Probabilistic Constrained Optimization: Methodology and Applications (S. P. Uryasev, Editor), pp. 1-25

Introduction to the Theory of Probabilistic Functions and Percentiles (Value-at-Risk)

S. Uryasev (uryasev@ise.ufl.edu)
University of Florida
474 Weil Hall, Gainesville, FL 32611-6595

Abstract

Probabilistic and quantile (percentile) functions are commonly used for the analysis of models with uncertainties or variabilities in parameters. In financial applications, the percentile of the losses is called Value-at-Risk (VaR). VaR, a widely used performance measure, answers the question: what is the maximum loss with a specified confidence level? Percentiles are also used for defining other relevant performance measures, such as Conditional Value-at-Risk (CVaR). CVaR (also called Mean Excess Loss, Mean Shortfall, or Tail VaR) is the average loss for the worst x% scenarios (e.g., 5%). CVaR risk measure has more attractive properties compared to VaR. This introductory paper gives basic definitions and reviews several topics:

- sensitivities of probabilistic functions;
- sensitivities of percentiles (VaR);
- optimization approaches for CVaR.

The emphasis of this paper is on issues which have been relatively recently developed.

1 Introduction

Probabilistic and quantile (percentile) functions are commonly used for the analysis of models with uncertainties or variabilities in parameters. For instance, in risk and reliability analysis, performance functions, characterizing operation of systems, are

formulated as probabilities of successful or unsuccessful accomplishment of their missions, e.g., core damage probability or frequency of a nuclear power plant, probability of successful landing of an aircraft, probability of profitable transactions in a stock market, or percentiles of the risks in public risk assessments. In financial applications, the percentile of the losses is called Value-at-Risk (VaR). VaR, a widely used performance measure, answers the question: what is the maximum loss with a specified confidence level? Percentiles are also used for defining other relevant performance measures, such as Conditional Value-at-Risk (CVaR). CVaR (also called Mean Excess Loss, Mean Shortfall, or Tail VaR) is the average loss for the worst x% scenarios (e.g., 5%). CVaR risk measure has more attractive properties compared to VaR.

This introductory paper gives basic definitions and reviews several topics:

- sensitivities of probabilistic functions;
- sensitivities of percentiles (VaR);
- optimization approaches for CVaR.

This paper is not intended as a comprehensive review. It covers only several recently developed topics and does not include many other important issues related to the analysis of the probabilistic and percentile functions. For more comprehensive discussions and analyses of probabilistic and quantile functions, see books by Prékopa [32], Kan and Kibzun [14], Pflug [29], Kall and Wallace [13], and Birge and Louveaux [3]. Also, many relevant topics and applications are subject of papers included in this volume.

2 Sensitivity Analysis of Probabilistic Functions

2.1 Background

The sensitivity analysis of probabilistic performance functions involves an evaluation of their derivatives with respect to (w.r.t.) parameters. Derivatives of the probability and quantile functions are used as well to solve stochastic optimization problems (e.g., [8, 9, 13, 48]), and to analyze the Discrete Event Dynamic Systems (DES) (e.g., [10, 11, 43]).

This chapter overviews basic results on differentiability of the probabilistic functions without going into details of formal conditions and theorems. More formal and comprehensive discussions of these topics can be found in the review paper by Kibzun and Uryasev [18]. Here, we only formulate statements and illustrate the results with examples. A probability function can be formally represented as an expectation of a discontinuous indicator function of a set, or as an integral over a domain depending upon parameters. Nevertheless, differentiability conditions of the probability function

do not follow from similar conditions of the expectations of continuous (smooth or convex) functions. A differentiation formula for an expectation of continuous functions can be obtained by interchanging the gradient and expectation operators [7, 12, 40].

The derivative of the probability function has many equivalent representations: it can be represented in general form as an integral over the surface, integral over the volume, or a sum of integrals over the volume and over the surface; it can be calculated using weak derivatives of the probability measures or conditional expectations. Here, we overview only mathematical results related to the first three representations.

The first general result on sensitivity of probability functions in general form was obtained by Raik [36] who represented the gradient of the probability function with one constraint in the form of the surface integral. For volume integrals with the domain depending upon a parameter, Roenko [41] and Simon [44] obtained the gradient formula in the form of the surface integral. Uryasev [50] extended Raik's formula for probability functions with many constraints; Kibzun and Tretyakov [17] extended it to the piece-wise smooth constraint and probability density function(see, also, [14]). Special cases of the probability function with normal and gamma distributions were investigated by Prékopa [31], and Prékopa and Szántai [34]. Pflug [29] represented the gradient of the probability function in the form of an expectation using weak probability measures.

The next general result was published by Uryasev [49], in which the gradient of the probability function was written as a volume integral. Later, he [50] generalized this formula to the case of several constraints, and established relations between formulas for the gradient in the form of the surface and the volume integrals. Using a change of variables, Marti [23, 24] derived the probability function gradient in the form of the volume integral. Rubinstein [42] used a change of variables to differentiate performance functions in analyses of DESs. Marti [23] also suggested approximations of the probability function gradient by an asymptotic expansion of the integrals.

Finally, a general analytical formula for the derivative of probability functions with many constraints was obtained by Uryasev [50]; it calculates the gradient as an integral over the surface, or an integral over the volume, or the sum of integrals over the surface and the volume. The general formula calculates the gradient using the solution of a system of nonlinear equations. Special cases of this formula correspond to the Raik formula [36], Uryasev formula [49], and the change–of–variables approach [24, 42].

2.2 Notations and Definitions

Let $f(x, y)$ be a performance function, for instance losses associated with the decision vector x, to be chosen from a subset X of $\mathbb{R}^n$, and the random vector y in $\mathbb{R}^m$. In finance applications, the vector x can be interpreted as representing a portfolio, with X as the set of available portfolios (subject to various constraints), but other

interpretations could be made as well. The vector y stands for the uncertainties, e.g. market parameters, that can affect the performance.

For each x, the performance function $f(x, y)$ is a random variable having a distribution in $\mathbb{R}$ induced by that of y. The underlying probability distribution of y in $\mathbb{R}^m$ will be assumed to have density, which we denote by $p(x, y)$. The density may depend upon the decision vector x. In finance applications, a two step procedure (see, for instance, RiskMetrics [37]) can be used to derive an analytical expression for $p(y)$ or construct a Monte Carlo simulation code for drawing samples from $p(x, y)$: (1) modeling of risk factors in $\mathbb{R}^{m_1}$,(with $m_1 < m$), (2) based on the characteristics of instrument i, $i = 1, \ldots, n$, the distribution $p(x, y)$ can be derived or code transforming random samples of risk factors to the random samples from density $p(x, y)$ can be constructed.

