}.\end{aligned}$$

To improve the efficiency over the modified $MLSA$, we try to take advantage of linearity. The dual of the problem (P) and the duality theorem are proposed. An algorithm which utilizes Network Flows is constructed that solves both the primal and the dual simultaneously after a finite number of iterations. It is also used to prove the existency of an optimal solution g^* and the duality theorem.

2. Dual problem

For each $x \in X$, the immediate successors of x and the immediate predecessors of x are the sets $U(x) = \{y \in X \mid x \ll y, x \neq y$ and there is no $z \in X$ such that $x \ll z \ll y\}$ and $L(x) = \{y \in X \mid y \ll x,\ x \neq y$ and there is no $z \in X$ such that $y \ll z \ll x\}$, respectively. Define the set $\mathcal{L}$ by $\mathcal{L} = \{(x, y) \mid x \in X, y \in U(x)\}$. We now rephrase the problem (P):

$$\begin{aligned}&\min \|f - g\|_1 \text{ subject to} \\ &\quad (P-1): g(x) \le g(y) \text{ whenever } (x, y) \in \mathcal{L}.\end{aligned}$$

Let h and F be two functions defined on X and $\mathcal{L}$ repectively. Consider the problem :

$$\begin{aligned}(\mathrm{D}): \ & \max \sum_{x \in X} h(x) f(x) \text{ subject to} \\ (\text{D-1}) \ & : -\omega(x) \le h(x) \le \omega(x), \ x \in X,\end{aligned}$$

$$\text{(D-2)} \quad : F(x,y) \geq 0, \ (x,y) \in \mathcal{L},$$
$$\text{(D-3)} \quad : h(x) = \sum_{y \in U(x)} F(x,y) - \sum_{z \in L(x)} F(z,x), \ x \in X.$$

It turns out that the problems (P) and (D) are dual to each other. We will make this more precise. Any function $g \in \mathcal{J}$ is said to be feasible for the primal (P) and any functions h on X and F on $\mathcal{L}$ satisfying the constraints (D-1), (D-2) and (D-3) are said to be feasible for the dual (D). Define $\text{sgn}(x) = 1$ if $x > 0; 0$ if $x = 0; -1$ if $x < 0$. Two conditions, which turn out to be optimal criteria, are defined as:

Condition A : $h(x) = \omega(x) sgn(f(x) - g(x))$ for all x with $f(x) \neq g(x)$.

Condition B : $\sum_{x \in X} h(x) g(x) = 0.$

LEMMA. *Let h be feasible for the dual* (D) *and g for the primal* (P). *The following inequality then holds:*

$$\sum_{x \in X} h(x) f(x) \leq \|f - g\|_1,$$

where equality holds if and only if both Conditions A and B are satisfied.

To prove Lemma, we define an incidence function e_x on $\mathcal{L}$ for each x in X:

$$e_x(y,z) = \begin{cases} 1 \text{ if } x = y, \\ -1 \text{ if } x = z, \\ 0 \text{ otherwise,} \end{cases}$$

and we rephrase the constraint (D-3) as:

$$h(x) = \sum_{(y,z) \in \mathcal{L}} F(y,z) e_x(y,z), \qquad x \in X$$

Proof of Lemma. It follows directly from the constraint (D-1) that the right hand side is greater than or equal to $\sum_{x\in X} h(x)[f(x)-g(x)]$ and that they are equal to each other if and only if Condition A holds. To complete the proof, it suffices to show that $\sum_{x\in X} h(x)g(x) \le 0$.

$$\begin{aligned}\sum_{x\in X} h(x)g(x) &= \sum_{x\in X} g(x)\left[\sum_{(y,z)\in\mathcal{L}} F(y,z)e_x(y,z)\right]\\ &= \sum_{(y,z)\in\mathcal{L}} F(y,z)\left[\sum_{x\in X} e_x(y,z)g(x)\right]\\ &= \sum_{(y,z)\in\mathcal{L}} F(y,z)[g(y)-g(z)]\\ &\le 0,\end{aligned}$$

where the inequality comes from the constraints (P-1) and (D-2).

We have shown that the minimum of the primal (P) is always greater than or equal to the maximum of the dual (D) and hence that the feasible functions are optimal if they are equal. Noting that Condition B is true if and only if $F(y,z)[g(y)-g(z)]=0$ for all $(y,z)\in\mathcal{L}$ in the proof of Lemma, Condition B may be equivalently described as:

Condition B′: $F(x,y)[g(x)-g(y)]=0$ for all $(x,y)\in\mathcal{L}$,

which is usually called *The Complimentary Slackness Condition* for the primal and dual problem.

