1 Primal-dual constraint aggregation method

Consider a convex optimization problem in the form

$$\min\{f(x)|x \in X\} \tag{P}$$

subject to

$$Ax = b,$$

where $f(\cdot) : \mathbf{R}^n \to \mathbf{R}$ is convex, $A : \mathbf{R}^n \to \mathbf{R}^m$, $b \in \mathbf{R}^m$, X is convex and compact. We will also use $\mathcal{X}$ to denote the feasible domain of problem (P).

The method described below relies on the following basic assumption:

A1: the set of optimal solutions to (P) is nonempty and if (x^*, p^*) is a primal-dual optimal solution pair, $(x^*, p^*) \in KKT(P)$, then

$$0 \in \partial f(x^*) + A^T p^* + N_X(x^*),$$

where $\partial f(u)$ is the subdifferential of f at u, and $N_X(u)$ is the normal cone to X at u:

$$N_X(u) = \{z \in \mathbf{R}^n | \langle z,\ y - u\rangle \le 0,\ y \in X\},$$

(see, e.g., [**?**]).

For the assumption to hold it is sufficient that any of the following conditions is satisfied ([**?**]):

i) ri $X \cap \{x|Ax = b\} \neq \emptyset$;

ii) X is polyhedral.

At each iteration of the method, the original feasible set $\mathcal{X}$ is replaced by a relaxed set formed by a number of aggregate constraints (even a single constraint is possible) that are dependent on the current solution approximation $x^c \in X$. Namely, let I_j^c, $j = 1, \ldots, J_c$ form a partition of the index set $\{1, \ldots, m\}$:

$$\cup_j I_j^c = \{1, \ldots, m\}, \quad I_j^c \cap I_l^c = \emptyset.$$

Define a set of aggregate constraints as follows

$$\langle [Ax^c - b]_{I_j^c},\ Ax - b\rangle = 0, \quad j = 1, \ldots, J_c, \tag{1}$$

where the ith component of $[h]_S$ is h_i if $i \in S$ and 0 otherwice. If for some j, the aggregate $[Ax^c - b]_{I_j^c}$ turns out to be identically 0 then the corresponding constraint is absent.

In the special case where the partition consists of a single element including all indices of the original set, the aggregate constraint obviously becomes

$$\langle Ax^c - b,\ Ax - b\rangle = 0. \tag{2}$$

Constraint aggregation method with the aggregate as (2) was proposed in [?] and its regularized version was studied in [?]. However, in [?, ?], quite strong requirements had to be imposed on parameters of the method. The present approach aims at overcoming this difficulty by considering the iterative process in the primal-dual space. At the same time, the computational burden required at each iteration is basically the same as in [?, ?].

Theoretical convergence analysis of the method will be conducted for the case of the single aggregate though all arguments remain true for multiple aggregates as well. Also note that in general neither the number of aggregates J_c nor their composition need remain constant in the course of the iteration process.

Denote by $\mathcal{X}_c$ the set defined by the (2):

$$\mathcal{X}_c = \{x \in X | \langle Ax^c - b,\ Ax - b\rangle = 0\} \tag{3}$$

Since aggregate constraints are linear combinations of the original constraints, one has $\mathcal{X} \subseteq \mathcal{X}_c$, i.e., the aggregate set $\mathcal{X}_c$ is indeed a relaxation of the original set.

Now define the subproblem to be solved at each iteration of the method. Let $x^c \in X$, $p^c \in \mathbf{R}^m$ be given. Consider the problem

$$\min\{f(x) + \langle p^c,\ Ax - b\rangle + \gamma_c|x - x^c|^2/2 | x \in \mathcal{X}_c\}. \tag{4}$$

Here $|\cdot|$ denotes the Euclidean norm, and γ_c is a parameter.

An important question concerning the subproblem is that how easily it can be solved. If $f(x)$ is linear and X is a nonnegative orthant or a box, the solution procedure is easy. Another case that can be reduced to the previus one (however, at some cost) is where

$$f(x) = \langle Hx,\ x\rangle + \langle \ell,\ x\rangle,$$

with $H = D^T PD$, and P being diagonal. Introducing new artificial variables $s = Dx$ one obtains a quadratic objective function with a diagonal hessian so that (4) becomes simple. An important example of the objective function f of the form just mentioned comes from portfolio optimization using mean-variance approach []. Here H is a variance-covariance matrix of random payoffs of assets.

However, if one wishes to consider a more general objective function and/or more aggregate constraints, one can only guarantee a solution of (4) up to a certain tolerance. To be more precise, we consider $u^c \in X$ to be an approximate solution to (4) if there exists $w \in \partial f(u^c)$ such that

$$\langle w + A^T p^c + \gamma_c(u^c - x^c),\ x - u^c\rangle \ \geq\ -\mu_c,\ x \in \mathcal{X}_c, \tag{5}$$

$$|\langle Ax^c - b,\ Au^c - b\rangle| \ \leq\ \eta_c \tag{6}$$

for some nonnegative μ_c and η_c. We will show that if μ_c and η_c are small enough then the convergence results will not be much affected by the presence of such inaccuracy. This in some sense characterizes stability of the method studied.

In sequel, we do not focus on particular technique for solving the subproblem at each iteration, thus making a "black-box" assumption. Note only that if major difficulties stem from a big number of constraints (the case we are mostly interested in), aggregation can help handling these problems because it reduces the dual space in the subproblem so that, for example, bundle methods (see, e.g., []) for the dual function can be efficiently applied.