Let an integral over the volume

$$F(x) = \int\limits_{f(x,y) \leq 0} p(x, y)\, dy \tag{1}$$

is defined on the Euclidean space $\mathbb{R}^n$, where $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^k$ and $p : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ are some functions. The inequality $f(x, y) \leq 0$ in the integral is a system of inequalities

$$f_i(x, y) \leq 0 \ , \quad i = 1, \ldots, k \ .$$

Both the kernel function $p(x, y)$ and the function $f(x, y)$ defining the integration set depend upon the parameter x. For example, let

$$F(x) = P\{f(x, \zeta(\omega)) \leq 0\} \tag{2}$$

be a *probability function*, where $\zeta(\omega)$ is a random vector in $\mathbb{R}^m$. The random vector $\zeta(\omega)$ is assumed to have a probability density $p(x, y)$ that depends on a parameter $x \in \mathbb{R}^n$. The probability function can be represented as an expectation of an indicator function, which equals one on the integration set, and equals zero outside of it. For example, let

$$F(x) \ = \ \mathbb{E}\big[I_{\{f(x,\boldsymbol{\zeta}) \leq 0\}}\, g(x, \boldsymbol{\zeta})\big] \ = \int\limits_{f(x,y) \leq 0} g(x, y)\, \rho(x, y)\, dy \ = \int\limits_{f(x,y) \leq 0} p(x, y)\, dy \ , \tag{3}$$

where $I_{\{\}}$ is the indicator function, and random vector $\boldsymbol{\zeta}$ in $\mathbb{R}^m$ has a probability density $\rho(x, y)$ that depends on the vector $x \in \mathbb{R}^n$.

2.3 Integral Over the Surface Formula.

The following formula calculates the gradient of integral (1) over the set given by nonlinear inequalities as the sum of the integral over the volume plus the integral

over the surface of the integration set. We call this *the integral over the surface formula* because if density $p(x,y)$ does not depend upon x the gradient of integral (1) equals the integral over the surface. This formula, for the case of one inequality, was obtained by Raik [36] and generalized for the case with many inequalities by Uryasev [50].

Let us denote by $\mu(x)$ the integration set

$$\mu(x) \;=\; \{\, y \in \mathbb{R}^m \;:\; f(x,y) \le 0 \,\} \;\stackrel{def}{=}\; \{\, y \in \mathbb{R}^m \;:\; f_l(x,y) \le 0\,,\; 1 \le l \le k \,\}\,,$$

and by $\partial\mu(x)$ the surface of this set $\mu(x)$. Also, let us denote by $\partial_i\mu(x)$ a part of the surface which corresponds to the function $f_i(x,y)$, i.e.,

$$\partial_i\mu(x) \;=\; \mu(x) \bigcap \{\, y \in \mathbb{R}^m \;:\; f_i(x,y) = 0 \,\}\,.$$

If the constraint functions are differentiable and the following integral exists, then the gradient of integral (1) equals

$$\nabla_x F(x) \;=\; \int\limits_{\mu(x)} \nabla_x p(x,y)\, dy \;-\; \sum_{i=1}^{k} \int\limits_{\partial_i \mu(x)} \frac{p(x,y)}{||\nabla_y f_i(x,y)||} \nabla_x f_i(x,y)\; dS\,. \qquad (4)$$

A potential disadvantage of this formula is that in a multidimensional space it is difficult to calculate the integral over a nonlinear surface. Standard numerical techniques, such as Monte-Carlo algorithms, are applicable to estimate volume rather than surface integrals. Nevertheless, this formula can be quite useful in various special cases, such as the linear case.

2.3.1 Example 1. Linear Case: Integral Over the Surface Formula [50]

Let $A(\omega)$, be a random $l \times n$ matrix with the joint density $p(A)$. Suppose that $x \in \mathbb{R}^n$ and $x_j \neq 0\,,\; j = 1, \ldots, n$. Let us define

$$F(x) = P\{\, A(\omega)x \le b\,,\; A(\omega) \ge 0 \,\}\,, \quad b = (b_1, \ldots, b_l) \;\in \mathbb{R}^l\,, \quad x \in \mathbb{R}^n\,, \qquad (5)$$

i.e. $F(x)$ is the probability that the linear constraints $A(\omega)x \le b\,,\; A(\omega) \ge 0$ are satisfied. The constraint, $A(\omega) \ge 0$, means that all elements $a_{ij}(\omega)$ of the matrix $A(\omega)$ are non-negativity. Let us denote by A_i and A^i the i-th row and column of the matrix A

$$A = \begin{pmatrix} A_1 \\ \vdots \\ A_l \end{pmatrix} \;=\; \left(A^1, \cdots, A^n\right)\,,$$

then

$$f(x,A) = \begin{pmatrix} f_1(x,A) \\ \vdots \\ f_k(x,A) \end{pmatrix} = \begin{pmatrix} A_1 x - b_1 \\ \vdots \\ A_l x - b_l \\ -A^1 \\ \vdots \\ -A^n \end{pmatrix} ,$$

$$k = l + l \times n \ .$$

The function $F(x)$ equals

$$F(x) = \int\limits_{f(x,A) \le 0} p(A)\, dA \ . \tag{6}$$

We use formula (4) to calculate the gradient $\nabla_x F(x)$ as an integral over the surface. The function $p(A)$ does not depend upon x and $\nabla_x p(A) = 0$. Formula (4) implies that $\nabla_x F(x)$ equals

$$\nabla_x F(x) \ = \ - \sum_{i=1}^{k} \int\limits_{\partial_i \mu(x)} \frac{p(A)}{||\nabla_A f_i(x,A)||} \nabla_x f_i(x,A) \ dS \ .$$

Since $\nabla_x f_i(x,A) = 0$ for $i = l+1, \ldots, k$, the gradient $\nabla_x F(x)$ equals

$$\nabla_x F(x) \ = \ - \sum_{i=1}^{l} \int\limits_{\partial_i \mu(x)} \frac{p(A)}{||\nabla_A f_i(x,A)||} \nabla_x f_i(x,A) \ dS \ = \ - \sum_{i=1}^{l} \int\limits_{\partial_i \mu(x)} \frac{p(A)}{||x||} A_i^T \ dS$$

$$= \ - \ ||x||^{-1} \sum_{i=1}^{l} \int\limits_{\substack{Ax \le b \, , \, A \ge 0 \\ A_i x = b_i}} p(A)\, A_i^T \ dS \ .$$

2.4 Integral Over the Volume Formula.

This section represents the gradient of function (1) in the form of volume integral. Let us introduce the following shorthand notations

$$f_{1l}(x,y) = \begin{pmatrix} f_1(x,y) \\ \vdots \\ f_l(x,y) \end{pmatrix} \ , \quad f(x,y) = f_{1k}(x,y) \ ,$$

$$\nabla_y f(x,y) = \begin{pmatrix} \frac{\partial f_1(x,y)}{\partial y_1} \ , & \cdots & , \ \frac{\partial f_k(x,y)}{\partial y_1} \\ & \vdots & \\ \frac{\partial f_1(x,y)}{\partial y_m} \ , & \cdots & , \ \frac{\partial f_k(x,y)}{\partial y_m} \end{pmatrix} \ .$$

Divergence for the $n \times m$ matrix H consisting of the elements h_{ji} is denoted by

$$div_y H = \begin{pmatrix} \sum\limits_{i=1}^{m} \frac{\partial h_{1i}}{\partial y_i} \\ \vdots \\ \sum\limits_{i=1}^{m} \frac{\partial h_{ni}}{\partial y_i} \end{pmatrix} .$$

Following [50], the derivative of function (1) is represented as the integral over the volume

$$\nabla_x F(x) = \int\limits_{\mu(x)} \nabla_x p(x,y)\, dy + \int\limits_{\mu(x)} div_y\big(p(x,y)H(x,y)\big)\, dy\,, \tag{7}$$

where a matrix function $H : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^{n \times m}$ satisfies the equation

$$H(x,y)\, \nabla_y f(x,y) + \nabla_x f(x,y) = 0\ . \tag{8}$$

The last system of equations may have many solutions. Therefore, formula (7) provides a number of equivalent expressions for the gradient. The following section gives analytical solutions of this system of equations. In some cases, this system does not have any solution, and formula (7) is not valid. The following section deals with such cases and provides a general formula where system of equations can be solved only for some of the functions defining the integration set.