DUALITY THEOREM. *Under the same assumptions in Lemma,the functions h and g are optimal if and only if they satisfy both Conditions A and B.*

The sufficiency of Duality Theorem is the immediate consequence of Lemma. Assume the necessity is proved. It then follows from the above Lemma that the optimal values are the same. Hence the two problems are dual to each other. The necessity will be proved by constructing an algorithm in the next section.

3. Dual algorithm

Notice that both problems always have the obvious feasible solutions $g \equiv 0, h \equiv 0$ and $F \equiv 0$, which satisfy Condition B. We thus start with them, seek improved feasible functions satisfying Condition B and stop if Condition A is also satisfied.

We may view the given set X with a partial order as a network with the node set X and with the oriented arc set $\mathcal{L}$. Let us augment this network by attaching a node, say x_0, and arcs (x_0, x), $x \in X$. From now on, the nodes in X and arcs in $\mathcal{L}$ are called original and those attached are called augmented. Let $N = X \cup \{x_0\}$ and $A = \mathcal{L} \cup \mathcal{L}_0$, where $\mathcal{L}_0 = \{(x_0, x) \mid x \in X\}$. The network under consideration is the one with the node set N and with the arc set A. With this setting, the funcitons h and ω on X may be regarded as ones on $\mathcal{L}_0$ but the notations $h(x)$ and $\omega(x)$ will be kept instead of the ones $h(x_0, x)$ and $\omega(x_0, x)$. Define a function $\mathcal{K}$ on A by $\mathcal{K} = h$ on $\mathcal{L}_0$ and $\mathcal{K} = F$ on $\mathcal{L}$.

We need some introduction of the painted network which is necessary for developing the algorithm here. This material can be found in Rockafellar [17, Chapters 1 and 2].

A *path* P in a network is a finite sequence : $x_1, J_1, x_2, J_2, x_3, \cdots, J_k, x_{k+1} (k > 0)$, where x_i denotes a node, J_1 an arc and either $J_i = (x_i, x_{i+1})$ or $J_i = (x_{i+1}, x_i)$. When $x_1 = x_{k+1}$, we call P a circuit. An elementary path is a path which uses no node more than once, except of course for the initial node and the terminal node when the path is a circuit. Form now on, by a path we mean an elementary path. The arc J_i in P is said to be traversed positively or negatively according to whether $J_i = (x_i, x_{i+1})$ or $J_i = (x_{i+1}, x_i)$. For a path P, P^+ is the set of positive arcs in P, P^- the set of negative arcs in P and the incidence function for P is definde as:

$$e_P(J) = \begin{cases} 1 \text{ if } J \in P^+ \\ -1 \text{ if } J \in P^- \\ 0 \text{ otherwise} \end{cases}$$

For a given node set S in a network, define the sets :

$$Q^+ = [S, N - S]^+ = \{(x, y) \in A \mid x \in S, \ y \in N - S\}$$
$$Q^- = [S, N - S]^- = \{(x, y) \in A \mid x \in N - S, \ y \in S\},$$

and a *cut* Q in the network as the set $Q = Q^+ \cup Q^-$, which is denoted by $Q = [S, N - S]$. Define the incidence function e_S for a node set S by $e_S(x) = 1$ if $x \in S; 0$ if $x \notin S$.

By *a painted network*, we mean a network each arc in which is painted one of the four colors (green, white, black and red)with the meaning: the green arc is traversable in either direction, the white arc only positively, the black arc only negatively and the red arc is forbidden. For given two nonempty disjoint node sets N^+ and N^- in the painted network, *the Painted Path Problem* involves determining a path $P : N^+ \to N^-$ such that each arc in P^+ is green or white and each arc in P^- is green or black and *the Painted Cut Problem* is to find a cut $Q = [S, N - S]$ with $N^+ \subset S$ and $N^- \cap S = \emptyset$ such that each arc in Q^+ is red or black and each arc in Q^- is red or white. *The Painted Network Algorithm (PNA)* used for the above two problems can be found in [17, pp 33–35].*The Painted Network Theorem* [17, p 39] reads : For given two nonempty disjoint node sets N^+ and N^- in the painted network, one and only one of the painted path problem and the painted cut problem has a solution. This means that the outcome of PNA is either a path or a cut, which is used in Step 3 of our algorithm below.