Now let us define the basic version of the primal-dual constraint aggregation method for solving (P). Some modifications of the method are to be specified later.

Method 1. Let (x^0, p^0) be fixed, $x^0 \in X$;

(1) given (x^k, p^k), $x^k \in X$, and the parameters $\gamma_k > 0$ and $\mu_k > 0$ and η_k, solve the subproblem (4) (with $c = k$) to obtain u^k such that (5) and (6) are satisfied;

(2) choose the values of the step-size multipliers α_k and β_k and set

$$x^{k+1} = x^k + \alpha_k(u^k - x^k),$$

$$p^{k+1} = p^k + \beta_k(Au^k - b);$$

(3) set $k = k + 1$ and continue with (1).

In the form just defined the method would produce an infinite sequence $\{(x^k, p^k)\}$. An implementable stopping rule is to be discussed below.

The following theorem establishes convergence of the method 1 under specific choice of the parameters α_k, β_k, γ_k, μ_k, η_k.

Theorem 1. *Let L_A be defined as follows*

$$L_A = \max\{|Ah|^2| \quad |h| = 1\};$$

(a) let γ_k be positive for all $k = 0, 1, \ldots$;

(b)

$\mu_k \leq (\gamma_k/2)|u^k - x^k|^2$ *(see remark 3 after the theorem),*

$\eta_k \leq (1/4)|Ax^k - b|^2$;

(c) α_k satisfy any of the following rules

A. $0 < \alpha_k \leq \frac{1}{2(1+L_A/\gamma_k^2)}$;

B. $0 < \alpha_k \leq \frac{|u^k - x^k|^2}{2(|u^k-x^k|^2 + |Au^k - b|^2/\gamma_k^2)}$;

(d) $\beta_k = \alpha_k/\gamma_k$.

Then

(i) the sequence $\{x^k\}$ generated by method 1 is bounded and its arbitrary limit point belongs to the set of optimal solutions of problem (P) $\mathcal{X}_*$;

(ii) for any $(x^*, p^*) \in KKT(P)$ and arbitrary k one has the estimate

$$|x^{k+1} - x^*|^2 + |p^{k+1} - p^*|^2 \leq |x^k - x^*|^2 + |p^k - p^*|^2$$
$$-(\alpha_k/4)|u^k - x^k|^2,$$

i.e., the distance to the optimal primal-dual solution set is monotonically decreasing on the trajectories of the method.

As a special case one may consider setting $\gamma_k \equiv \gamma$. Then rule A would result in a constant positive step-size. As will be shown, rule B will also guarantee a step-size uniformly bounded from zero.

Proof. Assumption A1 implies that for some $w^* \in \partial f(x^*)$

$$\langle w^* + A^T p^*,\ u^k - x^* \rangle \geq 0 \tag{7}$$

where (x^*, p^*) is an arbitrary KKT(P) point and k is an arbitrary iteration index. By definition of p^{k+1}

$$\begin{aligned} \langle A^T p^*,\ u^k - x^* \rangle &= \langle p^* - p^k,\ Au^k - b \rangle + \langle A^T p^k,\ u^k - x^* \rangle \\ &= (1/\beta_k)\langle p^* - p^k,\ p^{k+1} - p^k \rangle + \langle A^T p^k,\ u^k - x^* \rangle. \end{aligned} \tag{8}$$

Substitute $x = x^*$ and $c = k$ in (5):

$$\langle w + A^T p^k + \gamma_k(u^k - x^k),\ x^* - u^k \rangle \geq -\mu_k.$$

Rearranging the terms and recalling condition (b) on μ_k write this as follows

$$\langle w + A^T p^k,\ x^* - u^k \rangle - (\gamma_k/2)|u^k - x^k|^2 \geq \gamma_k \langle u^k - x^k,\ x^k - x^* \rangle \tag{9}$$

Now substitute (8) in (7) and multiply (7) and (9) by β_k. Noting that $\beta_k = \alpha_k/\gamma_k$ and $\langle w - w^*,\ x^* - u^k \rangle \leq 0$ add up (8) and (9) which results in the following estimate

$$\begin{aligned} &-(\alpha_k/2)|u^k - x^k|^2 \geq \\ &\alpha_k \langle u^k - x^k,\ x^k - x^* \rangle + \langle p^{k+1} - p^k,\ p^k - p^* \rangle. \end{aligned} \tag{10}$$

Consider the merit function $d(x,p) = |x - x^*|^2 + |p - p^*|^2$, where x^* and p^* are the same as above.

