***American Journal of Computational Mathematics*, 2011, 1, 294-302**
doi:10.4236/ajcm.2011.14036 Published Online December 2011 (http://www.SciRP.org/journal/ajcm)

First Order Convergence Analysis for Sparse Grid Method in Stochastic Two-Stage Linear Optimization Problem

Shengyuan Chen
Department of Mathematics and Statistics, York University, Toronto, Canada
E-mail: *chensy@mathstat.yorku.ca*

Received August 27, 2011; *revised September* 20, 2011; *accepted September* 30, 2011

Abstract

Stochastic two-stage linear optimization is an important and widely used optimization model. Efficiency of numerical integration of the second stage value function is critical. However, the second stage value function is piecewise linear convex, which imposes challenges for applying the modern efficient spare grid method. In this paper, we prove the first order convergence rate of the sparse grid method for this important stochastic optimization model, utilizing convexity analysis and measure theory. The result is two-folded: it establishes a theoretical foundation for applying the sparse grid method in stochastic programming, and extends the convergence theory of sparse grid integration method to piecewise linear and convex functions.

Keywords: Convergence Analysis, Stochastic Optimization, Scenario Generation, Convex Analysis, Measure Theory

1. Introduction

Stochastic two-stage linear optimization, also called stochastic two-stage linear programming, models a sequential decision structure, where the first stage decisions are made now before the random variable manifests itself; and the second stage decisions are made adaptively to the realized random variable and the first stage decisions. The adaptive decision model has been applied many important application areas. For example, in the introductory farmer's problem [1], a farmer needs to divide the land for different vegetables in spring. The farmer's objective is to maximize profit in the harvest season. The profit is related to the market price at that time and the weather dependent yield. Neither the price nor the weather is known at the present time, hence the farmer's decision in spring has to take into account multiple scenarios. It is not a simple forecasting problem though, since the farmer's second stage decision in fall, which adapts to different scenarios, also jointly determines the profit. The second stage decision problem is also called "recourse" problem. [2] collects more recent applications in engineering, manufacture, finance, transportation, telecommunication *et al*.

A stochastic two-stage linear problem with recourse has the following general representation:

$$\begin{gathered}\min_{x\in X} c^T x + \int_{\Xi} \rho(x,\xi) P(\mathrm{d}\xi), \text{and} \\ \rho(x,\xi) = \min s^T y \\ s.t.\, Qy = \xi + Tx, y \ge 0,\end{gathered} \tag{1}$$

where ξ is a random vector properly defined on $(\Xi, \mathcal{F}, P)$, X is a polytope feasible region for the first stage, $Q \in \mathbb{R}^{m\times n}$, s and T are a vector and a matrix of proper sizes, $\rho : (X, \Xi) \to \mathbb{R} \cup \{\infty, -\infty\}$ is a real valued function.

The high dimensional integration in (1) is difficult and is usually approximated by using a set of scenarios and weights $\{\xi^k, w^k\}, k = 1, \cdots, K$ as:

$$\begin{gathered}\min_{x\in X} c^T x + \sum_{k=1}^{K} w^k \rho(x,\xi^k), \text{and} \\ \rho(x,\xi^k) = \min d^T y \\ s.t.\, Qy = \xi^k + Tx, y \ge 0.\end{gathered} \tag{2}$$

Under this scenario approximation, the optimal objective value $\tilde{z}_K^*$ of (2) provides an approximation of the optimal objective value z^* of (1). An optimal solution $\tilde{x}_K^*$ of (2) provides an approximation of an optimal solution x^* of (1).

Monte Carlo (MC) method has been widely used in this approximation, where $\xi^k, k = 1, \cdots, K$ are random sampling points and $w^k = 1/K$. The convergence theory of Monte Carlo method has been extensively studied [3-6]. The core result is the epi-convergent theorem: under mild assumptions, $\tilde{z}_K^*$ converges to z^* w.p.1 as $K \to \infty$; and any clustering point of the $\{x_K^*\}_{K=1}^{\infty}$, which is the sequence of optimal solutions to (2), is in the optimal solution set of the original problem. Quasi Monte-Carlo (QMC) method has also been recently studied [7], and similar convergence result has been achieved.

The sparse grid (SP) method is an established high dimensional quadrature rule, which was originally proposed by Smolyak [8], and has been studied by many authors in the context of numerical integration [9] (and references therein). Its application in the stochastic two-stage linear optimization is only shown in a recent numerical study in [10]. Though [10] shows the superior numerical performance of sparse grid method, compared with both MC and QMC, the convergence analysis is based on an assumption that the recourse function is in a Sobolev space, which only holds for a very narrow subset of the two-stage linear problems, *i.e.*, separable problems. The contribution of this paper are 1) establishing the epi-convergence of the sparse grid method for this important decision model; 2) prove the first order convergence rate of the method.

We first introduce the spare grid approximation error for integrand functions in Sobolev spaces.

Let D_j denote the partial derivative operator

$$\left(D_j f\right)(x) = \frac{\partial f}{\partial x_j}(x)$$

Let $\alpha = \left(\alpha_1, \cdots, \alpha_d\right), \alpha_j \in \mathbb{N}^+$ be a multi-index, and define

$$D^\alpha = \prod_{j=1}^{d} D_j^{\alpha_j} = \frac{\partial^{|\alpha|}}{\prod_{j=1}^{d} \partial x_j^{\alpha_j}},$$

where $|\alpha| = \alpha_1 + \cdots + \alpha_d$. The Sobolev space with smoothness parameter $r \geq 1$ is defined as

$$\mathcal{W}_d^r = \left\{ f : D^\alpha f \in \mathcal{L}_2 [0,1]^d \text{ for all } \alpha \leq r \right\},$$

where $\alpha \leq r$ means component-wisely $\alpha_j \leq r,$ $\forall j = 1, \cdots, d$. Sobolev spaces could also be defined using $\mathcal{L}_p$ norms, see Evans [11]. The derivatives in the definition of Sobolev space are weak derivatives. Formally, $D^\alpha f$ is the α-derivative of f if for all $v \in C_0^\infty [0,1]^d$, *i.e.*, infinitely differentiable function on $(0,1)^d$, $D^\alpha f$ satisfies the following equation:

$$\int_{[0,1]^d} \left(D^\alpha f\right)(x) v(x) \mathrm{d}x$$
$$= (-1)^{|\alpha|} \int_{[0,1]^d} f(x) \left(D^\alpha v\right)(x) \mathrm{d}x.$$

For example, $f(x) = |x|$ defined in $[0,1]$ has the first order weak derivative function $Df = \operatorname{sign}(x)$; but the function is nondifferentiable at 0 in the usual strong sense. It has been shown that weak derivative is essentially unique, and coincides with the classical strong derivative when it exists. Various properties of strong derivatives carry over to weak derivative as well, for example, $D^{\alpha+\beta} f = D^\alpha \left(D^\beta f\right) = D^\beta \left(D^\alpha f\right)$ for all multi-index $\alpha, \beta, |\alpha| + |\beta| \leq r$. For more calculus rules regarding weak derivative, including the extended Leibniz theorem, see Evans ([11], Section 5.2.3).