Notice that without loss of generality, one may assume $f(x) > 0$ for all $x \in X$. One more assumption is: For any $x, y \in X$ with $x \neq y$, a path $P : x \to y$ can be found with the colors disregarded. Otherwise, one could partition X into two or more subsets for each of which our assumption is satisfied and solve the same problem for each partition.

Dual Agorithm

Initially, set $g \equiv 0, h \equiv 0$ and $F \equiv 0$.

Step 1: Given arbitrary functions g and h, set
$UN = \{x \in X \mid g(x) < f(x), -\omega(x) < h(x) < \omega(x)\}$.
Stop if $UN = \emptyset$.

Step 2: Given arbitrary functions g, h, and F, paint the network:

1) Any original arc $(x, y) \in \mathcal{L}$ is painted;
red if $g(x) < g(y), F(x, y) = 0$,
white if $g(x) = g(y), F(x, y) = 0$,

green if $g(x) = g(y), F(x,y) > 0$.

2) Any augmented arc $(x_0, x) \in \mathcal{L}_0$ is painted;

red if $h(x) = \omega(x)\ \mathrm{sgn}(f(x) - g(x)), f(x) \neq g(x)$,

black if $g(x) = f(x), h(x) = \omega(x)$,

white if $[g(x) = f(x), h(x) = -\omega(x)]$ or

$[g(x) < f(x), -\omega(x) < h(x) < \omega(x)]$,

green if $g(x) = f(x), -\omega(x) < h(x) < \omega(x)$.

Step 3: Select $x^* \in UN$ and apply PNA(Painted Network Algorithm) with $N^+ = \{x^*\}$ and $N^- = \{x_0\}$. The same node x^* should be selected as long as it is still in UN at next iteration. If PNA ends up with a circuit P, then go to Step 4. If PNA ends up with a cut $Q = [S, N - S]$, then go to Step 5.

Step 4: Calculate

$$\alpha = min \begin{cases} \omega(x^*) - h(x^*) \\ \omega(y^*) + h(y^*) \\ F(x,y) \text{ for } (x,y) \in P^-, \end{cases}$$

and update $\mathcal{K} : \mathcal{K}' = \mathcal{K} + \alpha e_P$.

Go to Step 1.

Step 5 : Calculate

$$\beta = min \begin{cases} g(y) - g(x) \text{ for } (x,y) \in Q^+ \\ f(x) - g(x) \text{ for } x \text{ with } (x_0, x) \in Q^- \text{ and} \\ \qquad - \omega(x) < h(x) \end{cases}$$

and update $g : g' = g + \beta e_S$.

Go to Step 1.

REMARK 1. Any circuit P has only two augmented arcs $(x_0, x^*) \in P^+$ and $(x_0, y^*) \in P^-$ because of N^+ and N^-.

REMARK 2. h and F are updated only when a circuit P at Step 3 and only at the arcs in P but $g' = g$ after a circuit.

REMARK 3. g is updated only after a cut Q at Step 3 and only on S but $h' = h$ and $F' = F$ after a cut.

Let's define $A_i, i = 1, 2, \cdots, 9$, subsets of A, as following:

$$\begin{aligned}
A_1 &= \{(x_0, x) \in \mathcal{L}_0 \mid g(x) < f(x),\ -\omega(x) < h(x) < \omega(x)\},\\
A_2 &= \{(x_0, s) \in \mathcal{L}_0 \mid g(x) = f(x),\ -\omega(x) < h(x) < \omega(x)\},\\
A_3 &= \{(x_0, x) \in \mathcal{L}_0 \mid g(x) < f(x),\ h(x) = \omega(x)\},\\
A_4 &= \{(x_0, x) \in \mathcal{L}_0 \mid g(x) = f(x),\ h(x) = -\omega(x)\},\\
A_5 &= \{(x_0, x) \in \mathcal{L}_0 \mid g(x) = f(x),\ h(x) = \omega(x)\},\\
A_6 &= \{(x_0, x) \in \mathcal{L}_0 \mid g(x) > f(x),\ h(x) = -\omega(x)\},\\
A_7 &= \{(x, y) \in \mathcal{L} \mid g(x) = g(y),\ F(x, y) = 0\},\\
A_8 &= \{(x, y) \in \mathcal{L} \mid g(x) < g(y),\ F(x, y) = 0\},\\
A_9 &= \{(x, y) \in \mathcal{L} \mid g(x) = g(y),\ F(x, y) > 0\}.
\end{aligned}$$