By definition of the method one may write

$$\begin{aligned} d^2(x^{k+1}, p^{k+1}) &= d^2(x^k, p^k) + 2\langle p^{k+1} - p^k,\ p^k - p^* \rangle \\ &+ 2\alpha_k \langle u^k - x^k,\ x^k - x^* \rangle + \alpha_k^2 |u^k - x^k|^2 \\ &+ \beta_k^2 |Au^k - b|^2 \end{aligned}$$

The right-hand side may be estimated using (10):

$$d^2(x^{k+1}, p^{k+1}) \leq d^2(x^k, p^k) - (\alpha_k - \alpha_k^2)|u^k - x^k|^2 + \beta_k^2|Au^k - b|^2 \qquad (11)$$

Using inequality (6) and condition (b) one can provide the following estimates for the residual of the system of constraints at x^k and u^k:

$$\begin{aligned} |Au^k - b|^2 &\leq -|Ax^k - b|^2 + |A(u^k - x^k)|^2 + 2\varepsilon_k \\ &\leq -(1/2)|Ax^k - b|^2 + |A(u^k - x^k)|^2 \end{aligned} \qquad (12)$$

From (11) one can see that if the step-size α_k is choosen according to the rule (B) then (11) transforms into

$$d^2(x^{k+1}, p^{k+1}) \leq d^2(x^k, p^k) - (\alpha_k/4)|u^k - x^k|^2.$$

Note that this step-size choice rule is derived from minimizing the right-hand side in (11) with respect to α_k. Because of (12) α_k is uniformly separated from zero at least if γ_k is uniformly positive.

Using again (12) one can further estimate (11) as follows

$$d^2(x^{k+1}, p^{k+1}) \leq d^2(x^k, p^k) - (\alpha_k - \alpha_k^2 - (L_A\alpha_k^2/\gamma_k^2))|u^k - x^k|^2.$$

From here one derives the same formula as above using the step-size choice rule A. Statement (ii) of the theorem is proved.

In particular, statement (ii) implies that the the sequences $\{x^k\}$, $\{p^k\}$ generated by the method is uniformly bounded and moreover, $|u^k - x^k| \to 0$. Then, inequality (12) guarantees that the limitting set of both sequences $\{x^k\}$ and $\{u^k\}$ lies in the fesible domain $\mathcal{X}$. Setting $x = x^*$ in (5) and recalling the definition of μ_k one obtains

$$f(u^k) - f_* \leq O(|u^k - x^k|)$$

which means that in fact, both sequences $\{x^k\}$ and $\{u^k\}$ converge to $\mathcal{X}^*$. The proof is complete.

Remark 1. Let θ_c denote the optimal dual multiplier to the aggregate constraint (2) at iteration c. It follows straightforwardly from (5) that if

$$\lim \theta_c(Ax^c - b) = 0$$

then the sequence of p^k is convergent to a dual optimal solution of (P). In this case, statement (ii) of the theorem would imply that the sequence of primal-dual pairs (x^k, p^k) converges to a particular point in KKT(P).

Remark 2. Step-size choice rule A is derived from B by using the "conservative" estimate (12) via the constant L_A. Of course, in practice B is preferable since allows a better use of the local information. Also note, that if one sets $\mu_k \equiv 0$, then α_k can be taken twice as big as in the rules A or B. This is directly seen from the proof of the theorem.

Remark 3. It is straightforward to generalize the proof to the case of multiple aggregate constraints. Note that for any iterate x^c adding up (1) yields $\langle Ax^c - b,\ Ax - b\rangle$. If one requires condition (6) as for the case of a single aggregate, the proof then goes without change.

Remark 4. The rule prescribing the choice of μ_k in condition (b) may seem to be unimplementable since it depends on u^k which is yet to be calculated. However, due to (12) one has a bound

$$|Ax^k - b|^2 \leq 2L_A|u^k - x^k|^2$$

whenever u^k satisfies (6). (Note that η_k in (6) depends on x^k already available.) Therefore it would be safe to require that

$$\mu_k \leq (\gamma_k/4L_A)|Ax^k - b|^2,$$

which is, however, a conservative estimate.

As is easily seen the errors μ_k and η_k allowed at each iteration of method 1 tend to zero and moreover, $\sum \mu_k < \infty$ and $\sum \eta_k < \infty$. However, we do not have to specify the entire sequences of μ_k and η_k apriori which would have posed a problem in practice. Instead, at a current iteration the magnitude of admissible errors can be determined using the past information from the method.

Remark 5. The method provides estimates of the quality of the current solution approximation and hence an implementable stopping rule. From statement (ii) of the theorem it follows that the limitting set of the sequences $\{x^k\}$ and $\{u^k\}$ coincide. Hence $\{u^k\}$ is also convergent to an optimal solution of problem (P) and we provide the error bounds for u^c at iteration number c. First note that by (12) for each c

$$L_A^{1/2}|u^c - x^c| \geq |Au^c - b|.$$

Consider (5) and set x to be an arbitrary optimal solution to (P). Then recalling (b)

$$((\gamma_c/2)|u^c - x^c| + \gamma_c d + |p^c|L_A^{1/2})|u^c - x^c| \geq f(u^c) - f_*,$$

with d being diameter of X. It follows from theorem 1 that p^c are bounded, and $|u^c - x^c| \to 0$. Therefore $|u^c - x^c|$ may serve as an easily computable measure of closeness of u^c to both feasible domain $\mathcal{X}$ and the optimal solution set of (P).