The norm of the defined Sobolev space is

$$\| f \|_{\mathcal{W}_d^r} = \left\| \left(\left\| D^\alpha f \right\|_{\mathcal{L}_2} \right)_{\alpha \leq r} \right\|_2,$$

where $\|\cdot\|_{\mathcal{L}_2}$ is the L_2 -norm of a function, $\alpha \leq r$ component-wisely, $\|\cdot\|_2$ is the Euclidian 2-norm of a finite vector:

$$\| f \|_{\mathcal{L}_2} = \left(\int_{[0,1]^d} \left| f(x) \right|^2 \mathrm{d}x \right)^{1/2}$$

$$\|h\|_2 = \left(\sum_{j=1}^{k} \left| h_j \right|^2 \right)^{1/2}.$$

For $f \in \mathcal{W}_d^r$, the sparse grid method achieves the following convergence rate [12]:

$$\left| \int_{[0,1]^d} f(\xi) \mathrm{d}_\xi - \sum_{k=1}^{K} w^k f\left(\xi^k\right) \right| \leq \beta_{r,d} \frac{(\log K)^{(d-1)(r+1)}}{K^r} \| f \|_{\mathcal{W}_d^r}, \tag{3}$$

where K is the number of function evaluations, $\beta_{r,d}$ is a constant independent of f, increasing with both d and r, see Brass [13]. Note that $f \in \mathcal{W}_d^r$ implies $f \in \mathcal{W}_d^i, i \leq r$. Since the norm $\| f \|_{\mathcal{W}_d^r}$ and $\beta_{d,r}$ are non-decreasing in r, and the term $K^{-r} (\ln K)^{(d-1)(r+1)}$ is non-increasing in r for large K, it is none trivial to tell which space will yield the tightest bound. The problem is called fat F problem in Wonzniakowski [14]. In this paper, as we shall see, only $r = 1$ is relevant for our discussion.

The convergence result in (3) only holds for the two-stage stochastic linear programming (1) in the trivial case, *i.e.*, when the integrand function $\rho(x, \cdot) : \Xi \to \mathbb{R}$ is separable. For example,

$$\rho\left(x_1, x_2, \xi_1, \xi_2\right) = \min_{y_1^+, y_1^-} y_1^+ + y_1^- + y_2^+ + y_2^-$$

$$y_1^+ - y_1^- = \xi_1 - x_1$$
$$y_2^+ - y_2^- = \xi_2 - x_2$$
$$y_1^+, y_1^-, y_2^+, y_2^- \geq 0$$

is equivalent to

$$\rho(x_1, x_2, \xi_1, \xi_2) = |\xi_1 - x_1| + |\xi_2 - x_2|,$$

where $|\xi_1 - x_1| \in \mathcal{W}_1^1$, $|\xi_2 - x_2| \in \mathcal{W}_1^1$ and the convergence result in (3) can be applied directly.

However, in general, $\rho(x, \cdot)$ is non-separable piecewise function, see Birge and Louveaux [1], and does not belong to $\mathcal{W}_d^r$ for any r. For example,

$$\rho(x_1, x_2, \xi_1, \xi_2) = \min_{y^+, y^-} y^+ + y^-$$
$$y^+ - y^- = \xi_1 - \xi_2 - x_1 - x_2$$
$$y^+, y^- \geq 0$$

is equivalent to

$$\rho(x_1, x_2, \xi_1, \xi_2) = |\xi_1 - \xi_2 - x_1 - x_2|,$$

and does not have the (weak) derivative $D^{(1,1)}\rho(x_1, x_2, \cdot, \cdot)$ since $D^{(0,1)} f$ is discontinuous and non-differentiable even in the weak sense, while $D^{(1,1)} f = D^{(1,0)}\left(D^{(0,1)} f\right)$ if $D^{(1,1)} f$ exists. Hence the error bound in (3) can not be applied to two-stage linear problem directly. The major contribution of this paper is to prove the convergence of (2) to (1) in the rate specified in (3) with $r = 1$, *i.e.*, the first order convergence rate, even though $\rho(x, \cdot) \notin \mathcal{W}_d^1$. On the other hand, this analysis extends the convergence theory of sparse grid method to convex multivariate piecewise linear functions since for such a function $f : \mathbb{X} \to \mathbb{R} := d_i x$ for $x \in B_i$; each B_i is polyhedron and $\{B_1, \cdots, B_m\}$ partitions $\mathbb{X}$, could be represented as[1]

$$f(x) = \min_y y$$
$$y \geq d_i x, i = 1, \cdots, m.$$

The paper is organized as the followings. In Section 2, we introduce a logarithmic mollifier function and prove its various properties. The mollifier function is quite familiar to the optimization community as it is the barrier function used in the Interior Point Method for linear programming. In Section 3, we use the limiting properties of the mollifier function to prove the uniform convergence and the first order convergence rate for the objective function. We also show the converging behaviour of the optimal solutions $\tilde{x}_K^*$ in a subsequence. Finally, Section 4 presents our conclusions.

In the coming sections, we assume $\Xi = [0,1]^d$. For a more general continuous distribution with a inverse cumulative function $F^{-1} : [0,1]^d \to \Xi$, one can apply transformation

$$\int_\Xi f(\xi) F(\mathrm{d}\xi) = \int_{[0,1]^d} f\left(F^{-1}(\omega)\right) \mathrm{d}\omega.$$

The transformation brings in more complexity in the analysis without changing our conclusion, hence we assume $\Xi = [0,1]^d$ in the following sections and extend the analysis through inverse transformation and truncation in the Appendix.

2. Mollifier Function

We make the following mild assumptions of the problem (1):

A1: X is compact with nonempty relative interior; $\forall x \in X, \xi \in \Xi, \rho(x, \xi) < \infty$, or relative;

A2: completeness, and $\rho(x, \xi)$ has nonempty relative interior;

A3: $\operatorname{rank}(Q) = m$;

A4: ξ is a continuous random vector with an invertible cumulative distribution function.

