ADVANCES IN OPTIMIZATION AND PARALLEL COMPUTING
P.M. Pardalos (Editor)

The Dual Active Set Algorithm

William W. Hager[1]

Abstract

A new algorithm, the Dual Active Set Algorithm, is presented for solving a minimization problem with equality constraints and bounds on the variables. The algorithm identifies the active bound constraints by maximizing an unconstrained dual function in a finite number of iterations.

1 Introduction

The dual approach to constrained optimization has much theoretical appeal. Complicated constraints can be imbedded in a dual variational form, and at an unconstrained maximum of the dual problem, the constraints of the primal problem are satisfied. On the other hand, numerical experience with dual optimization problems can be disappointing. For example, in the numerical experiments reported in [5], the Conjugate Gradient Method applied to a dual optimal control problem converged extremely slowly. In [5] we presented a new algorithm for dual optimal control problems, and we proved local quadratic convergence. In this note the algorithm of [5] is modified in order to obtain a globally convergent scheme that we call the Dual Active Set Algorithm. This scheme can be used to solve general optimization problems with equality constraints and bounds on the variables. In a separate paper [4], this new algorithm is applied to quadratic networks. It is observed that a 2-phase procedure consisting of a few iterations of the Conjugate Gradient Method followed by the Dual Active Set Algorithm yields extremely rapid convergence.

2 The algorithm

Given a real-valued function F and a function h mapping $\mathbb{R}^n$ to $\mathbb{R}^m$, the associated Lagrangian L is

$$L(\lambda, x) = F(x) + \sum_{i=1}^{m} \lambda_i h_i(x).$$

Let l and u be vectors in $\mathbb{R}^n$ with $l < u$ (vector equalities and inequalities are interpreted componentwise). We consider the constrained dual function $L(\lambda)$ defined by

$$L(\lambda) = \inf L(\lambda, x) \text{ subject to } l \leq x \leq u. \tag{1}$$

The Dual Active Set Algorithm is designed to solve the problem:

$$\text{maximize } L(\lambda) \text{ over } \lambda. \tag{2}$$

[1] Department of Mathematics, University of Florida, Gainesville, FL 32611. This research was supported by the U.S. Army Research Office Contract DAAL03-89-G-0082.

Problem (2) arises when one takes the dual to the constrained optimization problem

$$\text{minimize } f(x) \text{ subject to } h(x) = 0,\ l \le x \le u, \tag{3}$$

where f is a real-valued function. The "ordinary" Lagrangian corresponds to the choice $F(x) = f(x)$. If f is convex, h is affine, and there exists a solution x^* to (3), then there exists a solution λ^* to the dual problem (2) and $x = x^*$ achieves the minimum in (1). Conversely, if for some given λ, a minimizing x in (1) has the property that $h(x) = 0$, then λ is optimal in the dual (2) and x is optimal in the primal (3). If f is strongly convex, h is affine, and the feasible set for (3) is nonempty, then there exists a unique solution to (3) that can be computed in the following way: First, we obtain a λ^* that achieves the maximum in (2). The x that attains the minimum in (1) when $\lambda = \lambda^*$ is the solution to the primal problem (3).

If f lacks convexity, duality can still be employed, however, we need to augment the Lagrangian with a penalty term. The "augmented" Lagrangian obtained by taking $F(x) = f(x) + p|h(x)|^2$ where p is a positive parameter and $|\cdot|$ denotes the Euclidean norm is often called the quadratic penalty augmented Lagrangian. In a neighborhood of an optimum and for p sufficiently large, an augmented Lagrangian has many of the properties of the ordinary Lagrangian associated with a convex cost function (see [1] or [10]).

In stating the Dual Active Set Algorithm, the following notation is employed: If B is a subset of $\{1, 2, \ldots, n\}$ and x is a vector, then x_B denotes the vector with components x_i associated with indices $i \in B$. Given a vector $z \in \mathbb{R}^n$, we define modifications L_B and L_B^z of L by the rule

$$L_B(\lambda) = \inf L(\lambda, x) \text{ subject to } l_B \le x_B \le u_B, \tag{4}$$

and

$$L_B^z(\lambda) = \inf L(\lambda, x) \text{ subject to } x_B = z_B. \tag{5}$$

Note that the variable x_i in (4) and (5) is unconstrained if $i \notin B$.