PROPOSITION 1. *Any arc $(x_0, x) \in \mathcal{L}_0$ is in one of $A_i, i = 1, 2, \cdots, 6$, and any arc $(x, y) \in \mathcal{L}$ is in one of $A_i, i = 7, 8, 9$. Furthermore, α in Step 4 and β in Step 5 are positive.*

Proof. At the beginning, the first assertion holds because of initial setting and of that f, $\omega > 0$ on X. Any original arc $(x, y) \in P^-$ is green and hence $F(x, y) > 0$. But $\omega(x^*) - h(x^*) > 0$ since $x^* \in UN$, and $\omega(y^*) + h(y^*) > 0$ since $(x_0, y^*) \in P^-$ is green or black. Any original arc $(x, y) \in Q^+$ is red and hence $g(y) - g(x) > 0$. Any augmented arc $(x_0, x) \in Q^-$ is red or white, and thus $f(x) - g(x) > 0$ if $-\omega(x) < h(x)$. We therefore have shown that $\alpha > 0$ and $\beta > 0$ provided the first assertion holds. Assume that the first assertion holds before Step 3 at a certain iteration. Let an arc (x_0, x) be in A_1. It then is white, and thus it is in P^+ if in a circuit P and it is in Q^- if in a cut Q. Remark 1 implies $x = x^*$ if $(x_0, x) \in P^+$. It then follows from the definition of α and Remark 2 that any $(x_0, x) \in A_1$ remains in A_1 or becomes belonging to A_3 after Step 4. After a cut, $g'(x) = g(x) + \beta$ since

$(x_0, x) \in Q^-$ implies $x \in S$, and it follows from the definition of β and Remark 3 that any $(x_0, x) \in A_1$ remains in A_1 or becomes belonging to A_2 after Step 5. By the similar manner, we can show where each arc belongs to after updating h, g and F at Step 4 or Step 5. See the table below:

A_1	A_2	A_3	A_4	A_5	A_6	A_7	A_8	A_9
A_2, A_3	A_4	A_5	A_6	A_2, A_4		A_8, A_9	A_7	A_7

In the above table, each A_i is assumed to be in the second row of the i-th column. For example, the first column reads: Any arc in A_1 before Step 3 is contained in A_1, A_2 or A_3 after Step 3.

By the result of Proposition 1, we may revise the painting condition for the red original arc in Step 2:

$$\text{red if } g(x) < g(y).$$

PROPOSITION 2. *The updated functions g', h' and F' are feasible.*

Proof. It is an easy corollary of Proposition 1 that (D-1) and (D-2) are satisfied by h' and F' and (P-1) by g'. Noticing Remarks 2 and 3, we see that it suffices to show that (D-3) holds for h' and F' after a circuit P. For any node $x (\neq x^*, y^*)$ in P, there are only two arcs, say J_1 and J_2, in P that use the node x. Let x_1 and x_2 be the nodes of J_1 and J_2 respectively. If both J_1 and J_2 are either in P^+ or in P^-, then $x_1 \in U(x)$ and $x_2 \in L(x)$ or vice versa. If one of J_1 and J_2 is in P^+ and the other in P^-, the nodes x_1 and x_2 should be either in $U(x)$ or in $L(x)$. In any case, the right hand side of (D-3) remains unchanged by updated F', and hence $h' = h$ except at x^*, y^* implies (D-3) for this case.There is only one original arc J in P that uses the node x^*. If $J \in P^+, J = (x^*, y)$ for $y \in U(x^*)$ and only the first term in the right hand side of (D-3) is increased by α. If $J \in P^-, J = (y, x^*)$ for $y \in L(x^*)$ and the second term is increased by α. But $h' = h + \alpha$ at x^*, which shows that (D-3) holds. For y^*, the same argument can be employed.

PROPOSITION 3. *Condition B is satisfied after each iteration.*

Proof. In Proposition 1, we have shown that any arc $(x, y) \in \mathcal{L}$ is contained in one of A_7, A_8, A_9 at each iteration. But each arc in them satisfies Condition B', which is equivalent to Condition B.

PROPOSITION 4. *h and g are optimal at the termination.*

The above Proposition 4 is a corollary of the sufficiency of Duality Theorm and Proposition 1 through 3 since $(x_0, x) \in A_1$ if and only if $x \in UN$, and since Condition A is violated by a node x only in UN. Up to now we have justified Dual Algorithm. Now we will show that the algorithm is a finite one and hence that the problem always has a solution and the problems (D) and (P) are really dual to each other.