Assumption A1 is necessary for our analysis using the Interior Point Method theory. Assumption A2 is for convenience since otherwise we need to discuss the case $\rho(x, \xi) = \infty$, which will drag our analysis to a different focus. Assumption A3 is implicitly assumed in many analysis of linear programming, since the rows of Q could be preprocessed such that the reduced Q has full row rank. Assumption A4 facilitates the conversion from a unit cube $[0,1]^d$ to Ξ through the inverse c.d.f. transformation.

We define a mollifier function $\rho_{\mu,\xi} : \Xi \to \mathbb{R}$:

$$\begin{aligned} \rho_{\mu,x}(\xi) = \min s^T y + \mu B(y) \\ s.t. Qy = \xi + Tx, \end{aligned} \tag{4}$$

where $B(y) = -\sum_{i=1}^n \log y_i$. In the following, we call $\rho(x, \xi)$ the recourse function, and $\rho_{\mu,x}(\xi)$ the mollifier function. We let

$$g(y) = \nabla B(y) = \left(-\frac{1}{y_1}, \cdots, -\frac{1}{y_n}\right)^T$$
$$H(y) = \nabla^2 B(y) = \operatorname{diag}\left(\frac{1}{y_1^2}, \cdots, \frac{1}{y_n^2}\right). \tag{5}$$

$B(y)$ is in fact a barrier function widely used in the Interior Point Method for linear programming, and its properties are well studied. As $\mu \to 0^+$, the convergence of $\rho_{\mu,x}(\xi)$ and $\left(y_{\mu,x}^*, u_{\mu,x}^*\right)$ is stated in the following theorem.

Theorem 2.1. *For any* $x \in X, \mu \in (0,1)$, *let* $\left(y_{\mu,x}^*, u_{\mu,x}^*\right)$ *be the optimal primal and dual solutions of the mollifier*

[1] We thank John Birge for pointing out this elegant argument.

function, then

$$\lim_{\mu\to 0^+}\rho_{\mu,x}(\xi)=\rho(x,\xi)$$

$$\lim_{\mu\to 0^+}\left(y^*_{\mu,x},u^*_{\mu,x}\right)=\left(y^*_x,u^*_x\right),$$

where $\left(y^*_x,u^*_x\right)$ *is an optimal primal and dual pair of the recourse function* $\rho(x,\xi)$.

Proof. Due to the barrier function $B(\cdot)$, the objective function of $\rho_{\mu,x}(\xi)$ is strictly convex. Together with the relative completeness assumption, it is clear that $\rho_{\mu,x}(\xi)$ has an unique optimal solution. Since the problem has non-empty relative interior by assumption, the central path $\left\{\left(y^*_{\mu,x},u^*_{\mu,x}\right)\right\}_{\mu}$ exists for each x, and converges to the analytical center of the optimal set, see Roos *et al.* [15] Theorem I.30 and its Definition I.20 for analytic center. □

The converging property of $\left(y^*_{\mu,x},u^*_{\mu,x}\right)$ (central path) as $\mu\to 0^+$ has been an important topic in Interior Point Method, and has been extensive studied by many authors. For interested readers, in addition to the reference given in the proof, we refer the extensive research in Megiddo [16], an early work Fiacco [17], and a survey of degeneracy and IPM by Güler *et al.* [18]. For readers interested in the interior point method in general, we refer Nesterov and Nemirovskii [19], Renegar [20] and Wright [21].

The KKT condition of the optimization problem appeared in (4) is

$$F_{\mu,x}(y,u,\xi)=\begin{pmatrix} s+\mu g+Q^T u \\ Qy-\xi-Tx\end{pmatrix}=0,$$

where $F_{\mu,x}:\mathbb{R}^n_+\times\mathbb{R}^m\times\Xi\to\mathbb{R}^{(m+1)}$. Clearly $F_{\mu,x}(\cdot,\cdot,\cdot)$ is infinitely differentiable. Furthermore,

$$\nabla_{(y,u)}F_{\mu,x}=\begin{pmatrix}\mu H\left(y^*_{\mu,x}\right) & Q^T \\ Q & 0\end{pmatrix}$$

$$\nabla_{\xi}F_{\mu,x}=\begin{pmatrix}0\\-I\end{pmatrix}, \tag{6}$$

where I is an identity matrix, $\nabla_{(y,u)}F_{\mu,x}$ is invertible since $H(\cdot)$ is positive definite. Hence by the implicit function theorem, $y^*_{\mu,x},u^*_{\mu,x}$ are infinitely differentiable functions of ξ. So $\rho_{\mu,x}(\xi)=s^Ty^*_{\mu,x}+\mu B\left(y^*_{\mu,x}\right)$ is infinitely differentiable.

Proposition 2.1. *For any* $x\in X,\mu\in(0,1)$,

$$\rho_{\mu,x}(\cdot):\Xi\to\mathbb{R}\in C^{\infty}.$$

In the following, we directly derive the (strong) partial derivatives of $D^{\alpha}\rho_{\mu,x}(\cdot)$ for all $\alpha\le 1$. Note that there are 2^d-1 number of α s satisfying $\alpha\le 1$. We also prove these partial derivatives are finite for all $\mu\in(0,1)$, $x\in X$. Finally, we show that their limits are finite when $\mu\to 0^+$. Hereinafter, a vector $v<\infty$ or a matrix $M<\infty$ means the inequality holds component-wisely.

Proposition 2.2. *For all* $x\in X,\mu\in(0,1)$,

$$\nabla\rho_{\mu,x}(\xi)=-u^*_{\mu,x}<\infty.$$

Furthermore, $\lim_{\mu\to 0^+}\nabla\rho_{\mu,x}(\xi)=-u^*_x<\infty$.

Proof. The Lagrangian function of the optimization problem in (4) is

$$L_{\mu,x}(y,u,\xi)=s^Ty+\mu B(y)+u^TQy-u^T\xi-u^TTx.$$

Since $L_{\mu,x}\left(y^*_{\mu,x},u^*_{\mu,x},\xi\right)=\rho_{\mu,x}(\xi)$,

$$\begin{aligned}\nabla\rho_{\mu,x}(\xi)&=\nabla L_{\mu,x}\left(y^*_{\mu,x},u^*_{\mu,x},\xi\right)\\
&=\frac{\partial L_{\mu,x}}{\partial\xi}+\frac{\partial L_{\mu,x}}{\partial y^*_{\mu,x}}\frac{\partial y^*_{\mu,x}}{\partial\xi}+\frac{\partial L_{\mu,x}}{\partial u^*_{\mu,x}}\frac{\partial u^*_{\mu,x}}{\partial\xi}\\
&=-u^*_{\mu,x}+\left(s+\mu g\left(y^*_{\mu,x}\right)+Q^Tu^*_{\mu,x}\right)\frac{\partial y^*_{\mu,x}}{\partial\xi}\\
&\quad+\left(Qy^*_{\mu,x}-\xi-Tx\right)\frac{\partial u^*_{\mu,x}}{\partial\xi}\\
&=-u^*_{\mu,x},\end{aligned}$$

where the last equality follows the KKT condition. Clearly $u^*_{\mu,x}(\xi)<\infty$. As $\mu\to 0^+$, $u^*_{\mu,x}\to u^*_x$ by Theorem 2.1.