2.4.1 Example 2. Linear Case: Integral Over the Volume Formula [50]

With formula (7), the gradient of probability function (5) with linear constrains considered in Example 1 can be represented as the integral over the volume. It can be shown that equation (8) does not have a solution in this case. Nevertheless, we can slightly modify the constraints, such that the integration set is not changed and equation (8) has a solution. In the vector function $f(x, A)$, we multiply column A^i by x^i if x^i is positive or multiply it by $-x^i$ if x^i is negative. Therefore, we have the following constraint function

$$f(x,A) = \begin{pmatrix} A_1 x - b_1 \\ \vdots \\ A_l x - b_l \\ -(+)x_1 A^1 \\ \vdots \\ -(+)x_n A^n \end{pmatrix} , \tag{9}$$

where $-(+)$ means that we take an appropriate sign. It can be directly checked that, the following matrix

$$H^*(x,A) = \big(h^1(x,A_1), \ldots, h^l(x,A_l)\big)\ , \quad h^i(x,A_i) = -\begin{pmatrix} a_{i1}x_1^{-1} & & 0 \\ & \ddots & \\ 0 & & a_{in}x_n^{-1} \end{pmatrix}$$

is a solution of system (8). As it will be shown in the next section, this analytical solution follows from the fact that the change of variables $Y^i = x_i A^i$, $i = 1, \ldots, n$, eliminates variables x^i , $i = 1, \ldots, n$, from constraints (9).

Since $\nabla_x p(A) \ = \ 0$ and $div_A \Big(p(A) H^*(x, A) \Big)$ equals

$$div_A \Big(p(A) H^*(x, A) \Big) \ = \ - \left(\begin{array}{c} x_1^{-1} \Big(l\, p(A) \ + \ \sum_{i=1}^{l} a_{i1} \frac{\partial}{\partial a_{i1}} p(A) \Big) \\ \vdots \\ x_n^{-1} \Big(l\, p(A) \ + \ \sum_{i=1}^{l} a_{in} \frac{\partial}{\partial a_{in}} p(A) \Big) \end{array} \right)$$

formula (7) implies

$$\frac{\partial F(x)}{\partial x_j} \ = \ -x_j^{-1} \int\limits_{\substack{Ax \le b \\ A \ge 0}} \left(l\, p(A) \ + \ \sum_{i=1}^{l} a_{ij} \frac{\partial}{\partial a_{ij}} p(A) \right) \, dA \ . \tag{10}$$

2.5 General Formula.

Further, we give the general formula [49, 50] for the derivative of integral (1). The gradient of the integral is represented as a sum of integrals taken over the volume and over the surface. This formula can be used when system of equations (8) does not have a solution. We split the set of constraints $K \stackrel{def}{=} \{1, \ldots, k\}$ into two subsets K_1 and K_2. Without loss of generality we suppose that

$$K_1 \ = \ \{1, \ldots, l\} \ , \quad K_2 \ = \ \{l+1, \ldots, k\} \ .$$

The derivative of integral (1) can be represented as the sum of the volume and surface integrals

$$\nabla_x F(x) \ = \ \int\limits_{\mu(x)} \nabla_x p(x, y) \, dy + \int\limits_{\mu(x)} div_y \Big(p(x, y) H_l(x, y) \Big) \, dy$$

$$- \ \sum_{i=l+1}^{k} \int\limits_{\partial_i \mu(x)} \frac{p(x, y)}{||\nabla_y f_i(x, y)||} \Big[\nabla_x f_i(x, y) \ + \ H_l(x, y) \, \nabla_y f_i(x, y) \Big] \, dS \ , \tag{11}$$

where the matrix $H_l : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^{n \times m}$ satisfies the equation

$$H_l(x, y) \, \nabla_y f_{1l}(x, y) \ + \ \nabla_x f_{1l}(x, y) \ = \ 0 \ . \tag{12}$$

The last equation can have a lot of solutions and we can choose an arbitrary one, differentiable w.r.t. the variable y .

The general formula contains as a special case the integral over the surface formula (4) and integral over the volume formula (7). When the set K_1 is empty, the matrix

H_l is absent and the general formula is reduced to the integral over the surface. Also, when the set K_2 is empty, we have the integral over the volume formula (7). In addition to these extreme cases, the general formula provides a number of intermediate expressions for the gradient in the form of the sum of the integral over the surface and the integral over the volume. There are many equivalent representations of the gradient corresponding to the various sets K_1 and K_2 and solutions of equation (12).

Equation (12) (and equation (8) which is a partial case of equation (12)) can be solved explicitly. Usually, this equation has many solutions. For instance, the matrix

$$-\nabla_x f_{1l}(x,y)\Big(\nabla_y^T f_{1l}(x,y)\times\nabla_y f_{1l}(x,y)\Big)^{-1}\nabla_y^T f_{1l}(x,y) \tag{13}$$

is a solution of equation (12). Also, in a number of cases, equation (12) can be solved using a change of variables. Suppose that there is a change of variables

$$y=\gamma(x,z)$$

which eliminates the vector x from the function $f(x,y)$, defining the integration set, i.e., the function $f(x,\gamma(x,z))$ does not depend upon the variable x. Denote by $\gamma^{-1}(x,y)$ the inverse function, defined by the equation

$$\gamma^{-1}\Big(x,\gamma(x,z)\Big) \;=\; z\ .$$

Let us show that the following matrix

$$H(x,y) \;=\; \nabla_x\gamma(x,z)|_{z=\gamma^{-1}(x,y)} \tag{14}$$

is a solution of (12). Indeed, the gradient of the function $\gamma(x,y(x,z))$ w.r.t. x equals zero; therefore

$$0 \;=\; \nabla_x f_{1l}(x,\gamma(x,z)) = \nabla_x\gamma(x,z)\nabla_y f_{1l}(x,y)|_{y=\gamma(x,z)} \;+\; \nabla_x f_{1l}(x,y)|_{y=\gamma(x,z)}\ ,$$

and the function $\nabla_x\gamma(x,z)|_{z=\gamma^{-1}(x,y)}$ is a solution of equation (12).

Formula (7) with matrix (14) gives the derivative formulas, which can be obtained with change of variables in the integration set [24].