Remark 6. Complexity bound. Based on remark 5 and statement (ii) of theorem 1 one can provide the following bound on the number of iterations to be performed by method 1 in order to achieve an absolute accuracy $\delta > 0$ in the sense that

$$|Au^c - b| \le \delta,$$
$$f(u^c) - f_* \le \delta.$$

Let r denote the distance from the initial solution approximation (x^0, p^0) in method 1 to the set KKT(P), and let C be the maximal of the multipliers standing on the right-hand side in the estimates in remark 5:

$$C = \max\{L_A^{1/2}, 2\gamma_c d + sL_A^{1/2}\},$$

(we have bounded $|p^c|$ with s and $|u^c - x^c|$ with d). Let N be a maximum iteration number such that $|u^c - x^c| > \delta/C$ for all $c \le N$. Statement (ii) of theorem 1 together with rules A and B for choosing the step-size multiplier imply that

$$N \le \frac{8(1 + L_A/\gamma^2)r^2C^2}{\delta^2}. \tag{13}$$

According to [?] such estimate can not be improved by an order of δ uniformly with respect to the variable space dimension by a method that only uses information provided by the functions involved and its gradients. Method 1, however, does not completely fall into this class because the subproblem requires global information about f at each iteration. Below we consider a version of method 1 that uses only local information while still having a similar complexity bound.

1.1 Method 2

In the previous subsection we have made a "black box" assumption about the procedure for solving the auxiliary subproblem. The question whether this assumption is justified, in particular, depends on the objective function f. Below we describe a modification of method 1 that gives an explicit procedure for dealing with possible complications due to f. More specifically, we incorporate in method 1 an inner cycle that solves the auxiliary problem (4) by successive linearization of the objective function. It will be shown that the inner cycle terminates after a finite number of steps and also we provide the complexity bound for the method with linearization similar to (13).

We make the following additional assumption:

A2: function f is differentiable and its gradient is Lipschtitz continuous with constant $L_{\nabla f}$:

$$|\nabla f(x) - \nabla f(y)| \le L_{\nabla f}|x - y|.$$

We also assume for simplicity that auxiliary subproblems of the inner cycle are solved precisely.

Method 2.

Let (x^0, p^0) be fixed, $x^0 \in X$;
(1) given (x^k, p^k), and $\gamma_k > 0$ set $x^{k,0} = x^k$;
(1.1) (inner cycle): given $x^{k,i}$, compute

$$v^{k,i} = \text{argmin}\{\langle \nabla f(x^{k,i}) + A^T p^k + \gamma_k (x^{k,i} - x^k),\ x - x^{k,i}\rangle$$
$$+|x - x^{k,i}|^2/2\} \tag{14}$$

subject to

$$x \in X$$
$$\langle Ax^k - b,\ Ax - b\rangle = 0,$$

(1.2) choose the step-size λ_i and set

$$x^{k,i+1} = x^{k,i} + \lambda_i (v^{k,i} - x^{k,i});$$

(1.3) for the first i such that STOP, set $u^k = x^{k,i}$;
end of inner cycle;
(2) choose α_k and β_k and set

$$x^{k+1} = x^k + \alpha_k (u^k - x^k),$$
$$p^{k+1} = p^k + \beta(Au^k - b);$$

(3) set $k = k + 1$, continue with (1).

STOP:

$$(\gamma_k/2)|x^{k,i} - x^k|^2) \geq d|v^{k,i} - x^{k,i}|,$$

d is the diameter of the set X.

Remark 2.1. Finiteness of the inner cycle. Under appropriate choice of the step-size multiplier λ_i the inner cycle becomes a gradient projection method for the problem

$$\min\{\phi_k(x) | x \in \mathcal{X}_k\}$$

with

$$\phi_k(x) = f(x) + \langle p^k,\ Ax - b\rangle + (\gamma_k/2)|x - x^k|^2.$$

Standard convergence results with regard to this method (see, e.g., []) yield that the sequence $\{x^{k,i}\}$ converges to the unique optimal solution of the above problem, and moreover, $v^{k,i} - x^{k,i} \to 0$. This implies that STOP will hold after only a finite number of steps.

Now we establish convergence of method 2.

Theorem 2. *Choose γ_k, α_k, and β_k as in theorem 1. Then for the sequence of x^k generated by method 2, statements analogous to (i) and (ii) of theorem 1 are true.*

Proof. Optimality conditions in (14) yield

$$\langle \nabla\phi_k(x^{k,i}) + v^{k,i} - x^{k,i}, x - v^{k,i}\rangle \geq 0$$

for $x \in \mathcal{X}_k$. After some simple tranformations this becomes

$$\begin{aligned} \langle \nabla\phi_k(x^{k,i}),\ x - x^{k,i}\rangle + (\gamma_k/2)|x^{k,i} - x^k|^2 \quad &\geq \\ (\gamma_k/2)|x^{k,i} - x^k|^2 - d|v^{k,i} - x^{k,i}| & \end{aligned}$$

For the index i such that STOP is reached, the right-hand side of the latter inequality is greater or equal zero. Setting $u^k = x^{k,i}$ as in method 2 and denoting $\mu_k = (\gamma_k/2)|u^k - x^k|^2$ leads us to relation (5) with the error μ_k satisfying condition (b) of theorem 1. Then application of theorem 1 yields the desired result. The proof is complete.

Note that method 2 differs from method 1 just in a more detailed description of step (1). Therefore, all remarks made after theorem 1 remain valid also for method 2. We would like to further develop the one concerning complexity analysis.

Remark 2.2. Complexity bound. Method 2 belongs to the class of methods specified in [**?**] for which the bound similar to (13) is tight. We show that if λ_i is chosen to be $1/L_{\nabla\phi}$ and $\gamma_k \equiv \gamma > 0$ then for method 2 one has a complexity bound that differs from (13) only by a factor of $c_1 \log(c_2/\delta)$.