We let

$$\mathcal{Y}_{\mu,x}=\operatorname{diag}\left(y^*_{\mu,x}\right),\mathcal{Y}_x=\operatorname{diag}\left(y^*_x\right).$$

Recall that the y^*_x refers to the limiting point defined in the Theorem 2.1, not an arbitrary optimal solution of $\rho(x,\xi)$.

Lemma 2.1. *For any* $x\in X,\mu\in(0,1)$,

$$\nabla u^*_{\mu,x}(\xi)=-\mu\left(Q\mathcal{Y}^2_{\mu,x}Q^T\right)^{-1},$$

$$\nabla y^*_{\mu,x}(\xi)=\mathcal{Y}^2_{\mu,x}Q^T\left(Q\mathcal{Y}^2_{\mu,x}Q^T\right)^{-1}.$$

Proof. $F\left(\xi,y^*_{\mu,x},u^*_{\mu,x}\right)=0$ defines implicit functions $\left(y^*_{\mu,x},u^*_{\mu,x}\right)$. With $\nabla_{(y,u)}F_{\mu,x},\nabla_{\xi}F_{\mu,x}$ given in (6), $\nabla_{(y,u)}F^{-1}_{\mu,x}$ can be explicitly computed as

$$\begin{pmatrix}\frac{1}{\mu}H^{-1}\left(I-Q^T\left(QH^{-1}Q^T\right)^{-1}QH^{-1}\right) & H^{-1}Q^T\left(QH^{-1}Q^T\right)^{-1}\\ \left(QH^{-1}Q^T\right)^{-1}QH^{-1} & -\mu\left(QH^{-1}Q^T\right)\end{pmatrix},$$

where H is a shorthand notation of $H\left(y^*_{\mu,x}\right)$. Hence by the implicit function theorem

$$\begin{pmatrix}\nabla y^*_{\mu,x}(\xi)\\ \nabla u^*_{\mu,x}(\xi)\end{pmatrix}=-\nabla_{(y,u)}F^{-1}_{\mu,x}\nabla_{\xi}F_{\mu,x},$$

and the conclusion follows straightforward computation and (5).

Proposition 2.3. *For all* $x \in X, \mu \in (0,1)$, $\nabla^2 \rho_{\mu,x}(\xi) = \mu\left(Q\mathcal{Y}^2_{\mu,x}Q^T\right)^{-1} < \infty$. *Furthermore,* $\lim_{\mu\to 0^+}\nabla^2 \rho_{\mu,x}(\xi) = 0$ *essentially.*

Proof. By Proposition 2.2 and Lemma 2.1,

$$\nabla^2 \rho_{\mu,x}(\xi) = \mu\left(Q\mathcal{Y}^2_{\mu,x}Q^T\right)^{-1} \qquad (7)$$

Hence $\nabla^2 \rho_{\mu,x}(\xi) < \infty, \forall \xi \in \Omega, x \in X$. By Theorem 2.1,

$$\lim_{\mu\to 0^+}\nabla^2 \rho_{\mu,x}(\xi) = 0\cdot\left(Q\mathcal{Y}^2_{x}Q^T\right)^{-1}.$$

If the optimal set of $\rho(x,\xi)$ is non-degenerate, $Q\mathcal{Y}^2_x Q^T$ is non-singular, hence the above limit is zero. If the optimal set is degenerate, then the limit $0\cdot\infty$ is not defined. However, in the following, we show that the degenerated case has zero Lebesgue measure, hence the limit is zero with probability one. Let's first consider a special case of degeneration. Let the first m columns of Q be an optimal basis B and $y_1 = 0, y_2, \cdots, y_m > 0, y_{m+1} = \cdots = y_n =$ be a degenerated optimal basic feasible solution. Define the set $E = \left\{\xi \mid y^*_B = B^{-1}(\xi + Tx), y_1 = 0, y_2, \cdots, y_m > 0\right\} 0$. Since set $Y = \left\{y \in \mathbb{R}^m \mid y_1 = 0, y_2, \cdots, y_m > 0\right\}$ has zero Lebesgue measure in $\mathbb{R}^m$, and B is both injective and surjective, hence $m(E) = m(B(Y)) = 0$. Clearly the same argument holds for an arbitrarily degenerated optimal basis B with an arbitrarily chosen degenerated basic column. Furthermore, there are only finitely many ways to choose basis and degenerated columns. Hence the total measure of the set ξ which leads to a degenerated $\rho(\xi,x)$ is zero. Hence we conclude that the limit is zero essentially. □

We continue to calculate higher order partial derivatives. Note that $\nabla^2 \rho_{\mu,x}(\xi)$ is a $m \times m$ matrix, and

$$\frac{\partial^3 \rho_{\mu,x}(\xi)}{\partial\xi_i \partial\xi_j \partial\xi_k} = \left[\frac{\partial}{\partial\xi_k}\nabla^2 \rho_{\mu,x}(\xi)\right]_{ij}.$$

In general, for any $k \le d$ and any set of $i_1, \cdots, i_k$, $\frac{\partial}{\partial\xi_{i_3}}\cdots\frac{\partial}{\partial\xi_{i_k}}\nabla^2 \rho_{\mu,x}(\xi)$ is a $m \times m$ matrix, and

$$\frac{\partial^k \rho_{\mu,x}(\xi)}{\partial\xi_{i_1}\partial\xi_{i_2}\cdots\partial\xi_{i_k}} = \left[\frac{\partial}{\partial\xi_{i_3}}\cdots\frac{\partial}{\partial\xi_{i_k}}\nabla^2 \rho_{\mu,x}(\xi)\right]_{i_1,i_2}.$$

Proposition 2.4. *For any* $x \in X, \mu \in (0,1)$, *and any set of indices* $i_1, \cdots, i_k$,

$$\frac{\partial^k}{\partial\xi_{i_1}\cdots\partial\xi_{i_k}}\rho_{\mu,x}(\xi) < \infty,$$

$$\lim_{\mu\to 0^+}\frac{\partial^k}{\partial\xi_{i_1}\cdots\partial\xi_{i_k}}\rho_{\mu,x}(\xi) = 0$$

essentially.