2.5.1 Example 3.

While investigating the operational strategies for inspected components (see [33]), the following integral was considered

$$F(x) = \int\limits_{\substack{b(y)\le x\ ,\\ y_i\ge\theta\ ,\ i=1,\dots,m}} p(y)\ dy\ , \tag{15}$$

where $x \in \mathbb{R}^1, y \in \mathbb{R}^m, p : \mathbb{R}^m \to \mathbb{R}^1$, $\theta > 0$, $b(y) = \sum\limits_{i=1}^{m} y_i^{\alpha}$. In this case,

$$f(x,y) = \begin{pmatrix} b(y) - x \\ \theta - y_1 \\ \vdots \\ \theta - y_m \end{pmatrix} ,$$

and

$$F(x) = \int\limits_{f(x,y) \le 0} p(y)\, dy \;=\; \int\limits_{\mu(x)} p(y)\, dy \; .$$

Let us consider that $l = 1$, i.e. $K_1 \;=\; \{1\}$ and $K_2 \;=\; \{2, \ldots, m+1\}$. The gradient $\nabla_x F(x)$ equals

$$\nabla_x F(x) \;=\; \int\limits_{\mu(x)} \Big[\, \nabla_x p(y) \;+\; div_y\Big(p(y) H_1(x,y)\Big)\,\Big]\, dy$$

$$- \sum_{i=2}^{m+1} \int\limits_{\partial_i \mu(x)} \frac{p(y)}{||\nabla_y f_i(x,y)||} \Big[\, \nabla_x f_i(x,y) \;+\; H_1(x,y)\, \nabla_y f_i(x,y)\,\Big]\; dS \; , \qquad (16)$$

where the matrix $H_1(x,y)$ satisfies equation (12). In view of

$$\nabla_y f_1(x,y) = \alpha \begin{pmatrix} y_1^{\alpha-1} \\ \vdots \\ y_m^{\alpha-1} \end{pmatrix} , \quad \nabla_x f_1(x,y) = -1 \; ,$$

a solution $H_1^*(x,y)$ of equation (12) equals

$$H_1^*(x,y) \;=\; h(y) \;\stackrel{def}{=}\; \Big(h_1(y_1), \ldots, h_m(y_m)\Big) \;=\; \frac{1}{\alpha m} \Big(y_1^{1-\alpha}, \ldots, y_m^{1-\alpha}\Big) \; . \qquad (17)$$

Let us denote

$$(\theta_i \mid y) \;=\; (y_1, \ldots, y_{i-1}, \theta, y_{i+1}, \ldots, y_m) \; ,$$

$$y^{-i} \;=\; (y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_m) \; , \quad b(\theta_i \mid y) \;=\; \theta^{\alpha} \;+\; \sum_{\substack{j=1 \\ j \ne i}}^{m} y_j^{\alpha} \; .$$

Also, let us denote by $y^{-i} \ge \theta$ the set of inequalities

$$y_j \ge \theta \; , \quad j = 1, \ldots, i-1, i+1, \ldots, m \; .$$

The sets $\partial_i \mu(x) \; , \; i = 2, \ldots, m+1$ have a simple structure

$$\partial_i \mu(x) \;=\; \mu(x) \bigcap \{\, y \in \mathbb{R}^m \;:\; y_i = \theta \,\} \;=\; \{ y^{-i} \in \mathbb{R}^{m-1} \;:\; b(\theta_i | y) \le x \; , \; y^{-i} \ge 0 \,\} \; .$$

For $i = 2, \ldots, m+1$, we have

$$(\nabla_y f_i(y))_j = 0,\ j = 1, \ldots, m,\ j \neq i-1, \tag{18}$$

$$(\nabla_y f_i(y))_{i-1} = -1\,,\ ||\nabla_y f_i(y)|| = 1. \tag{19}$$

The function $p(y)$ and the functions $f_i(y)\,,\ i = 2, \ldots, m+1$ do not depend on x, consequently,

$$\nabla_x p(y) = 0\,, \tag{20}$$

$$\nabla_x f_i(y) = 0\,, \quad i = 2, \ldots, m+1\,. \tag{21}$$

Equations (16) - (21) imply

$$\nabla_x F(x) \;=\; \int\limits_{\mu(x)} div_y\Big(p(y)h(y)\Big)\,dy \;-\; \sum_{i=2}^{m+1} \int\limits_{\partial_i \mu(x)} \frac{p(y)}{||\nabla_y f_i(y)||}\, h(y) \nabla_y f_i(y)\, dS$$

$$= \int\limits_{\mu(x)} div_y\Big(p(y)h(y)\Big)\,dy \;+\; \sum_{i=2}^{m+1} h_{i-1}(\theta) \int\limits_{\partial_i \mu(x)} p(y)\, dS$$

$$= \int\limits_{\substack{b(y)\le x\,,\\ y_i \ge \theta\,,\, i=1,\ldots,m}} div_y\Big(p(y)h(y)\Big)\,dy \;+\; \sum_{i=1}^{m} \frac{\theta^{1-\alpha}}{\alpha m} \int\limits_{\substack{b(\theta_i|y)\le x\,,\\ y^{-i} \ge \theta}} p(\theta_i \mid y)\, dy^{-i}\,.$$

Since

$$div_y\Big(p(y)h(y)\Big) \;=\; h(y)\nabla_y p(y) + p(y) div_y h(y) = \;\frac{1}{\alpha m} \sum_{i=1}^{m} \frac{\partial p(y)}{\partial y_i} y_i^{1-\alpha} + p(y) \frac{1-\alpha}{\alpha m} \sum_{i=1}^{m} y_i^{-\alpha}\,,$$

we, finally, obtain that the gradient $\nabla_x F(x)$ equals

$$\nabla_x F(x) \;=\; \int\limits_{\substack{b(y)\le x\,,\\ y_i \ge \theta,\, i=1,\ldots,m}} \sum_{i=1}^{m} \frac{y_i^{-\alpha}}{\alpha m} \Big[y_i \frac{\partial p(y)}{\partial y_i} + (1-\alpha) p(y)\Big] dy + \frac{\theta^{1-\alpha}}{\alpha m} \sum_{i=1}^{m} \int\limits_{\substack{b(\theta_i|y)\le x\,,\\ y^{-i} \ge \theta}} p(\theta_i \mid y)\, dy^{-i}\,.$$

The formula for $\nabla_x F(x)$ is valid for an arbitrary sufficiently smooth function $p(y)$.

3 The Gradient of the Quantile (Value-at-Risk)

The quantile or Value-at-Risk (VaR) answers the question: what is the maximum loss (or other measure of the performance) with a specified confidence level? A description of various methodologies for the modeling of VaR can be seen, along with related resources, at www.gloriamundi.org. Mostly, approaches to calculating VaR rely on linear approximation of the risk factors and assume a joint normal (or log-normal) distribution of the underlying parameters, see, for instance, Duffie and Pan

[6], Pritsker [35], RiskMetrics [37], Simons [45], Stambaugh [46]). In financial applications, historical or Monte Carlo simulation-based tools are used when the portfolio contains nonlinear instruments such as options ([4, 26, 35, 37, 46]). Discussions of optimization problems involving VaR can be found in papers by Litterman [20, 21], Kast et al. [15], Lucas and Klaassen [22].

The gradient of the quantile function was obtained by Kibzun et al. [16]. The approach, considered in this section, for simultaneous calculating of quantiles and their gradients with Monte Carlo simulation algorithms was proposed in [53].