Recall (see, e.g., [**?**]) that under assumption of strong convexity, for the gradient projection method (inner cycle) with the step-size $\lambda_i = 1/L_{\nabla\phi}$ one has

$$|x^{k,i} - \bar{u}^k| \leq dq^{i/2},$$

where $\bar{u}^k$ is the optimal solution to (14) and $q = (1 - \gamma/L_{\nabla\phi})$. Local analysis of the objective function decrease at each iteration yields

$$\phi(x^{k,i+1}) \leq \phi(x^{k,i}) - (1/2L_{\nabla\phi})|v^{k,i} - x^{k,i}|^2.$$

Hence

$$\begin{aligned} (1/2L_{\nabla\phi})|v^{k,i} - x^{k,i}|^2 \quad &\leq \quad \phi(x^{k,i}) - \phi(\bar{u}^k) \\ &\leq \quad L_\phi|x^{k,i} - \bar{u}^k| \\ &\leq \quad L_\phi dq^{i/2} \end{aligned} \tag{15}$$

As in remark 6, let N be a maximum number of the outer iteration of method 2 such that $|u^k - x^k| > \delta/C$ for all $k \leq N$. This means that for $k \leq N$, the following bound

$$d|v^{k,i} - x^{k,i}| \leq \frac{\gamma\delta^2}{2C^2}$$

guarantees that the STOP in the inner cycle has already been reached. In turn, (15) guarantees that the latter inequality will be satisfied for the first i such that

$$(2L_{\nabla\phi}L_\phi d)^{1/2} q^{i/4} \leq \frac{\gamma\delta^2}{2C^2}.$$

This implies that for each $k \leq N$ the number the inner cycle iterations is bounded by $c_1 \log(c_2/\delta)$. The estimate for the number N is given by (13), therefore, the overall number of iterations is bounded by

$$O\left(\frac{\log(1/\delta)}{\delta^2}\right).$$

1.2 Method 3

In this section we describe another version of primal-dual constraint aggregation method. It differs from method 1 in that it uses an additional aggregate constraint $\langle p^c, Ax^c - b\rangle = 0$ rather than the term $\langle p^c, Ax^c - b\rangle$ at the objective function. A main motivation for considering such a version here is that in practice it outperforms notably method 1. Therefore, the results of numerical tests are presented for method 3. From theoretical point of view, method 3 provides somewhat better aposteriory bounds for the quality of the current solution approximation.

Method 3

Let (x^0, p^0) be fixed, $x^0 \in X$;
(1) given (x^k, p^k), $\gamma_k > 0$, find

$$u^k = \text{argmin}\{f(x) + (\gamma_k/2)|x - x^k|^2\}$$

subject to

$$\begin{aligned} & x \in X \\ \langle Ax^k - b,\ Ax - b\rangle &= 0, \\ \langle p^k,\ Ax - b\rangle &= 0; \end{aligned} \tag{16}$$

(2) choose α_k and β_k and set

$$x^{k+1} = x^k + \alpha_k(u^k - x^k),$$
$$p^{k+1} = p^k + \beta_k(Au^k - b);$$

Badly placed ()'s

(3) set $k = k + 1$, continue with (1).

Note that the feasible set of the subproblem in method 3 (throughout this subsection we shall denote it $\mathcal{X}_k$) is a relaxation of the original feasible domain $\mathcal{X}$.

Convergence results for method 3 are analogous to those of method 1. We assume that the auxiliary subproblem is solved precisely.

Theorem 3. *Choose γ_k, α_k, and β_k as in theorem 1. Assume that the auxiliary subproblem is solved precisely. Then, statements (i) and (ii) of theorem 1 carry over for the sequence of x^k generated by method 3.*

Proof. The argument basically follows the same line as for theorem 1. In view of (16), instead of equation (8) one has

$$\langle A^T p^*,\ u^k - x^* \rangle \quad = \quad (1/\beta_k)\langle p^* - p^k,\ p^{k+1} - p^k \rangle \tag{17}$$

and instead of (9)

$$\langle w,\ x^* - u^k \rangle - (\gamma_k/2)|u^k - x^k|^2 \geq \gamma_k \langle u^k - x^k,\ x^k - x^* \rangle. \tag{18}$$

Substitute (17) in (7). Then multiply (7) and (18) by β_k and add them up. As a result, one arrives at (10). The rest is the same as in the proof of theorem 1. Proof is complete.

Note that the remarks made after the proof of theorem 1 are completely relevant for method 3 as well. In particular, we would like to make the following observation regarding the stopping rule. First we obtain a lower bound for the directional derivative of the objective function in terms of $|u^c - x^c|$.

Optimality conditions in the subproblem in method 3 yield

$$\langle w,\ x - u^c \rangle \geq \gamma_c \langle u^c - x^c,\ u^c - x \rangle \geq -\gamma_c |u^c - x^c||u^c - x| \tag{19}$$

where w is some element from $\partial f(u^k)$. Divide both sides by $|u^c - x|$ and maximize (19) with respect to $w \in \partial f(u^c)$. Taking x to be an arbitary point from $\mathcal{X}$ one has

$$\max_{w \in \partial f(u^c)} \langle w,\ h \rangle \geq -\gamma_c |u^c - x^c|,$$

with $|h| = 1$. This means that a unit step in an arbitrary direction pointing towards the feasible domain could improve the objective function by no more than $\gamma_c |u^c - x^c|$. Obviously, in order to bound the optimality gap, set $x = x^*$ in (19):

$$\gamma_c d |u^c - x^c| \geq f(u^c) - f_*,$$

where d is diameter of X. The latter estimate appears to be more tight than analogous one in remark 4 since it does not include the constant L_A and an upper bound for the dual variables. Finally note that for residual of the system of constraints, the bound remains the same as in remark 4.