Proof. We first prove the conclusion for $k = 2$, and extend by induction.

$$\begin{aligned}
\frac{\partial}{\partial\xi_k}\nabla^2 \rho_{\mu,x}(\xi) &= \frac{\partial}{\partial\xi_k}\mu\left(Q\mathcal{Y}^2_{\mu,x}Q^T\right)^{-1} \\
&= \mu\left\{\left(Q\mathcal{Y}^2_{\mu,x}Q^T\right)^{-1} Q\left(\frac{\partial}{\partial\xi_k}\mathcal{Y}^2_{\mu,x}\right)Q^T\left(Q\mathcal{Y}^2_{\mu,x}Q^T\right)^{-1}\right\} \\
&= 2\mu\left(Q\mathcal{Y}^2_{\mu,x}Q^T\right)^{-1} Q\mathcal{Y}\begin{pmatrix} \frac{\partial y^*_{1,\mu,x}}{\partial\xi_k} & 0 & \cdots & 0 \\ 0 & \frac{\partial y^*_{2,\mu,x}}{\partial\xi_k} & \cdots & 0 \\ 0 & \cdots & \ddots & 0 \\ 0 & \cdots & \cdots & \frac{\partial y^*_{n,\mu,x}}{\partial\xi_k} \end{pmatrix} \qquad (8) \\
&\quad \cdot Q^T\left(Q\mathcal{Y}^2_{\mu,x}Q^T\right)^{-1},
\end{aligned}$$

where

$$\frac{\partial y^*_{j,\mu,x}}{\partial\xi_k} = \left[\nabla y^*_{\mu,x}(\xi)\right]_{jk} = \left[\mathcal{Y}^2_{\mu,x}Q^T\left(Q\mathcal{Y}^2_{\mu,x}Q^T\right)^{-1}\right]_{jk}$$

is shown in lemma 2.1. Clearly (8) is finite for $\mu > 0$. Now taking limit $\mu \to 0$, by Theorem 2.1 and Proposition 2.3:

$$\lim_{\mu\to 0^+}\frac{\partial}{\partial\xi_k}\nabla^2 \rho_{\mu,x}(\xi) = 0 \text{ essentially.}$$

Furthermore, we prove by induction that $\frac{\partial}{\partial\xi_{i_3}}\cdots\frac{\partial}{\partial\xi_{i_k}}\nabla^2 \rho_{\mu,x}(\xi)$ is in the form $2\mu\cdot S\left(Q, \mathcal{Y}_{\mu,x}, \mathcal{Y}_{\mu,x}\right)$, where $S(\cdot)$ is an algebraic expression involving multiplication and summation of $\mathcal{Y}_{\mu,x}, \left(Q\mathcal{Y}^2_{\mu,x}Q^T\right)^{-1}$, Q. The claim is true for $\frac{\partial}{\partial\xi_k}\nabla^2 \rho_{\mu,x}(\xi)$ in (8). Suppose it is true for $\frac{\partial}{\partial\xi_{i_3}}\cdots\frac{\partial}{\partial\xi_{i_{k-1}}}\nabla^2 \rho_{\mu,x}(\xi)$, then by the chain rule, the term $\left(Q\mathcal{Y}^2_{\mu,x}Q^T\right)^{-1}$ will be expanded into multiplication of the same terms $\mathcal{Y}_{\mu,x}, \left(Q\mathcal{Y}^2_{\mu,x}Q^T\right)^{-1}$, Q, hence the form $2\mu\cdot S\left(Q, \mathcal{Y}_{\mu,x}, \mathcal{Y}_{\mu,x}\right)$ is established. It is clear that $\lim_{\mu\to 0^+}2\mu\cdot S\left(Q, \mathcal{Y}_{\mu,x}, \mathcal{Y}_{\mu,x}\right) = 0$ essentially by the Theorem 2.1 and Proposition 2.3. We also derive higher order partial derivatives explicitly in the Appendix A. □

Theorem 2.2. *For any* $x \in X, \mu \in (0,1)$,

$$\left\|\rho_{\mu,x}(\xi)\right\|_{\mathcal{W}_n^1} < \infty;$$

$$\lim_{\mu\to 0^+}\left\|\rho_{\mu,x}(\xi)\right\|_{\mathcal{W}_n^1} = \left(\left\|u_{1,x}^*\right\|_{\mathcal{L}_2}^2 + \cdots, \left\|u_{m,x}^*\right\|_{\mathcal{L}_2}^2\right)^{\frac{1}{2}} < \infty.$$

Furthermore,

$$\mathcal{C} = \sup_{x\in X}\left(\left\|u_{1,x}^*\right\|_{\mathcal{L}_2}^2 + \cdots, \left\|u_{m,x}^*\right\|_{\mathcal{L}_2}^2\right)^{\frac{1}{2}} < \infty.$$

Proof. follows the definition of the norm $\|\cdot\|_{\mathcal{W}_n^1}$, Propositions 2.2, 2.3, 2.4, and Theorem 2.1. The finiteness of $u_{j,x}^*$, $j = 1,\cdots,m$, follows the relative completeness assumption of the two-stage linear problem and duality theory. Furthermore, X is compact, hence $\mathcal{C} < \infty$. □

3. First Order Convergence Rate

We first discuss the convergence of the objective function specified in the approximation model (2) to the objective function of the true model (1), and the convergence rate.

Theorem 3.1. *For all* $x \in X$,

$$\left|\int_{[0,1]^d}\rho(x,\xi)\,d_\xi - \sum_{k=1}^{K} w^k \rho\left(x,\xi^k\right)\right| \le \mathcal{C}\beta_{1,d}\frac{(\log K)^{2(d-1)}}{K}.$$

Hence,

$$\sum_{k=1}^{K} w^k \rho\left(x,\xi^k\right) \xrightarrow{K\to\infty} \int_{[0,1]^d}\rho(x,\xi)\,d_\xi \textit{ uniformly.}$$

Proof. For notation convenience, define operators E and A_K as

$$E(f) = \int_{[0,1]^d} f(\xi)\,\mathrm{d}_\xi,$$

$$A_K(f) = \sum_{k=1}^{K} w^k f\left(\xi^k\right).$$

Then for any $x \in X, \mu \in (0,1), K \in \mathbb{N}$,

$$\begin{aligned}&\left|E(\rho(x,\cdot)) - A_K(\rho(x,\cdot))\right|\\ &\le \left|E(\rho(x,\cdot)) - E\left(\rho_{\mu,x}(\cdot)\right)\right|\\ &\quad + \left|E\left(\rho_{\mu,x}(\cdot)\right) - A_K\left(\rho_{\mu,x}(\cdot)\right)\right|\\ &\quad + \left|A_K\left(\rho_{\mu,x}(\cdot)\right) - A_K(\rho(x,\cdot))\right|\end{aligned}$$