If λ_k is the current approximation to a solution of (2), then λ_{k+1} is computed by the following procedure:

Dual Active Set Iteration

Given λ_k, let $j = 0$, $\nu_0 = \lambda_k$, and define

$$B_0 = \{i : z_i = l_i \text{ or } z_i = u_i\}$$

where

$$z = x(\lambda_k) = \arg\min L(\lambda_k, x) \text{ subject to } l \le x \le u.$$

Subiteration: Let μ_j maximize $L^z_{B_j}(\lambda)$ over λ and define $\mu(t) = \nu_j + t(\mu_j - \nu_j)$. Determine the largest interval $[0, \bar{t}\,]$, $\bar{t} \geq 0$, such that

$$L^z_{B_j}(\mu(t)) = L_{B_j}(\mu(t)) \text{ for every } t \in [0, \bar{t}\,]. \tag{6}$$

If $\bar{t} < 1$, then put $\nu_{j+1} = \mu(\bar{t})$. The set B_{j+1} is obtained by deleting from B_j those indices $i \in B_j$ with the property that

$$\frac{\partial L(\lambda, x)}{\partial x_i}\bigg|_{\lambda=\bar{\lambda},\, x=\bar{x}} = 0,$$

where $\bar{x}$ is a minimizer in (5) associated with $\lambda = \bar{\lambda} = \mu(\bar{t})$. Increment j and repeat the subiteration.

If $\bar{t} \geq 1$, then set $\lambda_{k+1} = \mu_j$, increment k, and proceed to the next iteration.

Note that the scheme in [5], which is equivalent to taking $\lambda_{k+1} = \mu_0$ in each iteration, only converges in special situations. The phrase "conceptual form" was inserted above since the maximizer μ_j may not exist. As discussed in Section 4, lack of existence can be handled using the "proximal point" regularization.

3 Convergence

The convergence of Dual Active Set Algorithm is examined under a strong convexity assumption. That is, we assume that there exists a constant $\alpha > 0$ such that

$$L(\lambda, y) \geq L(\lambda, x) + \nabla_x L(\lambda, x)(y - x) + \alpha|y - x|^2, \tag{7}$$

where α is independent of x, y, and λ.

Theorem 1 *If F and h are continuously differentiable on $I\!R^n$, L satisfies the strong convexity assumption* (7), *and there exists a maximizer μ_j of $L^z_{B_j}$ for each j, then the Dual Active Set Algorithm reaches a solution of* (2) *in a finite number of iterations and subiterations.*

Proof. If z attains the minimum in (1) when $\lambda = \lambda_k$, then by the convexity assumption, $L^z_{B_0}(\lambda_k) = L_{B_0}(\lambda_k) = L(\lambda_k)$. By the continuous differentiability of F and h and by (7), there exists a minimizer $x(t)$ in (5) associated with $\lambda = \mu(t)$, and $x(t)$ is a continuous function of t (see [3], Theorem 4.1). If $\bar{t} < 1$ in the subiteration and

$$\frac{\partial L(\mu(\bar{t}), x(\bar{t}))}{\partial x_i} \neq 0$$

for every $i \in B_j$, then by the continuity of $x(t)$ and by convexity, it follows that $L^z_{B_j}(\mu(t)) = L_{B_j}(\mu(t))$ for t in a neighborhood of $\bar{t}$. But this violates the fact that $[0,\bar{t})$ was the largest interval where (6) holds. We conclude that B_{j+1} is strictly contained in B_j, and the subiteration terminates in a finite number of steps.

By the convexity assumption, the fact that $L^z_B(\lambda)$ is a concave function of λ for any choice of B, and μ_j maximizes $L^z_{B_j}(\lambda)$ over λ, we have

$$L^z_{B_j}(\nu_j) \le L^z_{B_j}(\nu_{j+1}) = L^z_{B_{j+1}}(\nu_{j+1}).$$

If C_k denotes the final set B_j generated in the subiteration, we have

$$L^z_{C_k}(\lambda_{k+1}) = L_{C_k}(\lambda_{k+1}) \le L(\lambda_{k+1})$$

since $L_B(\lambda) \le L(\lambda)$ for any choice of B. Combining these inequalities gives

$$L(\lambda_k) \le L^z_{B_0}(\nu_1) \le L^z_{B_1}(\nu_2) \le \cdots \le L^z_{C_k}(\lambda_{k+1}) = L_{C_k}(\lambda_{k+1}) = L(\lambda_{k+1}).$$

If $\nabla L(\lambda_k) = 0$, then by the concavity of L, λ_k maximizes the dual function. Conversely, if $\nabla L(\lambda_k) \neq 0$, then

$$\nabla L(\lambda_k) = \nabla L_{B_0}(\lambda_k) = \nabla L_{B_0}(\nu_0) \neq 0$$

(see Clarke [2], Proposition 1.13 and Theorem 2.1). Since $L^z_{B_j}$ is concave, μ_j maximizes $L^z_{B_j}$, and ν_{j+1} lies between ν_j and μ_j, it follows that $L^z_{B_j}(\nu_j) < L^z_{B_j}(\nu_{j+1})$ for either $j = 0$ or $j = 1$. Hence, when $\nabla L(\lambda_k) \neq 0$, we have

$$L(\lambda_k) < L^z_{C_k}(\lambda_{k+1}) \le L(\lambda_{k+1}).$$

Since the components of z_{C_k} are chosen from a finite set, the Dual Active Set Algorithm reaches a solution of (2) in a finite number of iterations, which completes the proof.

