A dual strategy for the implementation of the aggregation principle in decision making under uncertainty

R.T. Rockafellar

Departments of Mathematics and Applied Mathematics

University of Washington

Roger J-B Wets

Department of Mathematics & Institute of Theoretical Dynamics

University of California, Davis

Abstract. A solution method for stochastic programs is proposed on the basis of the aggregation principle, which allows one to find the solution of a stochastic program by aggregating the solutions of individual deterministic scenario problems. The method concentrates on finding good estimates of the dual variables associated with the nonanticipativity constraints.

Keywords: decomposition, stochastic programming, aggregation principle, approximations

Date: Revised, November, 1991

Although multistage decision problems will be treated in the later part of this paper, the ideas will first be developed in the simplest case. Consider the 1-stage stochastic optimization problem

$$(\mathcal{P}) \qquad \text{minimize } F(x) := \sum_{s\in S} p_s f(x,s) \text{ over all } x \in \mathbb{R}^n,$$

where S is the set of possible realizations, i.e. scenarios, and is finite, p_s is the probability attached to scenario s, and $f : \mathbb{R}^n \times S \to \overline{\mathbb{R}}$ is a cost function with the (usual) interpretation that $f(x,s) = \infty$ means x is not a feasible solution under scenario s. We shall initially be concerned with a method for finding an optimal solution to this reduced problem. The method is based on the aggregation of solutions of individual scenario problems of the form

$$\text{minimize } f^a(x,s) \text{ over all } x \in \mathbb{R}^n \text{ for fixed } s \in S,$$

where f^a is a perturbed version of f. The basic difference with the progressive hedging algorithm [**7, 9**], which relies on a similar aggregation, is that the accent is placed on getting better estimates of the dual variables associated with the nonanticipativity constraints.

1. Approximation Scheme

Let $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ and suppose that $h \not\equiv \infty$ and $h \geq -\alpha_1|\cdot - x_0|^2 + \alpha_0$ for some $\alpha_1 \geq 0$, $x_0 \in \mathbb{R}^n$ and $\alpha_0 \in \mathbb{R}$ i.e., h is quadratically bounded from below. For $\lambda > 0$ let

$$h_\lambda(x) = \inf_u \big\{h(u) + \frac{1}{2\lambda}\, |x-u|^2\big\}, \quad x \in \mathbb{R}^n.$$

The function h_λ is the *Moreau-approximate (of index λ)*. For λ sufficiently small, h_λ is a finite-valued function [**1**]. In general it is subsmooth, and when h is convex, h_λ is convex and differentiable [**1**]. Moreover, as $\lambda \to 0$ with $\lambda > 0$ not only does h_λ converge pointwise to h, i.e.,

$$h_\lambda(x) \to h(x) \text{ as } \lambda \to 0 \text{ for all } \quad x \in \mathbb{R}^n,$$

it also h_λ epi-converges to h, i.e., we also have that for all $x \in \mathbb{R}$

$$\liminf_{\lambda\to 0} h_\lambda(x^\lambda) \geq h(x) \text{ for all } x^\lambda \to x.$$

In particular, epi-convergence means that if

$$x^k \in \operatorname{argmin} h_{\lambda_k} \text{ for } \quad k = 1, 2, \ldots$$

for some sequence $\lambda_k \to 0$ then every cluster point of the sequence $\{x^k\}_{k=1}^{\infty}$ minimizes h [**1**], i.e.,

$$\{\text{ cluster points of } \{x^k\}\} \subset \operatorname{argmin} h.$$

Returning to the stochastic optimization problem $(\mathcal{P})$, we approximate it by the following class of problems with parameter $\lambda > 0$:

$$(P_\lambda) \qquad \text{minimize } F_\lambda(x) := \sum_{s\in S} p_s f_\lambda(x,s) \text{ over all } x \in \mathbb{R}^n,$$