Taking limiting $\mu \to 0^+$ on both sides, then the first and third term on the right hand side go to zero by Theorem 2.1, and the second term is bounded by the classical convergence rate of sparse grid method, see (3), Proposition 2.3 and Theorem 2.2. □

Let the objective function of the true problem (1) and the approximated problem (2) be

$$z(x) = c^T x + \int_{[0,1]^d}\rho(x,\xi)\,P(\mathrm{d}\xi),$$

$$\tilde{z}_K(x) = c^T x + \sum_{k=1}^{K} w^k \rho\left(x,\xi^k\right),$$

and let the optimal objective value and optimal solution set of the true model and approximated model be $z^*, X^*, \tilde{z}_K^*, X_K^*$ respectively. Theorem 3.2 and Theorem 3.3 state the results for the optimal objective value and the optimal sets separately.

Theorem 3.2. *The optimal objective value converges, i.e.,* $\lim_{K\to\infty}\tilde{z}_K^* \to z^*$, *and the rate of convergence is*

$$\left|z^* - \tilde{z}_K^*\right| \le \mathcal{C}\beta_{1,d}\frac{(\log K)^{2(d-1)}}{K}.$$

Proof. For the minimization problem we note that $z\left(x^*\right) \le z(x)$ for any $x^* \in X^*, x \in X$, and $\tilde{z}_K\left(\tilde{x}_K\right) \le \tilde{z}_K(x)$ for any $\tilde{x}_K \in X_K^*, x \in X$. Let $\varepsilon_K = \mathcal{C}\beta_{1,d}\frac{(\log K)^{2(d-1)}}{K}$, then

$$\begin{aligned}z\left(x^*\right) - \tilde{z}_K\left(\tilde{x}_K\right) &= z\left(x^*\right) - z\left(\tilde{x}_K\right) + z\left(\tilde{x}_K\right) - \tilde{z}_K\left(\tilde{x}_K\right)\\ &\le z\left(\tilde{x}_K\right) - \tilde{z}_K\left(\tilde{x}_K\right) \le \varepsilon_K\end{aligned}$$

$$\begin{aligned}z\left(x^*\right) - \tilde{z}_K\left(\tilde{x}_K\right) &= z\left(x^*\right) - \tilde{z}_K\left(x^*\right) + \tilde{z}_K\left(x^*\right) - \tilde{z}_K\left(\tilde{x}_K\right)\\ &\ge z\left(x^*\right) - \tilde{z}_K\left(x^*\right) \ge -\varepsilon_K,\end{aligned}$$

where the inequalities follow Theorem 3.1. □

Theorem 3.3. *For any* $\tilde{x}_K \in X_K^*, x^* \in X^*$,

1) $\tilde{x}_K$ *is feasible*;

2) $\left|z\left(\tilde{x}_K\right) - z\left(x^*\right)\right| \le 2\mathcal{C}\beta_{1,d}\frac{(\log K)^{2(d-1)}}{K}$;

3) *For any clustering point* $\tilde{x}^*$ *of a subsequence* $\tilde{x}_{K_t}, t \in \mathbb{N}, \tilde{x}_{K_t} \in X_{K_t}^*$, $\tilde{x}^* \in X^*$; *furthermore, if* $X^* = \left\{x^*\right\}$ *is a singleton, then* $\tilde{x}^* = x^*$.

Proof. $\tilde{x}_K$ satisfies the first stage constraints and by the relative completeness assumption, $\tilde{x}_K$ is feasible. To show that $\tilde{x}_K$ is also very close to any optimal solution, we apply the similar technique used in Theorem 3.1.

$$\begin{aligned}z\left(\tilde{x}_K\right) - z\left(x^*\right) &= z\left(\tilde{x}_K\right) - \tilde{z}_K\left(\tilde{x}_K\right) + \tilde{z}_K\left(\tilde{x}_K\right) - z\left(x^*\right)\\ &\in \left[-2\varepsilon_K, 2\varepsilon_K\right],\end{aligned}$$

since $-\varepsilon_K \le z\left(\tilde{x}_K\right) - \tilde{z}_K\left(\tilde{x}_K\right) \le \varepsilon_K$ by Theorem 3.1, and $-\varepsilon_K \le \tilde{z}\left(\tilde{x}_K\right) - z\left(x^*\right) \le \varepsilon_K$ by the steps in the proof of Theorem 3.3. Since $\tilde{x}^*$ is a clustering point, $\left|z\left(\tilde{x}^*\right) - z\left(x^*\right)\right| = \lim_{t\to\infty}\left|z\left(\tilde{x}_{K_t}\right) - z\left(x^*\right)\right| = 0$ by the inequality above. As a special case, if $X = \left\{x^*\right\}$, then $\tilde{x}^* = x^*$.

The third result of Theorem 3.3 is a classical result based the uniform convergence, see Römisch [22]. The result is stated in subsequence since the optimal sets are not necessarily singletons, and one can only expect optimality of clustering points. The result can also be proved by epi-convergence, see Attouch [23,24]. Epi-convergence is implied by uniform convergence, see Kall [25].

4. Conclusions

The modern sparse grid method is very efficient in numerical integration for integrant functions the Sobolev space $\mathcal{W}_d^r$. However, the integrand function in two-stage linear programming does not belong to $\mathcal{W}_d^r$. We prove that the sparse grid method for the stochastic two-stage linear programming not only converges but also converges in the first order rate. Our constructive proof uses a logarithmic mollifier function from interior point method.

5. References

[1] J. R. Birge and F. Louveaux, “Introduction to Stochastic Programming,” Springer, New York, 1997.

[2] S. W. Wallace and W. T. Ziemba, Eds., “Applications of Stochastic Programming,” Society for Industrial and Applied Mathematics, Philadelphia, 2005.

[3] A. J. King and R. J.-B Wets, “Epi-Convergency of Convex Stochastic Programs,” *Stochastic and Stochastic Reports*, Vol. 34, 1991, pp. 83-92.