The probability of the performance function $f(x, y)$ not exceeding a threshold α is denoted by

$$\Psi(x, \alpha) \;=\; \int\limits_{f(x,y)\le\alpha} p(y)\, dy\,. \tag{22}$$

As a function of α for fixed x, $\Psi(x, \alpha)$ is the cumulative distribution function for the performance associated with x. It completely determines the behavior of this random variable and is fundamental in defining quantile (VaR). In general, $\Psi(x, \alpha)$ is nondecreasing w.r.t. α and continuous from the right, but not necessarily from the left because of the possibility of jumps. We assume however in what follows that the probability distributions are such that no jumps occur, or in other words, that $\Psi(x, \alpha)$ is everywhere continuous w.r.t. α. The required continuity follows from properties of the function $f(x, y)$ and the density $p(y)$. As it follows from derivative formula (4), the function $\Psi(x, \alpha)$ is differentiable in α if the function $f(x, y)$ is smooth, $||\nabla_y f_i(x, y)|| > 0$, and the appropriate surface integral exists.

The β-quantile (or β-VaR) for the performance random variable associated with x and any specified probability level β in $(0, 1)$ will be denoted by $\alpha_\beta(x)$. In this setting, the quantile is given by

$$\alpha_\beta(x) \;=\; \min\{\, \alpha \in \mathbb{R} :\; \Psi(x, \alpha) \ge \beta \,\}\ .$$

So, the quantile $\alpha_\beta(x)$ is the left endpoint of the nonempty interval consisting of the values α such that actually $\Psi(x, \alpha) = \beta$. This follows from $\Psi(x, \alpha)$ being continuous and nondecreasing w.r.t. α. The interval might contain more than a single point if Ψ has "flat spots."

Further, we suppose that for considered x, and β

$$\left.\frac{\partial\Psi(x, \alpha)}{\partial\alpha}\right|_{\alpha=\alpha(x,\beta)} \neq 0\ .$$

With this assumption, the quantile $\alpha(x, \beta)$ is a unique solution of the following equation

$$\Psi(x, \alpha) = \beta \tag{23}$$

w.r.t. α . The gradient of the quantile function $\alpha(x,\beta)$ in x can be obtained by differentiating the left and right parts of equation (23), i.e.,

$$\nabla_x \Psi(x,\alpha(x,\beta)) \;=\; \nabla_x \Psi(x,\alpha)\mid_{\alpha=\alpha(x,\beta)} + \nabla_\alpha \Psi(x,\alpha)\mid_{\alpha=\alpha(x,\beta)} \nabla_x \alpha(x,\beta) \;=\; 0\ .$$

Consequently,

$$\nabla_x \alpha(x,\beta) = -\frac{\nabla_x \Psi(x,\alpha)\mid_{\alpha=\alpha(x,\beta)}}{\nabla_\alpha \Psi(x,\alpha)\mid_{\alpha=\alpha(x,\beta)}}\ . \tag{24}$$

Analogous, the derivative of quantile function $\alpha(x,\beta)$ in the parameter β can be obtained by differentiating equation (23) w.r.t. β, i.e.,

$$\nabla_\beta \Psi(x,\alpha(x,\beta)) \;=\; \nabla_\alpha \Psi(x,\alpha)\mid_{\alpha=\alpha(x,\beta)} \nabla_\beta \alpha(x,\beta) \;=\; 1\ ,$$

and

$$\nabla_\beta \alpha(x,\beta)) \;=\; (\nabla_\alpha \Psi(x,\alpha)\mid_{\alpha=\alpha(x,\beta)})^{-1}\ . \tag{25}$$

Formulas (24) and (25) involve derivatives of the probability function $\Psi(x,\alpha)$ w.r.t. x and α. These derivatives can be calculated using formulas of the previous section. We can apply various formulas to the denominator and the numerator in (24) and (25). For example, the gradient of the quantile function can be represented as the ratio of surface or volume integrals. Usually, we prefer to work with volume integrals because they can be evaluated with standard numerical methods or Monte Carlo type techniques. However, surface integrals are convenient for low dimension cases (for instance, if $x \in \mathbb{R}^1$, the surface is reduced to a few points), or when the constraint function $f(x,y)$ is linear w.r.t. y .

We finalize this section with an example showing how formulas (24) and (25) can be used to evaluate derivatives of quantile functions with Monte Carlo simulation algorithms. In this case, the derivative of the probability function is calculated using integrals over the volume.

The probability function $\Psi(x,\alpha)$, given by formula (22), is an integral over the m-dimensional volume. Suppose that with Monte Carlo simulations we evaluated the quantile $\alpha_\beta(x)$. Using formulas (7) or (11), derivatives of probability function $\Psi(x,\alpha)$ w.r.t. x and α can be represented as integrals over the same volume with the same density $p(x)$ (for instance, such representation for linear functions gives formula (10)) i.e.,

$$\nabla_x \Psi(x,\alpha) \;=\; \int\limits_{f(x,y)\le\alpha} a(x,\alpha,y)\, p(y)\, dy\ , \qquad \nabla_\alpha \Psi(x,\alpha) \;=\; \int\limits_{f(x,y)\le\alpha} b(x,\alpha,y)\, p(y)\, dy\ . \tag{26}$$

The m-dimensional kernel function $a(x,\alpha,y)$ in the first integral and the kernel function $b(x,\alpha,y)$ in the second integral can be evaluated during the same simulation run. So, using the same random samples $y(\omega)$ from density $p(y)$ we can evaluate simultaneously gradients of the probability function with formulas (26), and gradients of the quantile function with formulas (24) and (25).

Example 4. The following percentile function was studied in evaluating incremental lifetime cancer risk for children due to ingesting benzene with soil [47]. The gradient for this percentile function is calculated in [53]. The constraint function $f(x,y)$ is the ratio of two random variables

$$f(x,y) \triangleq \frac{1}{x_3 + x_4 y_2} e^{x_1 + x_2 y_1} \ , \tag{27}$$

where $x \in \mathbb{R}^4$, $y \in \mathbb{R}^2$, and y_1, y_2 are independent random variables having the same normal distribution $N(0,1)$. We denote densities for these random variables by $p_1(y_1)$ and $p_2(y_2)$, respectively. The numerator of the constraint function is the lognormally distributed random variable with parameters (x_1, x_2). The denominator of the constraint function is the normally distributed random variable with parameters $N(x_3, x_4)$. Thus, the probability function $\Psi(x,\alpha)$ is a two-dimensional integral with the constraint $f(x,y) \le \alpha$ and the density function $p(y) = p_1(y_1)\, p_2(y_2)$. The change of variables

$$(y_1, y_2)^T = \gamma(x,\alpha,y) \stackrel{def}{=} \Big((y_1 - x_1)/x_2,\ -(\alpha\, x_3 - y_2)/(\alpha\, x_4) \Big)^T$$

eliminates x and α from the function $f(x,y) - \alpha$. Therefore, formula (7) with matrix (14) gives the following expressions for the four components of the vector $a(x,\alpha,y)$ and $b(x,\alpha,y)$

$$a_1 = \frac{y_1}{x_2}\ , \quad a_2 = \frac{-1 + y_1^2}{x_2}\ , \quad a_3 = \frac{y_2}{x_4}\ , \quad a_4 = \frac{-1 + y_2^2}{x_4}\ , \quad b = \frac{-x_4 + x_3 y_2 + x_4 y_2^2}{\alpha\, x_4}\ .$$

Thus, in this case, we have six volume integrals, i.e., the probability function is calculated with formula (22), and the five partial derivatives are calculated with formulas (26). All integrals have the same density $p(y)$ and the same constraint function $f(x,y) \le \alpha$ defining the integration domain. Therefore, all integrals can be evaluated simultaneously with different values of α during the same simulation runs of the Monte Carlo algorithm. By solving the equation $\Psi(x,\alpha) = \beta$ w.r.t. α, we find the quantile $\alpha_\beta(x)$. Further, with formulas (24) and (25), we can evaluate derivatives of the quantile function $\alpha_\beta(x)$.