2 Numerical results

This section reviews the results of numerical tests of the primal-dual constraint aggregation method. As mentioned above, method 3 turns out to be superior to method 1 and therefore, we further focus on method 3 and the results to be found below pertain to this method.

Two important issues were given special care while conducting the tests: (1) choice of the aggregate constraints, their number and composition, (2) choice of the prox-parameter γ_k and the step-size multiplier α_k (subproblems were solved by QP-subroutine from the NAG library (see []) and it is quite safe to assume that μ_k and η_k are zero). Empirical observations with regard to (1) and (2) can be briefly summarized as follows. It seems better to choose parameter α_k as big as possible, in particular, in all tests it was set on its theoretical upper bound, i.e., twice as big as in the rule B in theorem 1 (see remark 2) except for one case reported below. Smaller values of the step-size did not improve behaviour of the method. Parameter γ_k should be kept as small as possible but big enough to provide that α_k is close to 1. In general, bigger values of γ_k tend to facilitate convergence in the residual and slow down convergence in the optimality gap. Various strategies can be devised in order to implement this general rule. In the tests reported, constant value in the range from 1 to 5 for γ_k was used throughout the iteration process.

The most cruicial question is the choice of aggregate constraints. To a great extent this depends on the structure of the problem in question. The multistage stochastic programming problems provide natural clustering schemes since one may consider various scenario bundles for aggregation. Below we give the results of tests with one such scheme. However, presently, a more satisfactory answer to this problem is largely a subject for future research.

In the following subsection we describe the problem used for the tests. Then we present the results of tests of two modifications of method 3 on two instances of the given problem of smaller and bigger dimension. The algorithms differ from each other by the constraint aggregation scheme. The first one uses the simplest scheme specified in the method 3, i.e. only two aggregates, while the second one uses a scenario bundling scheme to be specified below in more details.

2.1 Test problem description

We consider a dynamic portfolio selection problem. Over a given time horizon $0, 1, \ldots, T$, a decision maker maintaines a portfolio of assets so that to maximaize his terminal risk-return utility function. Risk is associated with future uncertainty of return on assets. We represent uncertainty as a random process

$$\mathbf{w}_T = (\omega_1, \dots, \omega_T) \in \mathbf{\Omega} = \Omega \times \dots \times \Omega,$$

where the outcomes $\mathbf{w}_T$ constitute the set of possible random events in the model. Also for each $0 < t \leq T$ denote

$$\mathbf{w}_t = (\omega_1, \dots, \omega_t).$$

At time t the history of the process $\mathbf{w}_t$ is already known to the decision-maker, thus he may rebalance the portfolio in order to reflect this information. Therefore, with each time t, possible realization of $\mathbf{w}_t$, and asset type $i = 1, \dots, a$ we associate a corresponding decision variable $x_t^i(\mathbf{w}_t) \geq 0$ standing for the part of his current wealth allocated for the asset i. Thus, for each time $t = 0, \dots, T-2$ and event $\mathbf{w}_t$, the budgetary constraint is expressed as follows

$$\sum_{i=1}^{a} x_{t+1}^i(\mathbf{w}_{t+1}) - \sum_{i=1}^{a} (1 + r_{t+1}^i(\mathbf{w}_{t+1})) x_t^i(\mathbf{w}_t) = 0,$$

for all $\omega_{t+1} \in \Omega$, where $r_t^i(\mathbf{w}_t)$ is the i-th asset return in time t in event $\mathbf{w}_t$. The initial portfolio allocation is given by the constraint

$$\sum_{i=1}^{a} x_0^i = 1.$$

Let $\pi_{\mathbf{w}_T}$ denote probability of each event $\mathbf{w}_T$. The objective of the decision maker in this example is to maximize

$$f(x) = E - \kappa \sum_{\mathbf{w}_T \in \mathbf{\Omega}} \pi_{\mathbf{w}_T} \left(\sum_{i=1}^{a} (1 + r_T^i(\mathbf{w}_T)) x_{T-1}^i(\mathbf{w}_{T-1}) - E \right)^2,$$

where E is the expected return

$$E = \sum_{\mathbf{w}_T \in \mathbf{\Omega}} \pi_{\mathbf{w}_T} \sum_{i=1}^{a} (1 + r_T^i(\mathbf{w}_T)) x_{T-1}^i(\mathbf{w}_{T-1})$$

and $\kappa > 0$ is the risk-aversion parameter.