[4] A. J. King and R. T. Rockafellar, “Asymptotic Theory for Solutions in Statistical Estimation and Stochastic Programming,” *Mathematics for Operations Research*, Vol. 18, No. 1, 1993, pp. 148-162. doi:10.1287/moor.18.1.148

[5] A. Shapiro, “Asymptotic Analysis of Stochastic Programs,” *Annals of Operations Resesrch*, Vol. 30, No. 1, 1991, pp. 169-186. doi:10.1007/BF02204815

[6] J. Dupacova and R. Wets, “Asymptotic Behavior of Statistical Estimators and of Optimal Solutions of Stochastic Optimization Problems,” *Annals of Statistics*, Vol. 16, No. 4, 1988, pp. 1517-1549. doi:10.1214/aos/1176351052

[7] T. Pennanen and M. Koivu, “Epi-Convergent Discretization of Stochastic Programs via Integration Quadratures,” *Numerische Mathematik*, Vol. 100, No. 1, 2005, pp. 141-163. doi:10.1007/s00211-004-0571-4

[8] S. A. Smolyak, “Interpolation and Quadrature Formula for the Class W_s^a and E_s^a,” *Doklady Akademii Nauk SSSR*, Vol. 131, 1960, pp. 1028-1031. (in Russian, English Translation: *Soviet Mathematica Doklady*, Vol. 4, 1963, pp. 240-243).

[9] T. Gerstner and M. Griebel, “Numerical Integration Using Sparse Grid,” *Numerical Algorithms*, Vol. 18, No. 3-4, 1998, pp. 209-232. doi:10.1023/A:1019129717644

[10] M. Chen and S. Mehrotra, “Epiconvergent Scenario Generation Method for Stochastic Problems via Sparse Grid,” *Stochastic Programming E-Print Series*, Vol. 2008, No. 7, 2008.

[11] L. C. Evans, “Partial Differential Equations,” *American Mathematical Society*, Vol. 37, No. 3, 1998, pp. 363-367.

[12] G. W. Wasilkowsi and H. Wozniakowski, “Explicit Cost Bounds of Algorithms for Multivariate Tensor Product Problems,” *Journal of Complexity*, Vol. 11, No. 1, 1995, pp. 1-56. doi:10.1006/jcom.1995.1001

[13] H. Brass and G. H¨ammerlin, Eds., “Bounds for Peano kernels,” Vol. 112, Birkhäuser, Basel, 1993, pp. 39-55.

[14] H. Wozniakowski, “Information-Based Complexity,” *Annual Review of Computer Science*, Vol. 1, No. 1, 1986, pp. 319-380. doi:10.1146/annurev.cs.01.060186.001535

[15] C. Roos, T. Terlaky and J.-P. Vial, “Interior Point Methods for Linear Optimization,” Springer, New York, 1997.

[16] N. Megiddo, “Progress in Mathematical Programming, Chapter Pathways to the Optimal Set in Linear Programming,” Springer-Verlag, New York, 1989, p. 132.

[17] A. V. Fiacco, “Introduction to Sensitivity and Stability Analysis in Nonlinear Programming,” Academic Press, New York, 1983.

[18] O. Güler, D. den Hertog, C. Roos and T. Terlaky, “Degeneracy in Interior Point Methods for Linear Programming: A Survey,” *Annals of Operations Research*, Vol. 46-47, No. 1, 1993, pp. 107-138. doi:10.1007/BF02096259

[19] Y. Nesterov and A. Nemirovskii, “Interior Point Polynomial Algorithms in Convex Programming,” Society for Industrial and Applied Mathematics, Philadelphia, 1994.

[20] J. Renegar, “A Mathematical View of Interior-Point Methods in Convex Optimization,” Society for Industrial and Applied Mathematics, Philadelphia, 2001.

[21] S. J. Wright, “Primal-Dual Interior-Point Methods,” Society for Industrial and Applied Mathematics, Philadelphia, 1997.

[22] W. Römisch, “An Approximation Method in Stochastic Optimal Control,” In: *Optimization Techniques, Part* 1, *Lecture Notes in Control and Information Sciences*, Springer-Verlag, New York, 1980, pp. 169-178.

[23] H. Attouch, “Variational Convergence for Functions and Operators,” Pitman (Advanced Publishing Programs), 1984.

[24] H. Attouch and R. J.-B. Wets, “Quantitative Stability of Variational Systems: I. The Epigraphical Distance,” *Transactions of the American Mathematical Society*, Vol. 328, No. 2, 1991, pp. 695-729. doi:10.2307/2001800

[25] P. Kall, “Approximation to Optimization Problems: An Elementary Review,” Mathematics of Operations Research, Vol. 11, No. 1, 1998, pp. 9-18. doi:10.1287/moor.11.1.9

[26] T.-W. Ma, “Higher Chain Formula Proved by Combinatorics,” *The Electronic Journal of Combinatorics*, Vol. 16, No. 21, 2009.

Appendix A. Inverse Transformation and Truncation

For a random vector on Ξ with invertible cumulative distribution function $F^{-1}:[0,1]^d \to \Xi$, the integration domain of a mollifier function can be transformed from Ξ to $[0,1]^d$:

$$\int_\Xi \rho_{\mu,x}(\xi)F(\mathrm{d}\xi)=\int_{[0,1]^d}\rho_{\mu,x}\left(F^{-1}(\omega)\right)\mathrm{d}\omega,$$

then we apply the sparse grid method to generate scenarios and weights for on the cube $[0,1]^d$. We need to check the properties of the integrand function $\rho_{\mu,x}\circ F^{-1}(\cdot)$. Its differentiability only depends on $F^{-1}(\cdot)$ since $\rho_{\mu,x}\in C^\infty(\cdot)$. Most commonly used invertible cumulative distribution functions, for example, inverse of normal distribution function $\phi^{-1}(\cdot)$, is also in C^∞. The finiteness of the partial derivatives of $\rho_{\mu,x}\circ F^{-1}(\cdot)$ also only depends on $F^{-1}(\cdot)$ since partial derivatives of $\rho_{\mu,x}(\cdot)$ are finite for any multi-index $\alpha\le 1$ component-wisely.

The higher order partial derivative of a composite function can be calculated explicitly. For $h:x\in X\subset\mathbb{R}\xrightarrow{f}y\in Y\subset\mathbb{R}\xrightarrow{g}z\in\mathbb{R}$, we can apply the Faà di Bruno's formula:

$$\frac{\mathrm{d}^n}{\mathrm{d}x^n}f\left(g(x)\right)=\left(f\circ g\right)^{(n)}(x)=\sum_{P\in\mathbb{P}_n}f^{(|P|)}\left(g(x)\right)\prod_{B\in P}g^{(|B|)}(x),$$

where $\mathbb{P}_n$ is the set of all partitions of the set J_n of integers $1,\cdots,n$. A partition of J_n is a family of pairwise disjoint nonempty subsets of J_n whose union is J_n. $|A|$ means the cardinality of the set A. For a vector composite function:

$$h:x\in X\subset\mathbb{R}^\nu\xrightarrow{f}y\in Y\subset\mathbb{R}^\gamma\xrightarrow{g}z\in\mathbb{R},$$