where the function

$$f_\lambda(x,s) = \inf_u \left\{ f(u,s) + \frac{1}{2\lambda}|x-u|^2 \right\}$$

is the Moreau-approximate to f with respect to the x-variable. All along, we are going to assume that the functions $f(\cdot\,,s)$ for $s \in S$ are minorized by the same quadratic form: $\alpha_0|\cdot - x_0| + \alpha_1$ with $\alpha_0 \geq 0$, $x_0 \in \mathbb{R}^n$, $\alpha_1 \in \mathbb{R}$. In practice this is not a significant restriction. Note that the objective function F_λ in $(\mathcal{P}_\lambda)$ is *not* the Moreau-approximate (with respect to the x-variable) of the (essential) objective function F in the stochastic optimization problem $(\mathcal{P})$. But we still have

$$F_\lambda(x) \to F(x) \quad \text{for all } x \text{ as } \lambda \to 0,$$

i.e., the approximations converge pointwise, and further that whenever $x^\lambda \to x$ (as λ tends to 0), then

$$\liminf_{\lambda\to 0} F_\lambda(x) \geq F(x),$$

as can readily be verified. This means that the (essential) objective functions for the approximating problems $(\mathcal{P}_\lambda)$ *epi-converge* to that of the given stochastic optimization problem $(\mathcal{P})$. In turn, this implies that the optimal solutions to the approximating problems can be accepted as approximate solutions to the given problem. Thus, if $x^k \in \operatorname{argmin}_x F_{\lambda_k}(x)$ for some parameter sequence $\lambda_k \to 0$ (with $\lambda_k > 0$), and if $x^k \to \bar{x}$, then $\bar{x} \in \operatorname{argmin}_x F(x)$.

The overall plan is to solve the approximating problems $(\mathcal{P}_\lambda)$ and let $\lambda \to 0$. In this short note we shall be exclusively interested in finding a solution to (P_λ) for some $\lambda > 0$ sufficiently close to 0.

In the case of $(\mathcal{P}_\lambda)$, we are dealing with an optimization problem whose objective function is finite-valued. Moreover, if the functions $f(\cdot\,,s)$ are convex, the functions $f_\lambda(\cdot\,,s)$ and F_λ are convex *differentiable* functions. In what follows, we are going to deal only with the convex case, because it allows for a full exploitation of duality.

2. A Duality Result

To come up with an appropriate dualization for the approximate problem, we consider perturbations of the basic *"constraint"* requiring that $x \in \mathbb{R}^n$ be chosen in advance of the scenario realization to occur, i.e. as a constant function of the variable $s \in S$. The problem we get is

$$(\mathcal{D}_\lambda), \qquad \text{maximize } G_\lambda(w) := -\sum_{s\in S} p_s f_\lambda^*(w_s, s) \text{ subject to } \sum_{s\in S} p_s w_s = 0,$$

where w denotes a family $\{w_s\}_{s\in S}$ of vectors in $\mathbb{R}^n$. We show that this is dual to problem $(\mathcal{P}_\lambda)$ in the sense that $\inf(\mathcal{P}_\lambda) \geq \sup(\mathcal{D}_\lambda)$, and if both problems admit optimal solutions equality holds, i.e., $\inf(\mathcal{P}_\lambda) = \sup(\mathcal{D}_\lambda)$. Here f_λ^* is the conjugate function of f_λ with respect to the first variable (under Legendre-Fenchel transform of convex analysis),

$$f_\lambda^*(w, s) := (f_\lambda)^*(w, s) = \sup_{x\in\mathbb{R}^n} \big\{wx - f_\lambda(x, s)\big\}.$$

Assuming that both $(\mathcal{P}_\lambda)$ and $(\mathcal{D}_\lambda)$ are feasible and that $\bar{x}$ is a feasible solution to $(\mathcal{P}_\lambda)$ and $\bar{w} = \{\bar{w}_s\}_{s\in S}$ is a feasible solution to $(\mathcal{D}_\lambda)$, we have

$$f_\lambda^*(\bar{w}_s, s) \geq \bar{w}_s \bar{x} - f_\lambda(\bar{x}, s) \quad \text{for all } s \in S$$

and consequently

$$-\sum_{s\in S} p_s f_\lambda^*(\bar{w}_s, s) + (\sum_{s\in S} p_s \bar{w}_s)\bar{x} \leq \sum_{s\in S} p_s f_\lambda(\bar{x}, s).$$

This translates to $G_\lambda(\bar{w}) \leq F_\lambda(\bar{x})$, because $\sum_{s\in S} p_s \bar{w}_s = 0$ by the feasibility of $\bar{w}$ in $(\mathcal{D}_\lambda)$. Nothing more is needed to conclude that $(\inf P_\lambda) \geq (\sup D_\lambda)$ since if either problem is infeasible the inequality is satisfied trivially.