4 Optimization with Conditional Value-at-Risk Performance Function and Constraints

This chapter reviews the approach for minimization of the Conditional Value-at-Risk, which was recently developed in papers [39, 28]. In presenting this material, we follow the review paper [52]. The methodology is quite general and can be used for any application involving optimization of quantiles. However, development of this

approach is mostly stimulated by finance applications. Therefore, it will be explained in the framework of the financial contents.

Value-at-Risk (VaR), which is a quantile of a loss distribution, calculates the maximum loss with a specified confidence level. Probably the most popular technique for the estimation of VaR is the RiskMetrics methodology [37]. Although VaR is a very popular measure of risk in finance applications, it has undesirable properties [2] such as lack of sub-additivity, i.e., VaR of a portfolio with two instruments may be greater than the sum of individual VaRs of these two instruments. Also, VaR is difficult to optimize when calculated using scenarios. In this case, VaR is non-convex (see definition of convexity in [38]), non-smooth as a function of positions, and it has multiple local extrema.

An alternative measure of losses, with more attractive properties, is Conditional Value-at-Risk (CVaR), which is also called Mean Excess Loss, Mean Shortfall, or Tail VaR. CVaR is a more consistent measure of risk since it is sub-additive and convex [2]. Recently, Pflug [30] proved that CVaR is a coherent risk measure having the following properties: transition-equivariant, positively homogeneous, convex, monotonic w.r.t. stochastic dominance of order 1, and monotonic w.r.t. monotonic dominance of order 2. Moreover, as it was shown in [39], it can be optimized using linear programming (LP) and nonsmooth optimization algorithms, which allows to handle portfolios with very large numbers of instruments and scenarios. Numerical experiments indicate that the minimization of CVaR also leads to near optimal solutions in VaR terms because CVaR is always greater than or equal to VaR. When the return-loss distribution is normal, these two measures are equivalent [39], i.e., they provide the same optimal portfolio.

CVaR can be used in conjunction with VaR and is applicable to the estimation of risks with non-symmetric return-loss distributions. Although CVaR has not become a standard in the finance industry, it is likely to play a major role as it currently does in the insurance industry. Similar to the Markowitz [25] mean-variance approach, CVaR can be used in return-risk analyses. For instance, we can calculate a portfolio with a specified return and minimal CVaR. Alternatively, we can constrain CVaR and find a portfolio with maximal return, see [28]. Also, rather than constraining the variance, we can specify several CVaR constraints simultaneously with various confidence levels (thereby shaping the loss distribution), which provides a flexible and powerful risk management tool.

Similar measures as CVaR have been earlier introduced in the stochastic programming literature, although not in financial mathematics context. The conditional expectation constraints and integrated chance constraints described in [32] may serve the same purpose as CVaR. The reader interested in other applications of optimization techniques in finance area can find relevant papers in [55].

Several case studies showed that risk optimization with the CVaR performance function and constraints can be done for large portfolios and a large number of scenarios with relatively small computational resources. For instance, a problem with 1,000

instruments and 20,000 scenarios can be optimized on a 300 MHz PC in less than one minute using the CPLEX LP solver. A case study on the hedging of a portfolio of options using the CVaR minimization technique is included in [39]. This problem was first studied at Algorithmics, Inc. with the minimum expected regret approach [26]. Also, the CVaR minimization approach was applied to credit risk management of a portfolio of bonds [1]. This portfolio was put together by several banks to test various credit risk modeling techniques. Earlier, the minimum expected regret optimization technique was applied to the same portfolio at Algorithmics, Inc.[27]; we have used the same set of scenarios to test the minimum CVaR technique. A case study on optimization of a portfolio of stocks with CVaR constraints is included in [28]. The reader interested in other applications of optimization techniques in the finance area can find relevant papers in [55].

4.1 Approach

This section outlines the approach suggested in [39] for simultaneous minimization of CVaR and calculation of VaR. The next section discusses how to extend this idea to problems with CVaR constraints.

Let $f(x, y)$ be a loss function depending upon the decision vector x and a random vector y. The decision vector x belongs to a feasible set of portfolios, X. For example, we may consider portfolios with non-negative positions (short positions are not allowed) and an expected return greater than 10%.

Example 5. A Two Instrument Portfolio.
A portfolio consists of two instruments (e.g., options). Let $x = (x_1, x_2)$ be a vector of positions of these two instruments, $m = (m_1, m_2)$ be a vector of initial prices, and $y = (y_1, y_2)$ be a vector of uncertain prices of these instruments in the next period. The loss function equals the difference between the current value of the portfolio, $(x_1 m_1 + x_2 m_2)$, and an uncertain value of the portfolio at the next period $(x_1 y_1 + x_2 y_2)$, i.e.,

$$f(x, y) = (x_1 m_1 + x_2 m_2) - (x_1 y_1 + x_2 y_2) = x_1(m_1 - y_1) + x_2(m_2 - y_2).$$

If we do not allow short positions, the feasible set of portfolios is a two-dimensional set of non-negative numbers

$$X = \{(x_1, x_2) : x_1 \geq 0, x_2 \geq 0\}.$$

In this case, the loss function is linear w.r.t. positions and the feasible set is defined by a set of linear inequalities.

For convenience, we assume that the random vector y has a probability density function $p(y)$. However, the existence of the density is not critical for the considered approach; this assumption can be relaxed. Denote by $\Psi(x, \alpha)$ the probability that the loss $f(x, y)$ does not exceed some threshold value α (see, (22)). The VaR function

$\alpha(x,\beta)$, which is the percentile of the loss distribution with confidence level β, is the smallest number such that $\Psi(x,\alpha(x,\beta)) = \beta$. CVaR, denoted by $\phi_\beta(x)$, which is by definition the conditional expected loss (under the condition that it exceeds VaR), is defined by

$$\phi_\beta(x) \;=\; (1-\beta)^{-1} \int\limits_{f(x,y)\geq\alpha_\beta(x)} f(x,y)\,p(y)\,dy. \tag{28}$$