We apply Tsoy-Wo Ma's higher chain formula [26]:

$$\frac{\partial^{|\alpha|}z}{\partial x^\alpha}=\alpha!\sum_{(s,p,m)\in\mathcal{D}}\frac{\partial^{|m|}z}{\partial y^m}\cdot\prod_{k=1}^{s}\frac{1}{m_k!}\left[\frac{1}{p_k!}\frac{\partial^{|p_k|}y}{\partial x^{p_k}}\right]^{m_k}$$

where $\mathcal{D}$ is the set of all *decompositions* of *multi-index* α with *multiplicities* m. *Calculation* involving a multi-index $\alpha=(\alpha_1,\cdots,\alpha_\nu)\in\mathbb{N}^\nu$ follows rules:

$$|\alpha|=\sum_{j=1}^{\nu}\alpha_j,\alpha!=\prod_{j=1}^{\nu}\alpha_j!,$$

$$x^\alpha=\prod_{j=1}^{\nu}x_j^{\alpha_j},\frac{\partial^{|\alpha|}z}{\partial x^\alpha}=\prod_{j=1}^{\nu}\frac{\partial^{\alpha_j}}{\partial x_j}z.$$

A multi-index $\alpha\in\mathbb{R}^\nu$ decomposes into s parts $p_1,\cdots,p_s$ in $\mathbb{N}^\nu$ with multiplicities $m_1,\cdots,m_s$ in $\mathbb{N}^\gamma$ respectively if the *decomposition equation*

$$\alpha=|m_1|\,p_1+|m_2|\,p_2+\cdots+|m_s|\,p_s$$

holds and all parts are different. The total multiplicity is defined as

$$m=m_1+m_2+\cdots+m_s.$$

The list (s,p,m) is called a γ-decomposition of α. To ensure all parts are different we may impose $0\ll p_1\ll p_2\ll\cdots\ll p_s$, where $\alpha\in\mathbb{R}^\nu\ll\beta\in\mathbb{R}^\nu$ means $\alpha_1=\beta_j,\cdots,\alpha_{j-1}=\beta_{j-1}$, but $\alpha_j<\beta_j$, for a $j\le\nu$.

For the problem under discussion, let $\alpha\le 1$ component-wisely, *i.e.*, $r=1$, note that are 2^d-1 number of such α (s). Following the higher chain rule formula, we get

$$\begin{aligned}&\frac{\partial^{|\alpha|}}{\partial\omega^\alpha}\rho_{\mu,x}\circ F^{-1}(\omega)\\&=\sum_{(s,p,m)\in\mathcal{D}}\frac{\partial^{|m|}}{\partial\xi^m}\rho_{\mu,x}(\xi)\prod_{k=1}^{s}\frac{1}{m_k!}\left[\frac{1}{p_k!}\frac{\partial^{p_k}\xi}{\partial\omega^{p_k}!}\right]^{m_k}.\end{aligned}$$

Furthermore, since $\lim_{\mu\to0^+}\nabla^2\rho_{\mu,x}(\xi)=0$ by Proposition 2.3, the computation of $\lim_{\mu\to0^+}\frac{\partial^\alpha}{\partial\omega^\alpha}\rho_{\mu,x}\circ F^{-1}(\omega)$ can be simplified significantly. In this case, only the decompositions $(1,\alpha,e_i)$, e_i is the ith unit basis of $\mathbb{R}^d$, correspond to non-zeros in the above formula. Otherwise, for $s\ge2$, $m=m_1+\cdots+m_s$, and $\lim_{\mu\to0^+}\frac{\partial^{|m|}}{\partial\xi^m}\rho_{\mu,x}(\xi)=0$.

Hence

$$\begin{aligned}\lim_{\mu\to0^+}\frac{\partial^\alpha}{\partial\omega^\alpha}\rho_{\mu,x}\circ F^{-1}(\omega)&=\lim_{\mu\to0^+}\sum_{i=1}^{d}\alpha_i\frac{\partial\rho_{\mu,x}(\xi)}{\partial\xi_i}\frac{\partial^{|\alpha|}\xi_i}{\partial w^\alpha}\\&=-\sum_{i=1}^{d}\alpha_i u_{i,x}^*\frac{\partial^{|\alpha|}\xi_i}{\partial w^\alpha}.\end{aligned}$$

Hence,

$$\lim_{\mu\to0^+}\left\|\rho_{\mu,x}\circ F^{-1}(\omega)\right\|_{\mathcal{W}_d^1}=\left\|\left(\sum_{i=1}^{d}\alpha_i\left\|u_{i,x}^*\frac{\partial^{|\alpha|}\xi_i}{\partial\omega^\alpha}\right\|_{\alpha_2}\right)_{\alpha\le1}\right\|_2<\infty,$$

if and only if $\frac{\partial^{|\alpha|}\xi_i}{\partial\omega^\alpha}<\infty$ almost surely $\forall\alpha\le1,\forall i=1,\cdots,d$. For some distributions, the condition might not hold. For example, the inverse of a cumulative function of the normal distribution does not have this property nearby 0 or 1. To remove the singularities, truncation of the cube [0, 1] could be applied:

$$\int_{[0,1]^d}h(x)\mathrm{d}x\approx\int_{[\varepsilon,1-\varepsilon]^\mathrm{d}}h(x)\,\mathrm{d}x,$$

where $0<\varepsilon<1$ is a small positive number. To compute the righthand side using the standard sparse grid method, we need to change the variable to y, where

$x=\varepsilon+(1-2\varepsilon)y:[0,1]^d \to [\varepsilon,1-\varepsilon]^d$. Hence

$$\int_{[\varepsilon,1-\varepsilon]^d} h(x)\,\mathrm{d}x=(1-2\varepsilon)^d \int_{[0,1]^d} h(\varepsilon+(1-2\varepsilon)y)\,\mathrm{d}y.$$

Hence for a two-stage linear problem with an invertible but unbounded cumulative distribution function F^{-1}, we shall first generate the standard grid points and weights $\left\{(\omega^k,w^k)\right\}_{k=1}^K$ using the sparse grid method, then scale and transform them to the original random variable ξ by

$$\xi^k=F^{-1}\left(\varepsilon+(1-2\varepsilon)\omega^k\right),\quad \tilde{w}^k=(1-2\varepsilon)^d w^k,$$

and finally use the $\left\{\xi^k,\tilde{w}^k\right\}$ in the approximation model (2).

The error of this approximation model is exactly the sum of truncation error e_t, and sparse grid approximation error e_s. Error e_t goes down with ε and error e_s goes down with increasing K at the first order rate of sparse grid method.