We shall argue next that if both problems admit optimal solutions, then $\inf(\mathcal{P}_\lambda) = \sup(\mathcal{D}_\lambda)$. In doing so, however, we are going to proceed in a way that starts only by assuming that $(\mathcal{P}_\lambda)$ is solvable.

Let $\bar{x}$ be an optimal solution to $(\mathcal{P}_\lambda)$. Because the functions $f_\lambda(\cdot, s)$ are differentiable, so is the function $F_\lambda = \sum_{s\in S} p_s f_\lambda(\cdot, s)$, and thus

$$0 = \nabla F_\lambda(\bar{x}) = \sum_{s\in S} p_s \nabla f_\lambda(\bar{x}, s).$$

Now, let

$$\bar{w}_s := \nabla f_\lambda(\bar{x}, s) \text{ for each } s \in S.$$

Clearly, $w := \{\bar{w}_s\}_{s\in S}$ is a feasible solution of the dual problem $(\mathcal{D}_\lambda)$, since $\sum_{s\in S} p_s \bar{w}_s = 0$. Moreover, from the definition of the conjugate function, along with the gradient inequality for convex functions, which we express here as

$$f_\lambda(x,s) - f_\lambda(\bar{x},s) \geq \bar{w}_s(x-\bar{x}) \quad \text{for all } x \in \mathbb{R}^n,$$

we have $-f^*_\lambda(\bar{w}_s, s) = f_\lambda(\bar{x}, s) - \bar{w}_s\bar{x}$, and hence

$$-\sum_{s\in S} p_s f^*_\lambda(\bar{w}_s, s) = \sum_{s\in S} p_s f_\lambda(\bar{x}, s), \text{ i.e., } G_\lambda(\bar{w}) = F_\lambda(\bar{x}).$$

This completes the proof and shows at the same time that when $(\mathcal{P}_\lambda)$ is solvable, so is $(\mathcal{D}_\lambda)$.

Before we pass on to the description of a solution procedure for $(\mathcal{D}_\lambda)$, let us note that for any function $h : \mathbb{R}^n \to \overline{\mathbb{R}}$ and any $\lambda > 0$, the Moreau-approximate h_λ defined in Section 1 has as its conjugate the function

$$h^*_\lambda = h^* + (\frac{1}{2\lambda}|\cdot|^2)^* = h^* + \frac{\lambda}{2}|\cdot|^2.$$

Thus in general, the functions $f^*_\lambda(\cdot, s)$ appearing in the dual problem $(\mathcal{D}_\lambda)$ are *not* necessarily finite-valued. We also observe that $h^*_\lambda \neq (h^*)_\lambda$.

3. Solving the Dual Problem

The vectors w_s in the dual problem $(\mathcal{D}_\lambda)$ can be interpreted as the marginal price vectors associated with the nonanticipativity constraint that the decision x cannot depend on foreknowledge of the scenario s. This has been elaborated in a series of papers [**3, 5, 6, 8**] and is now a well-understood theory. What is novel in our approach is that we plan to solve the stochastic optimization problem by concentrating on finding "good" solutions to the dual problem.

To obtain such solutions we rely on a technique introduced by Wolfe and Dantzig, called generalized linear programming [**2**]. In our setting the method takes on the following form. Let

$$\widehat{w}^k = \{\widehat{w}^k_s\}_{s\in S} \text{ for } k = 0, \ldots, \nu - 1$$

be a collection of dual elements (typically generated during the first ν steps of the procedure) such that the system of linear relations

$$\sum_{k=0}^{\nu-1} \beta_k (\sum_{s\in S} p_s \widehat{w}^k_s) = 0, \qquad \sum_{k=0}^{\nu-1} \beta_k = 1, \qquad \beta_k \geq 0,$$

can be solved for the coefficients β_k. (Initially, one could take $\widehat{w}_s^0 = 0$ for all s, for instance.) For $k = 0, \ldots, \nu - 1$ define

$$\bar{w}^k = \sum_{s \in S} p_s \widehat{w}_s^k \ \text{ and } \ \alpha_k = \sum_{s \in S} p_s f_\lambda^*(\widehat{w}_s^k, s);$$

we shall return later to the calculation of these quantities.