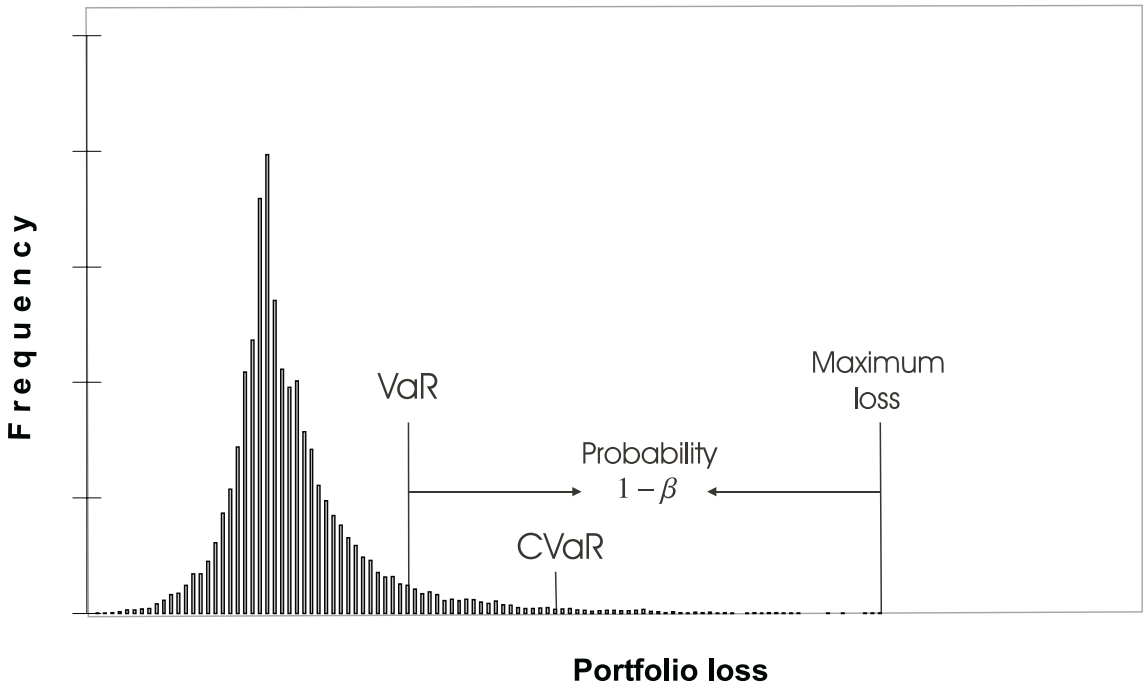

Figure 1: Portfolio Loss Distribution, VaR, and CVaR.

It is difficult to handle CVaR because of the VaR function $\alpha_\beta(x)$ involved in its definition, unless we have an analytical representation for VaR. The main idea of our approach is that we can define a much simpler function

$$F_\beta(x,\alpha) \;=\; \alpha \;+\; (1-\beta)^{-1} \int\limits_{f(x,y)\leq\alpha} (f(x,y)-\alpha)\,p(y)\,dy\,, \tag{29}$$

which can be used instead of CVaR. It can be proved: (1) the function $F_\beta(x,\alpha)$ is convex w.r.t. α; (2) VaR is a minimum point of this function w.r.t. α; and (3) that minimizing $F_\beta(x,\alpha)$ w.r.t. α gives CVaR

$$\phi_\beta(x) \;=\; F_\beta(x,\alpha_\beta(x)) \;=\; \min_{\alpha\in\mathbb{R}} F_\beta(x,\alpha). \tag{30}$$

This follows from the fact that the derivative of the function $F_\beta(x,\alpha)$ w.r.t. α equals

$$1 + (1-\beta)^{-1}(\Psi(x,\alpha) - 1) \ ,$$

see details in [39]. By equating the derivative to zero we immediately obtain that VaR minimizes the function $F_\beta(x,\alpha)$ w.r.t. α. Furthermore, we can use the function $F_\beta(x,\alpha)$ for the simultaneous calculation of VaR and the optimization of CVaR, i.e.,

$$\min_{x\in X} \phi_\beta(x) \ = \ \min_{(x,\alpha)\in X\times \mathbb{R}} F_\beta(x,\alpha) \ . \tag{31}$$

Indeed, minimization of the function $F_\beta(x,\alpha)$ w.r.t. both variables optimizes CVaR and finds VaR in "one shot". Let (x^*,α^*) be a solution of the above minimization problem. Then, $F_\beta(x^*,\alpha^*)$ equals an optimal CVaR, the optimal portfolio equals x^*, and the corresponding VaR equals α^*. Under quite general conditions (see, Chapter 1 of this paper) the function $F_\beta(x,\alpha)$ is smooth. Moreover, if the function $f(x,y)$ is convex w.r.t. x, then the function $F_\beta(x,\alpha)$ is also convex w.r.t. x. Thus, if we want to minimize CVaR, we can use the convex smooth function $F_\beta(x,\alpha)$. Therefore, if the feasible set X is also convex, we need to solve a smooth convex optimization problem.

4.2 Optimization Problems with CVaR Constraints

Banks, investment companies, and other businesses tolerate different levels of risk, depending upon their objectives and capital. The adequate representation and management of risk is a critical task for business success. A typical approach in risk management is to estimate and control VaR with a specified confidence level, such as 0.95, 0.99, or 0.999. VaR is estimated for various periods, depending upon the risk management objectives - short term VaR is estimated usually for one day or two weeks, longer terms may include one, two, or five years. The problem of controlling VaR can be formalized as a mathematical programming problem with VaR constraints. However, such a problem is very difficult to solve using formal optimization methods because VaR is non-convex w.r.t. the portfolio positions and it has many local minima. In this section, we show that in contrast to VaR constraints, CVaR constraints can be easily handled using formal optimization approaches. Constraining CVaR also restricts VaR because CVaR $\geq$ VaR. Therefore, VaR constraints can be replaced by more conservative CVaR constraints.

Similar to CVaR minimization, we can include CVaR in constraints and replace it by the function $F_\beta(x,\alpha)$, see [28]. For instance, let us consider the problem of minimizing the mean losses, $\mu(x) = E\, f(x,y)$, subject to some balance constraints $x \in X$, and two CVaR constraints with confidence levels β and γ. In this case, the optimization problem can be stated as follows

$$\min \ \ \mu(x)$$

subject to

$$x \in X \ ,$$

$$\phi_\beta(x) \leq C_\beta \ ,$$

$$\phi_\gamma(x) \leq C_\gamma \ ,$$

where C_β and C_γ are some constants constraining CVaR at different confidence levels. The last two constraints can be replaced by constraints

$$F_\beta(x, \alpha_1) \leq C_\beta \ ,$$

$$F_\gamma(x, \alpha_1) \leq C_\gamma \ .$$

Indeed, if these constraints are satisfied for some α_1 and α_2, then they are satisfied for the minimal values

$$\min_{\alpha_1} F_\beta(x, \alpha_1) = \phi_\beta(x) \ ,$$

and

$$\min_{\alpha_2} F_\gamma(x, \alpha_2) = \phi_\gamma(x) \ .$$

Optimization with these constraints assures that the CVaR values are properly restricted. Moreover, if a risk constraint is active, e.g., in the first constraint, $F_\beta(x^*, \alpha_1^*) = C_\beta$, then the optimal value α_1^* equals β-VaR.