In the tests Ω consisted of three random outcomes $\{1, 0, -1\}$, thus, the process as a whole can be represented as a trinomial brunching structure. The number of assets $a = 4$. Asset returns were generated according to the following relation

$$1 + r_t^i(\mathbf{w}_t) = e_i + \sigma_i(t)\omega_t, \ \omega_t \in \Omega,$$

with e_i being average return on asset i and $\sigma_i(t)$ being volatility of return. We have chosen $\sigma_i(t)$ to be slightly decresing over time. We considered two problem instances (P4) and (P5) with $T = 4$ and $T = 5$ having the number of scenario paths 81 and 243 respectively. In the method which we further refer to as method 3.1, all budgetary constraints were treated as $Ax = b$ in terms of defenition of problem (P), and nonnegativity constraints were treated as X. In method 3.1 we have used only two aggregate constraints as in definition of method 3. In the method 3.2 only those budgetory constraints were considered as $Ax = b$ that correspond to time $T - 2$ and the rest were considered as X. The aggregates were constructed by merging the "subtrees" emanating from the "nodes" $\mathbf{w}_{T-2}$. As a result, total number of linear equality constraints in the relaxed subproblem was nearly half as much as in the original problem.

2.2 Results of tests

Here we give the problem parameters and the numerical data obtained. Problem (P4) has in total 161 variables and 41 constraints (nonnegativity constraints on variables are not counted). The optimal value $f_* = 1.248$. The aggregate subproblem in method 3.1 has 2 constraints and in method 3.2 it has 24 constraints. The initial approximations for (x, p) were chosen either all zero or all ones and this does not seem to affect strongly behaviour of the methods. The following tables summurize the data. For convinience we take negative of the objective function and consider an equivalent minimization problem. Thus, the values in column $f(u^k) - f_*$ are positive if the corresponding value of the utility function lies below the maximum and negative otherwice.

Table 1. Problem (P4), method 3.1, $\gamma_k \equiv 5$.

k	$\lvert Ax^k - b\rvert$	$\lvert u^k - x^k\rvert$	$f(u^k) - f_*$
0	4.462	2.811	-0.857
100	0.160	0.041	0.064
200	0.024	0.009	0.056
300	0.005	0.004	0.051
400	0.001	0.003	0.045
500	0.002	0.003	0.040

Table 2. Problem (P4), method 3.2, $\gamma_k \equiv 1$.

k	$\lvert Ax^k - b\rvert$	$\lvert u^k - x^k\rvert$	$f(u^k) - f_*$
0	4.462	9.154	0.041
100	0.013	0.022	0.050
200	0.009	0.016	0.037
300	0.007	0.012	0.030
400	0.006	0.011	0.025
500	0.006	0.010	0.021

Note that the step $|u^k - x^k|$ turns out to be reasonably good approximation for the optimality gap $f(u^k) - f_*$. For bigger values of γ_k, assymptotically the step tends to underestimate $f(u^k) - f_*$, and as one can see from the data, the gap is in most cases about 2-10 times the step. Also, the greater γ_k, the less accurate is the approximation of the gap via $|u^k - x^k|$. However, bigger values of γ_k result in a better convergence in the residual.

Table 3. Problem (P4), method 3.2, $\gamma_k \equiv 5$.

k	$\|Ax^k - b\|$	$\|u^k - x^k\|$	$f(u^k) - f_*$
0	4.462	9.109	0.853
100	9.6e-5	0.005	0.061
200	6.8e-5	0.003	0.052
300	6.1e-5	0.003	0.046
400	5.2e-5	0.003	0.040
500	3.1e-5	0.002	0.036

Similar observation can be made for the bigger problem as well. Problem (P5) has in total 485 variables and 122 constraints. The optimal value $f_* = 1.318$. The aggregate subproblem in method 3.2 has 69 constraints. Since method 3.2 outperforms method 3.1 we give the results for method 3.2 only.

Table 4. Problem (P5), method 3.2, $\gamma_k \equiv 1$.

k	$\|Ax^k - b\|$	$\|u^k - x^k\|$	$f(u^k) - f_*$
0	4.720	15.594	0.060
100	0.007	0.015	0.066
200	0.005	0.011	0.055
300	0.004	0.010	0.048
400	0.004	0.009	0.042
500	0.003	0.008	0.037

Table 5. Problem (P5), method 3.2, $\gamma_k \equiv 5$.

k	$\|Ax^k - b\|$	$\|u^k - x^k\|$	$f(u^k) - f_*$
0	4.720	15.567	0.911
100	0.005	0.004	0.075
200	4.7e-5	0.003	0.070
300	3.1e-5	0.002	0.065
400	4.1e-5	0.002	0.062
500	1.7e-5	0.002	0.059

Additionally in Table 6 we provide the results of a heuristic experiment with method 3.2. As mentioned in the beginning of this section, primal-dual constraint aggregation method tends to perform better when the step-size α_k

is close to 1. The results in table 4 are produced by method 3.2 with $\alpha_k \equiv 1$. At the same time, the prevoius results show that convergence in the optimality gap suffers from big γ_k. In this test we have set $\gamma_k \equiv 0.1$. Despite the unit step is not substantiated by the theory, after several iterations method tends to stabilize and provide a good convergence. Similar effect was also observed with problem (P4).

Table 6. Problem (P5), method 3.2, $\gamma_k \equiv 0.1$.

k	$\|Ax^k - b\|$	$\|u^k - x^k\|$	$f(u^k) - f_*$
0	4.720	15.667	-0.406
100	0.428	0.256	0.016
200	0.030	0.025	0.008
300	0.002	0.015	0.005
400	5.1e-4	0.012	0.003
500	0.001	0.009	0.002

The results given in table 6 suggest that a closer investigation of the strategies for choosing the step-size and the prox-parameter should be taken.