We consider the following linear programming problem in the variables β_k:

$$\begin{aligned} \text{minimize } & \sum_{k=0}^{\nu-1} \beta_k \alpha_k \\ \text{subject to } & \sum_{k=0}^{\nu-1} \beta_k \bar{w}^k = 0 \\ & \sum_{k=0}^{\nu-1} \beta_k = 1 \\ & \beta_k \geq 0, \quad k = 0, \ldots, \nu - 1. \end{aligned}$$

With the condition that has been imposed on the choice of the vectors $(\widehat{w}_s^k, s \in S)$ we know that this problem is feasible and bounded. Thus there exist multipliers (z^ν, θ^ν) such that

1. $\alpha_k - z^\nu \bar{w}^k - \theta^\nu \geq 0$ for $k = 0, \ldots, \nu - 1$
2. $\theta^\nu =$ optimal value of the problem.

If we merely had to find the best solution to the dual problem when restricted to all elements $w = \{w_s\}_{s \in S}$ that can be obtained as a convex combination of the vectors $\widehat{w}_s^k$ for $s \in S$, $k = 0, \ldots, \nu - 1$, we would simply use as weights the values β_k obtained as an optimal solution to this linear program. The difficulty with accepting this convex combination as the overall solution to $(\mathcal{D}_\lambda)$, however, is that the current collection of vectors $\widehat{w}_s^k$ may not be rich enough.

If a new element $\widehat{w}^\nu = \{\widehat{w}_s^\nu\}_{s \in S}$ is going to be introduced the collection it should be done in such a way that the resulting linear program (which would involve one additional column) will yield an improved solution. Thus we should choose $\widehat{w}^\nu$ such that

$$\alpha_\nu - z^\nu \bar{w}^\nu - \theta^\nu < 0,$$

where $\bar{w}^\nu = \sum_{s \in S} p_s \widehat{w}_s^\nu$ and $\alpha_\nu = \sum_{s \in S} p_s f_\lambda^*(\widehat{w}_s^\nu, s)$. Equivalently, we should find vectors $\widehat{w}_s^\nu$ for $s \in S$ such that

$$\sum_{s \in S} p_s (f_\lambda^*(\widehat{w}_s^\nu, s) - z^\nu \widehat{w}_s^\nu) < \theta^\nu.$$

Such vectors exist, provided that

$$\sum_{s\in S} p_s \inf_v \{f^*_\lambda(v,s) - z^\nu v\} < \theta^\nu.$$

Thus the search for $\widehat{w}^\nu = \{\widehat{w}^\nu_s\}_{s\in S}$ boils down to solving for each $s\in S$ the optimization subproblem

$$\text{minimize } \ f^*_\lambda(v,s) - z^\nu v, \ \ v\in \mathbb{R}^n.$$

We are going to show next that the calculation of an optimal solution to this problem along with the corresponding optimal value does not require a direct formula for the function f^*_λ. The optimal solution, once obtained, is the vector we choose for $\widehat{w}^\nu_s$. If the dual element $\widehat{w}^\nu = \{\widehat{w}^\nu_s\}_{s\in S}$ which is so obtained satisfies the desired condition, a new column is introduced in the linear programming problem, after computing α_ν and $\bar{w}^\nu$. We stop if $\widehat{w}^\nu$ satisfies

$$\alpha_\nu - z^\nu \bar{w}^\nu \geq \theta^\nu.$$

In the latter case, with $\beta^o = (\beta^o_0,\dots,\beta^o_{\nu-1})$ as the optimal solution of the linear program, the vectors

$$w^o_s = \sum_{k=0}^{\nu-1} \beta^o_k \ \widehat{w}^k_s$$

furnish an optimal solution to $(\mathcal{D}_\lambda)$ while z^ν is an optimal solution to $(\mathcal{P}_\lambda)$; this follows immediately from the relationship between $(\mathcal{P}_\lambda)$ and $(\mathcal{D}_\lambda)$.