Now, let us consider the existence of a maximum for $L^z_B(\lambda)$. Under standard convexity assumptions and a constraint qualification, we have

$$\max_{\lambda} L^z_B(\lambda) = \inf\{F(x) : h(x) = 0,\ x_B = z_B\}. \tag{8}$$

However, when the minimization problem in (8) is infeasible, the maximum in (8) is often $+\infty$. Hence, whenever the Dual Active Set Algorithm encounters a set B for which the minimization problem in (8) is infeasible, there does not exist a maximum in the subiteration.

4 Implementable algorithm

The existence of a maximum in the subiteration can be ensured using the proximal point regularization. References for the Proximal Point Algorithm include the papers [8] and [9] by Martinet, [11] by Rockafellar, [7] by Luque, [12] by Spingarn, and [6] by Ha. The basic idea behind the proximal point regularization is the following: The dual function is concave since it is defined in terms of a minimization. Hence, if Λ is a fixed vector in $\mathbb{R}^m$ and ϵ is a positive constant, then the function M defined by

$$M(\lambda) = L(\lambda) - \epsilon|\lambda - \Lambda|^2$$

is a strongly concave function that achieves a maximum. If Λ_k is the current iterate in the Proximal Point Algorithm, then the new iterate Λ_{k+1} is computed by maximizing the M associated with the choice $\epsilon = \epsilon_k$ and $\Lambda = \Lambda_k$. If the ϵ_k are sufficiently small, then the Proximal Point Algorithm often convergences linearly; if the ϵ_k tend to zero, then the convergence is superlinear. For example, suppose that (2) has a maximizer λ^* and there exists a neighborhood Ω of λ^* and an $\alpha > 0$ such that

$$L(\lambda) \leq L(\lambda^*) - \alpha|\lambda - \lambda^*|^2 \text{ for each } \lambda \in \Omega. \tag{9}$$

Then we have (see [11]):

Lemma 1 *If* (9) *holds and the* ϵ_k *are uniformly bounded, then for* k *sufficiently large,* Λ_{k+1} *satisfies the inequality*

$$|\Lambda_{k+1} - \lambda^*| \leq \frac{\epsilon_k}{\epsilon_k + \alpha}|\Lambda_k - \lambda^*|.$$

The Dual Active Set Algorithm applies to M as well as to L, we only need to replace each L in the statement of the algorithm by an M. The regularized forms of the modified functions are

$$M_B(\lambda) = L_B(\lambda) - \epsilon|\lambda - \Lambda|^2 \text{ and } M_B^z(\lambda) = L_B^z(\lambda) - \epsilon|\lambda - \Lambda|^2.$$

The convergence proof given earlier also applies to the regularized problem, but with each L in the proof replaced by an M.

Corollary 1 *If* F *and* h *are continuously differentiable on* $\mathbb{R}^n$ *and* L *satisfies the strong convexity assumption* (7), *then the Dual Active Set Algorithm with* L *replaced by* M *generates a maximum of* M *in a finite number of iterations and subiterations.*

References

[1] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, 1982.

[2] F. H. Clarke, Generalized gradients and applications, *Transactions of the American Mathematical Society* 205 (1975), 247-262.

[3] W. W. Hager, Inequalities and approximation, In *Constructive Approaches to Mathematical Models*, C. V. Coffman and G. J. Fix, eds., Academic Press, New York, 1979, 189-202.

[4] W. W. Hager and D. W. Hearn, The dual active set algorithm and quadratic networks, Report No. 90-7, Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL, June, 1990.

[5] W. W. Hager and G. D. Ianculescu, Dual approximations in optimal control, *SIAM Journal on Control and Optimization* 22 (1984), 423-465.

[6] C. D. Ha, A generalization of the proximal point algorithm, *SIAM Journal on Control and Optimization* 28 (1990), 503-512.

[7] F. J. Luque, Asymptotic convergence analysis of the proximal point algorithm, *SIAM Journal on Control and Optimization* 22 (1984), 277-293.

[8] B. Martinet, Regularisation d'inèquations variationelles par approximations successives, *Revue Francaise Informatique et Recherche Opérationnelle* (1970), 154-159.

[9] B. Martinet, Determination approachée d'un point fixe d'une application pseudo-contractante, *Comptes Rendus des Séances de l'Académie des Sciences, Paris* 274 (1972), 163-165.

[10] R. T. Rockafellar, Augmented Lagrange multiplier functions and duality in non-convex programming, *SIAM Journal on Control* 12 (1974), 268-285.

[11] R. T. Rockafellar, Monotone operators and the proximal point algorithm, *SIAM Journal on Control and Optimization* 14 (1976), 877-898.

[12] J. E. Spingarn, Submonotone mappings and the proximal point algorithm, *Numerical Functional Analysis and Optimization* 4 (1982), 123-150.