Recall that a set of numbers is said to be *commensurable* if they can all be expressed as whole multiples of a certain number $\lambda > 0$. Certainly, any finite set of rational numbers is commensurable. The commensurability condition in Proposition 5 below is no harm in practice since, for computations, numbers are always rounded off to something rational.

PROPOSITION 5. *Dual Algoritm is finite if the values $f(x)$ and $\omega(x)$ are commensurable.*

Proof. From the table in the proof of Proposition 1, we know that any arc which is not in A_1 remains outside A_1. From that $(x_0, x) \in A_1$ if and only if $x \in UN$, it follows that the updated UN is a subset of UN at previous iteration. Since the set X, and hence UN, is finite, it therefore sufficient to show that once a node $x^* \in UN$ is selected at Step 3 of a certain iteration, it is no longer a member of UN after a finite number of iteration thereafter. Whit a circuit P as the outcome of PNA, the value $h(x^*)$ is increased by α since $(x_0, x) \in P^+$, and with a cut $Q = [S, N - S]$, the value $g(x^*)$ is increased by β because $(x_0, x^*) \in Q^-$ and $x^* \in S$. But the values of f, ω, h, g and F are commensurable at the outset of the algorithm. They are then all multiples of a certain number $\lambda > 0$, and hence so will be the numbers α, β, h', F' and g'. The situation now is self-perpetuating, and it follows that at every iteration either $h(x^*)$ or $g(x^*)$ is increased by at least λ provided the same node x^* is selected in a row as long as it is still in UN. Therefore we can conclude that either $h(x^*) = \omega(x^*)$ or $g(x^*) = f(x^*)$, whichever comes first, will occur after a finite number of iterations, which completes the proof.

Proof of the Necessity for Duality Theorem. We have shown that whenever the algorithm starts with feasible funtions it produces the optimal solutions and the optimal values are the same. Together with the fact that there always exist feasible functions, Lemma completes the proof.

References

1. R. E. Barlow, D. J. Bartholomew, J. M. Bremner and H. D. Brunk, *Statistical Inference Under Order Restrictions,* New York: Wiley, 1972.
2. H. D. Brunk, G. M. Ewing and W. R. Utz, *Minimizing integrals in certain classes of monotone functions,* Pacif. J. Math. **7** (1957), 833–847.
3. H. D Brunk, *Maximum Likelihood estimates of monotone parameters,* Ann. Math. Statist. **26** (1955), 607–616.
4. J. B. Kruskal, *Nonmetric multidimensional scaling: a numerical method,* Psychometrica **29** (1964), 115–129.
5. M. Ayer, H. D. Brunk, G. M. Ewing, W. T. Reid and E. Silverman, *An empirical distribution functions of sampling with incomplete information,* Ann. Math. Statist. **26** (1955), 641–647.
6. T. Robertson and P. Waltman, *On estimating monotone parameters,* Ann. Math. Statist. **39** (1968), 1030–1039.
7. J. D. Cryer, T. Robertson, F. T. Wright and R. J. Cassidy, *Monotone median regression,* Ann. Math. Statist. **43** (1972), 1459–1469.
8. T. Robertson and F. T. Wright, *Multiple isotonic median regression,* Ann. Statist. **1** (1973), 422–432.
9. ——— *A norm reducing property for isotonized Cauchy mean value functions,* Ann. Statist. **2** (1974), 1302–1307.
10. ——— *Consistency in generalized isotonic regression,* Ann. Statist. **3** (1975), 350–362.
11. ——— *Statistical inferences for ordered parameters: A personal view of isotonic regression since the work by Barlow, Bartholomew, Bremner and Brunk,* Proceedings of Symposia in Applied Mathematics **23** (1980), 55–71.
12. ——— *Algorithms in order restricted statistical inference and the Cauchy mean value property,* Ann. Statist. **8** (1980), 645–651.
13. R. E. Barlow and V. A. Ubhaya, *Isotonic approximation,* InProceedings of the Symposium on Optimizing Methods in Statistics, New York: Academic Press, 1971.
14. V. A. Ubhaya, *Isotonic Optimization I,* J. Approximation Theory **12** (1974), 146–159.

15. ——— *Isotonic Optimization II*, J. Approximation Theory **12** (1974), 315–331.
16. S. Y. Chung, *Dual algorithm for L_1 isotonic optimization on a partially ordered set*, Dissertation for the Degree of Ph. D, ISU Math. Dept.,Ames, Iowa,1987.
17. R. T. Rockafellar, *Network flows and monotropic optimization*, New York: Wiley-Interscience, 1984.

Department of Mathematics,Chungnam National University, Taejon 305-764, Korea