It remains only to demonstrate that the construction of a solution to the subproblem

$$\text{minimize } \ f^*_\lambda(v,s) - z^\nu v \ \text{ over all } \ v\in\mathbb{R}^n$$

does not require an auxiliary formula for $f^*_\lambda(\cdot,s)$. Indeed, one has

$$\begin{aligned}
\inf_y \{f^*_\lambda(y,s) - z^\nu y\} &= \inf_y \sup_x \{(x-z^\nu)y - f_\lambda(x,s)\} \\
&= \inf_y \sup_x \sup_u \{(x-z^\nu)y - f(u,s) - \frac{1}{2\lambda}|x-u|^2\} \\
&= \inf_y \sup_u \ \big\{ \sup_x \{(x-z^\nu)y - \frac{1}{2\lambda}|x-u|^2\} - f(u,s)\big\} \\
&= \inf_y \sup_u \ \big\{(\lambda y + u - z^\nu)y - \frac{1}{2\lambda}|\lambda y + u - u|^2 - f(u,s)\big\} \\
&= \inf_y \sup_u \ \big\{(\lambda/2)|y|^2 + (u-z^\nu)y - f(u,s)\big\} \\
&= \sup_u \inf_y \ \big\{(\lambda/2)|y|^2 + (u-z^\nu)y - f(u,s)\big\},
\end{aligned}$$

where for the last equality we have used a minimax theorem [**4,** theorem 37.3]; we note that the function in the final braces is concave (and lsc) in u, and that it is convex and inf-compact in y. From this we get

$$\inf_y \{f^*_\lambda(y,s) - z^\nu y\} = \sup_u \big\{\frac{\lambda}{2}|\lambda^{-1}(z^\nu - u)|^2 + (u - z^\nu)\lambda^{-1}(z^\nu - u) - f(u,s)\big\}$$
$$= -\inf_u \big\{f(u,s) + \frac{1}{2\lambda}|z^\nu - u|^2\big\}.$$

Therefore, with

$$u^\nu_s \in \operatorname*{argmin}_u \big\{f(u,s) + \frac{1}{2\lambda}|z^\nu - u|^2\big\},$$

we set

$$\widehat{w}^\nu_s = \lambda^{-1}(z^\nu - u^\nu_s),$$

and

$$\alpha^\nu_s = z^\nu \widehat{w}^\nu_s - f(z^\nu - \lambda\widehat{w}^\nu_s, s) - \frac{\lambda}{2}|\widehat{w}^\nu_s|^2.$$

4. Summary of the Algorithm

We now review the operations involved in the procedure, skipping the justification that passes through $f^*_\lambda(\cdot, s)$.

Step 0. Initialize by setting $\nu = 1$ picking $\{\widehat{w}^0_s\}_{s\in S}$ in such a way that the vector $\bar{w}^0 := \sum p_s \widehat{w}^0_s$ is 0, and $\alpha_0 = \sum_{s\in S} p_s[\lambda/2|\widehat{w}^0_s|^2 + \sup_u(\widehat{w}^0_s u - f(u,s))]$.

Step 1. Find an optimal solution to the linear programming problem

$$\begin{aligned} \min_\beta \quad & \sum_{k=0}^{\nu-1} \beta_k \alpha_k \\ \text{subject to} \quad & \sum_{k=0}^{\nu-1} \beta_k \bar{w}^k = 0, \\ & \sum_{k=0}^{\nu-1} \beta_k = 1, \\ & \beta_k \ge 0,\ k = 0, \ldots, \nu - 1. \end{aligned}$$

Let $(z^\nu, \theta^\nu) \in \mathbb{R}^{n+1}$ be the multipliers obtained as the solution to the corresponding dual linear programming problem (namely, minimize $-\theta$ subject to $\theta \ge \bar{w}^k z - \alpha_k$ for $k = 0, \ldots, \nu - 1$).