Graphical illustration of convergence of the methods is given in figures 1 and 2. Figure 1 compares behaviour of methods 3.1 and 3.2 with $\gamma_k \equiv 5$ on problem (P4). The graph shows the trajectory of the step $|u^k - x^k|$ for both methods: black thiner line is for the method 3.1 and gray thicker one is for method 3.2.

Figure 2 displays two trajectories of the steps $|u^k - x^k|$ generated by method 3.2 with $\gamma_k \equiv 5$ for problem (P4) (black thiner line) and for problem (P5) (gray thicker line). The figure shows that the behaviour is much the same despite the problems are of different dimensions.

3 Conclusions

In the paper a basic convergence analysis of the primal-dual constraint aggregation method was developed and an application to a multistage stochastic programming was considered. The method suggests an approach for incorporating a constraint reduction scheme within an iterative process that converges to an optimal solution of the original problem. The error bounds showing the quality of the solution approximation are provided. Complexity analysis demonstrates that the method achieves an unimproveable by order of δ theoretical complexity bound $O(\delta^{-2})$ which is independent on the space dimension. Preliminary computational experiments show that the behaviour of the method strongly depends on the choice of the aggregates and the parameters of the method such as step-size and prox multipliers.

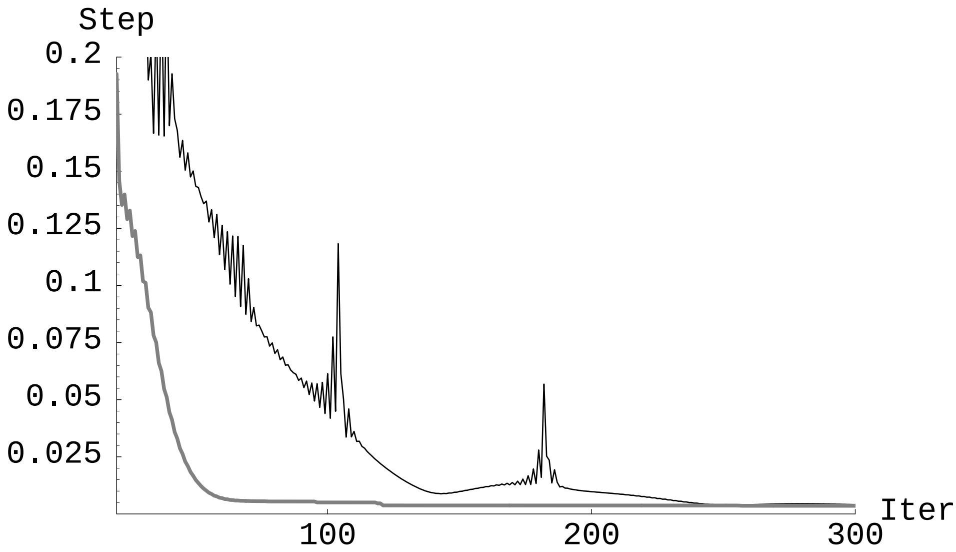

Figure 1.

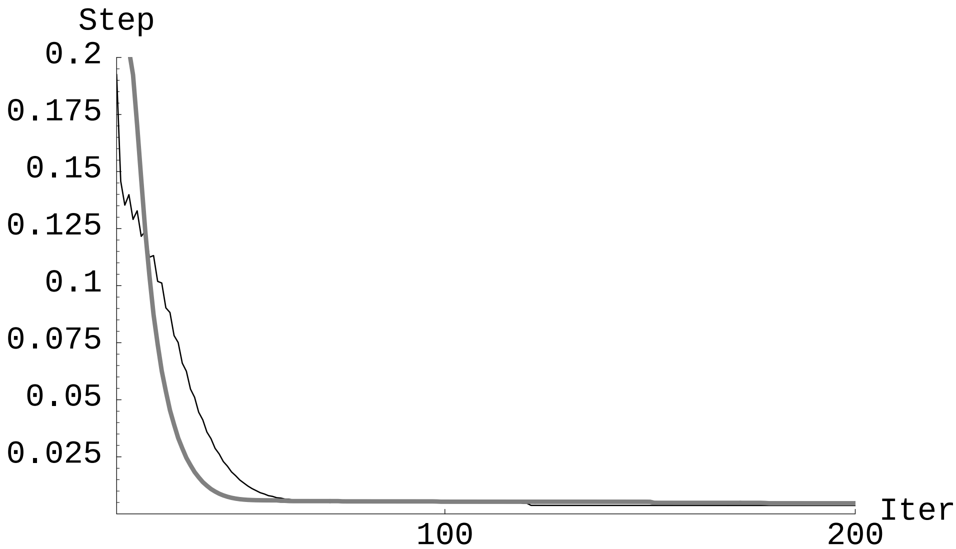

Figure 2.

Computational effort required by the method is dependent on whether a smaller number of aggregated constraints correctly represent major structural properties of the original set and whether it is at all possible. This seems to be particularly relevant for multistage stochastic problems where big dimension results from a big number of scenarios, while certain different scenarios may carry only slightly different information. Therefore, existence of efficient aggregation schemes seems possible. In view of that, an important problem for future research is development of adaptive dynamic strategy for updating iteratively the list and the composition of the aggregate consraints as well as the parameters of the method.