Step 2. For each $s \in S$ let

$$u_s^\nu \in \operatorname*{argmin}_u \left\{ f(u,s) + \frac{1}{2\lambda}|z^\nu - u|^2 \right\}$$
$$\widehat{w}_s^\nu = \lambda^{-1}(z^\nu - u_s^\nu), \quad \bar{w}^\nu = \lambda^{-1} \sum_{s\in S} p_s (z^\nu - u_s^\nu),$$
$$\alpha_s^\nu = z^\nu \widehat{w}_s^\nu - f(z^\nu - \lambda \widehat{w}_s^\nu, s) - \frac{\lambda}{2}|\widehat{w}_s^\nu|^2, \quad \alpha_\nu = \sum_{s\in S} p_s \alpha_s^\nu.$$

Step 3. If $\alpha_\nu < \bar{w}^\nu z^\nu + \theta^\nu$ return to Step 1 with the counter ν increased by 1.
If $\alpha_\nu \geq \bar{w}^\nu z^\nu + \theta^\nu$ the problem is solved; z^ν is the optimal solution.

In Step 3, if λ is not small enough, and $\bar{w}^\nu z^\nu + \theta^\nu \leq \alpha_\nu$ one may consider adjusting λ, recalculating the coefficients $\alpha_k, k = 0, \dots, \nu - 1$ and restarting at Step 1. In view of the formula for the f_λ^* (at the end of section 2), only minor calculations will be necessary to adjust the coefficients of the linear program in Step 1 when passing from λ to a new $\lambda' > \lambda$. Indeed

$$\begin{aligned} \alpha_s^k(\lambda') - \alpha_s^k &= f_{\lambda'}^*(\widehat{w}_s^\nu, s) - f_\lambda^*(\widehat{w}_s^\nu, s) \\ &= f^*(\widehat{w}_s^\nu, s) + \frac{\lambda'}{2}|\widehat{w}_s^\nu|^2 - f^*(\widehat{w}_s^\nu, s) - \frac{\lambda}{2}|\widehat{w}_s^\nu|^2 \\ &= \frac{\lambda' - \lambda}{2}|\widehat{w}_s^\nu|^2. \end{aligned}$$

If the strategy is to work first with a large value for λ, to be decreased later on, Step 3 would have to consider readjusting λ on the basis of the difference between α_ν and $\bar{w}^\nu z^\nu + \theta^\nu$.

This approach can also be viewed as a "cutting" plane method. To see this simply note that the rows of the dual of the linear program to be solved in Step 1 determine affine functions majorized by F_λ.

5. The Multistage Case

Let $x = (x_1, \ldots, x_T)$ with $x_t \in \mathrm{I\!R}^{n_t}$ correspond to the sequence of decisions to be made in stage t with $t = 1, \ldots, T$. Denote by

$$\vec{x}^{\,t} = (x_1, \ldots, x_t) \quad \text{for } t \leq T$$

the subvector that corresponds to the decisions to be made up to time t. Similarly, let $s = (s_1, \ldots, s_T)$ denote a scenario with component $s_t \in \mathrm{I\!R}^{N_t}$ symbolizing the events associated with the t-th stage, i.e., s_t is observed after x_t is selected. Let

$$\vec{s}^{\,t} = (s_1, \ldots, s_t) \quad \text{for } t \leq T.$$

Because a decision at state t can only depend on the information available, x_t can only depend on $\vec{s}^{\,t-1}$ and not on future events, not even those revealed during period t. We shall express this in the following fashion: we consider x_t as function of s, i.e.,

$$x_t(\cdot) : S \to \mathrm{I\!R}^{n_t}$$

and demand that

$$x_t(s) = (E^{t-1}x_t)(s) := E\{x_t | \;\vec{s}^{\,t-1}\} \text{ for all } s \in S$$

where $E\{\cdot \,|\, \cdot\}$ denotes conditional expectation.

Let J be the operator that takes a decision mapping $x : S \times \ldots \times S \to \mathrm{I\!R}^n$ aggregates it into a decision mapping that satisfies the preceding condition of nonanticipativity:

$$Jx = (E^0 x_1, E^1 x_2, \ldots, E^{t-1} x_t, \ldots, E^{T-1} x_T)$$

with $E^0 = E$ denoting the usual expectation. We can formulate the multistage stochastic programming problem as follows:

$$\begin{aligned} \text{minimize } & \sum_{s \in S} p_s f(x(s), s), \\ \text{subject to } & x_t(\cdot) = E\{x_t \,|\; \vec{s}^{\,t-1}\}(\cdot), \; t = 1, \ldots, T, \end{aligned}$$

or equivalently

$$\begin{aligned} \text{minimize } & \sum_{s \in S} p_s f(x(s), s), \\ \text{subject to } & x(\cdot) = Jx(\cdot). \end{aligned}$$

We note that J is a projection operator; in particular $J^2x = Jx$.

From this point on, the approach is similar to the one we followed for the 1-stage model. The given problem is replaced by an approximate one involving the parameter value $\lambda > 0$,

$$\begin{aligned} &\text{minimize } \sum_{s\in S} p_s f_\lambda(x(s), s) \\ &\text{subject to } x = Jx. \end{aligned}$$

The dual of this problem is then shown to be

$$\begin{aligned} &\text{maximize } \ -\sum_{s\in S} p_s f_\lambda^*(w(s), s) \\ &\text{subject to } \ Jw = 0; \end{aligned}$$

the constraint $Jw = 0$ is equivalent to $E^{t-1}w_t = 0$ for $t = 1, \ldots, T$. This problem is now solved by the method of generalized linear programming as for the 1-stage model, the only change being in the structure of the linear program that occurs in Step 1.

6. Algorithm for the Multistage Problem

The justification given for the 1-stage case still applies, but we shall later provide some details in connection with Step 2. A value of $\lambda > 0$ must be selected in advance; here we are not discussing the iterative adjustment of this parameter, the comments at the end of section 4 remain valid.

Step 0. Initialize by setting $\nu = 1$ and picking $\{\widehat{w}_s^0\}_{s\in S}$ such that $J\widehat{w}^0 = 0$. Calculate $\bar{w}^0(s) = (J\widehat{w}^0)(s)$ for each $s \in S$, and set $\alpha_0 = \sum_{s\in S} p_s[\lambda/2|\widehat{w}_s^0|^2 + \sup_u(\widehat{w}_s^0 u - f(u, s))]$.

Step 1. Find an optimal solution to the linear programming problem

$$\begin{aligned} &\text{minimize } \sum_{k=0}^{\nu-1} \beta_k \alpha_k \\ &\text{subject to } \sum_{k=0}^{\nu-1} \beta_k \bar{w}^k(s) = 0, \ s \in S \\ &\qquad\qquad\ \sum_{k=0}^{\nu-1} \beta_k = 1 \\ &\qquad\qquad\ \beta_k \geq 0, \ k = 0, \ldots, \nu - 1. \end{aligned}$$

Let $z^\nu(s)$ for $s \in S$ and θ^ν be the associated multipliers.

Step 2. For each $s \in S$ take

$$\begin{aligned}
u^\nu(s) &\in \operatorname*{argmin}_u \left\{ f(u,s) + \frac{1}{2\lambda}|z^\nu(s) - u|^2 \right\} \\
w^\nu(s) &:= \lambda^{-1}(z^\nu(s) - u^\nu(s)), \\
\alpha^\nu(s) &:= -f(u^\nu(s), s) - (1/2\lambda)|z^\nu(s) - u^\nu(s)|^2 + z^\nu(s)w^\nu(s), \\
\bar{w}_t^\nu(s) &:= (E^{t-1}w_t)(s), t = 1, \dots, T;\ s \in S, \\
\alpha_\nu &:= \sum_{s\in S} p_s \alpha_\nu(s).
\end{aligned}$$

Step 3. If $\alpha_\nu < \sum_{s\in S} z^\nu(s)\bar{w}^\nu(s) + \theta^\nu$ return to Step 1 with counter ν increased by 1. Otherwise, stop; z^ν is the optimal solution.

7. Justification of the Multistage Version

As for the 1-stage model, the linear program formulated in Step 1 of the algorithm is an inner approximation of the dual problem (up to a sign change for the objective value). Step 2 corresponds to the construction of a new column for this linear program which would enable us to get a better solution to its dual problem. Thus, we are looking for

$$\{w(s)\}_{s\in S} \text{ with } \bar{\alpha} = \sum_{s\in S} p_s f_\lambda^*(w(s), s), \qquad \bar{w} = Jw,$$

such that

$$\bar{\alpha} - \sum_{s\in S} z^\nu(s)\bar{w}(s) < \theta^\nu,$$

or equivalently, for a dual element $\{w(s)\}_{s\in S}$ such that

$$\sum_{s\in S} p_s f_\lambda^*(w(s), s) - \sum_{t=1}^{T}\sum_{s\in S} z_t^\nu(s) E\left\{w_t \mid \vec{s}^{\,t-1}\right\}(s) < \theta^\nu,$$

which can also be written as

$$\sum_{s\in S} p_s \left\{ f_\lambda^*(w(s), s) - z^\nu(s)w(s) \right\} < \theta^\nu.$$

From this point on, the analysis is the same as that for the 1-stage case.

8. Comments

The linear program to be solved in Step 1 of the algorithm for the multistage problem is much larger than in the 1-stage case. It involves a large number of columns: $(\nu\times\ |S|)$ with $|S|$ = cardinality of S. By exploiting the properties of J, we see that this linear program can be substantially simplified.

As formulated, the constraints of the linear program in Step 1 are

$$\sum_{k=0}^{\nu-1} \beta_k = 1, \quad \beta_k \geq 0, k = 0, \ldots, \nu - 1,$$

and

$$\sum_{k=0}^{\nu-1} \beta_k \bar{w}_t^k(s) = 0, \quad t = 1, \ldots, T, s \in S,$$

with

$$\bar{w}_t^k(s) \in \mathbb{R}^{n_t}.$$

However for all t, we have

$$\bar{w}_t^k(s) = \bar{w}_t^k(s') \quad \text{if} \quad \vec{s}^{\,t-1} = \vec{s'}^{\,t-1}$$

as follows from the definition of J. For all $t = 1, \ldots, T-1$, the S equations $\sum_{k=0}^{\nu-1} \beta_k \bar{w}_t^k(s) = 0, s \in S$ can be replaced by a much smaller number of equations. In fact, it suffices to have 1 equation for each element in the partition of S that will represent all s that have the same "past", i.e., for given $\vec{s}^{\,t-1}$, all s' such that $\vec{s'}^{\,t-1} = \vec{s}^{\,t-1}$. This means that when $t = 1$ only 1 equation (instead of $|S|$ equations) is needed, since we have $\bar{w}_1^k(s) = \bar{w}_1^k(s')$ for all $s, s' \in S$. For $t = 2$, one only needs to include as many equations as there are different realizations of s_1, and so on. Note that the resulting z^ν will automatically satisfy the nonanticipativity condition.

References

1. Attouch, Hedy and Roger J-B Wets, "Epigraphical analysis" in *Analyse Non Linéaire* ed. I. Ekeland, pp. 73-100, Gauthier-Villars, Paris, 1989.
2. Dantzig, George B., *Linear Programming and Extensions*, Princeton University Press, 1973.
3. Dempster, Michael, "On stochastic programming II: Dynamic problems under risk," *Stochastics*, vol. 25, pp. 15-42, 1988.
4. Rockafellar, R.T., *Convex Analysis*, Princeton University Press, Princeton, 1970.
5. Rockafellar, R.T. and Roger J-B Wets, "Nonanticipativity and $\mathcal{L}^1$-martingales in stochastic optimization problems," *Mathematical Programming Study*, vol. 6, pp. 170-187, 1976, also in *Stochastic Systems: Modelling, Identification and Optimization*, Roger J-B Wets (ed.), North-Holland, Amsterdam, 1976.
6. Rockafellar, R.T. and Roger J-B Wets, "Deterministic and stochastic optimization problems of Bolza type in discrete time," *Stochastics*, vol. 10, pp. 273-312, 1983.
7. Rockafellar, R.T. and Roger J-B Wets, "Scenarios and policy aggregation in optimization under uncertainty," *Mathematics of Operations Research*, vol. 16, pp. 1-29, 1991.
8. Roger J-B Wets, "On the relation between stochastic and deterministic optimization," in *Control Theory, Numerical Methods and Computer Systems Modelling*, A. Bensoussan and J.L. Lions (eds.), pp. 350-361, Springer Verlag, Berlin, 1975. Lecture Notes in Economics and Mathematical Systems, 107.
9. Roger J-B Wets, "The aggregation principle in scenario analysis and stochastic optimization," in *Algorithms and Model Formulations in Mathematical Programming*, S. Wallace (ed.), pp. 91-113, Springer-Verlag, NATO ASI Vol. 51, Berlin, 1989.