4.3 Minimizing CVaR with Finite Number of Scenarios: Linear Programming

Let us consider now the case in which an analytical representation of the density function $p(y)$ is not available, but we have J scenarios, $y_j, j = 1, ..., J$, sampled from the density $p(y)$. For instance, we may have historical observations of prices of instruments in the portfolio, or we may use Monte Carlo simulations to price the instruments. In this case, the function $F_\beta(x, \alpha)$ can be calculated approximately as follows

$$\tilde{F}(x, \alpha) \ \stackrel{def}{=} \ \alpha \ + \ \nu \sum_{j=1}^{J} \Big(f(x, y_j) - \alpha \Big)^+ \ ,$$

where the constant ν equals $\nu = ((1 - \beta)J)^{-1}$ and $t^+ = \max(0, t)$. If the function $f(x, y)$ is convex w.r.t. x, then the function $\tilde{F}(x, \alpha)$ is a convex nonsmooth function w.r.t. the vector (x, a). Therefore, if the feasible set X is convex, the optimization problem with the CVaR performance function can be solved using non-smooth optimization techniques. Moreover, if the function $f(x, y)$ is linear w.r.t. x, this problem can be solved using LP techniques. LP approaches are routinely used in portfolio

optimization with various criteria, such as mean absolute deviation [19], maximum deviation [54], and mean regret [5].

Let us first explain how LP techniques can be used for the minimization of CVaR. Indeed, after replacing in $\tilde{F}(x,\alpha)$ the terms $\left(f(x,y_j)-\alpha\right)^+$ by auxiliary variables z_j, and imposing constraints $z_j \ \geq \ f(x,y_j)-\alpha\,,\ z_j \ \geq \ 0\,,\ j=1,\ldots,J\,,$ we can reduce the minimization of the function $\tilde{F}(x,\alpha)$ to the following optimization problem

$$\min_{x\in\mathbb{R}^n,z\in\mathbb{R}^J,\alpha\in\mathbb{R}} \quad \alpha \ + \ \nu\sum_{j=1}^{J} z_j \tag{32}$$

subject to

$$x\in X\ , \tag{33}$$

$$z_j \ \geq \ f(x,y_j)-\alpha\ , \quad z_j \ \geq \ 0\ , \quad j=1,\ldots,J\ . \tag{34}$$

Several case studies (see, [1, 28, 39]) have demonstrated that this formulation provides a very powerful and numerically stable technique which can solve problems with a large number of instruments and scenarios.

Example 6. CVaR Minimization with a Constraint on Mean Losses.
Suppose that we want to minimize CVaR of the small portfolio described in Example 5. We are interested in minimizing one day CVaR under the condition that the mean daily portfolio losses are less than or equal to $-R$ (i.e., the mean profit is bigger than or equal to R). Suppose that for two instruments in the portfolio, we have prices for J previous days. From this historical data, we can estimate J daily returns and calculate J scenarios for the next day prices, $y_j=(y_{j1},y_{j2}), j=1,...,J$. The mean portfolio loss equals

$$\mu(x) \ = J^{-1}\ \sum_{j=1}^{J} f(x,y_j) \ = \ J^{-1}\sum_{j=1}^{J}(x_1(m_1-y_{j1})+x_2(m_2-y_{j2}))\ .$$

The constraint on the mean losses is formulated as follows

$$J^{-1}\sum_{j=1}^{J}(x_1(m_1-y_{j1})+x_2(m_2-y_{j2})) \ \leq \ -R\ . \tag{35}$$

The CVaR minimization problem can be easily solved by minimizing the linear function (32) subject to linear constraints (33),(34), and (35). This problem can be solved using standard LP solvers such as CPLEX.

4.4 Linearization of CVaR Constraints with Finite Number of Scenarios

The previous section showed that the nonlinear CVaR function can be minimized using a linear objective function and linear constraints. Here, we show that a CVaR

constraint in optimization problems can be approximated by a set of linear constraints. Let J scenarios, y_j, $j = 1, ..., J$, be sampled from the density $p(y)$. Suppose that a CVaR constraint, $\phi_\beta(x) \leq C_\beta$ needs to be satisfied. As it was earlier discussed, this constraint can be replaced by the constraint $F_\beta(x, \alpha) \leq C_\beta$ using the additional variable α. Further, we can approximate this constraint by the constraint $\tilde{F}_\beta(x, \alpha) \leq C_\beta$ using scenarios y_j, $j = 1, ..., J$. Finally, the last constraint can be equivalently represented by the set of constraints

$$\alpha \ + \ \nu \sum_{j=1}^{J} z_j \ \leq \ C_\beta \ , \tag{36}$$

$$z_j \ \geq \ f(x, y_j) - \alpha \ , \quad z_j \ \geq \ 0 \ , \quad j = 1, \ldots, J \ . \tag{37}$$

If constraint (36) is active, then the optimal value α^* equals VaR. A case study on the application of these techniques to the optimization of the portfolio consisting of the S&P100 stocks can be found in [28].

5 Conclusion

This introductory paper reviewed several topics related to the analysis of the probabilistic and quantile functions: (1) sensitivity analysis of probabilistic functions; (2) sensitivity analysis of quantiles (VaR); and (3) optimization approaches for the risk management problems with the CVaR performance function and constraints.

This overview is very far from being comprehensive and many important issues are beyond the scope of this paper. For instance, convexity of the probability and quantile functions is not discussed in this paper. However, this topic is well studied in the literature, see [32] and [14] and a reader interested in this material can find it in these and other publications. The emphasis of this paper is on issues which have been relatively recently developed. Also, this volume provides a lot of new results in the area of theory and applications of the probabilistic and quantile functions.

References

[1] Andersson, F., Mausser, H., Rosen, D., and S. Uryasev (2000) Credit Risk Optimization With Conditional Value-At-Risk Criterion. *Mathematical Programming*. To appear.

[2] Artzner, P., Delbaen F., Eber, J.M., and D. Heath (1999) Coherent Measures of Risk. *Mathematical Finance*, 9, 203-228.

[3] Birge J.R. and F. Louveaux (1997) Introduction to Stochastic Programming. Springer, New York.

[4] Bucay, N. and D. Rosen (1999) Credit Risk of an International Bond Portfolio: a Case Study. *ALGO Research Quarterly.* Vol.2, No. 1, 9–29.

[5] Dembo, R.S. and A.J. King (1992) Tracking Models and the Optimal Regret Distribution in Asset Allocation. *Applied Stochastic Models and Data Analysis.* Vol. 8, 151-157.

[6] Duffie, D. and J. Pan (1997) An Overview of Value-at-Risk. *Journal of Derivatives.* 4, 7–49.

[7] Ermoliev, Yu. (1976) *Stochastic Programming Methods.* Nauka, Moscow. (in Russian).

[8] Ermoliev, Yu. (1983) Stochastic Quasi-Gradient Methods and Their Applications to System Optimization. *Stochastics*, 4, 1-36.

[9] Ermoliev, Yu. and R. J-B Wets (Eds.) (1988): *Numerical Techniques for Stochastic Optimization*, Springer Series in Computational Mathematics, 10.

[10] Glasserman P. (1991) *Gradient Estimation Via Perturbation Analysis.* Kluwer, Boston.

[11] Ho, Y.C. and X.R. Cao (1991) *Perturbation Analysis of Discrete Event Dynamic Systems.* Kluwer, Boston.

[12] Ioffe, A.D. and V.L. Levin (1972) Subdifferentials of Convex Functions. *Papers of Moscow Mathematical Society* (Trudy MMO), V. 26, 3-73 (in Russian).

[13] Kall, P. and S.W. Wallace (1994) *Stochastic Programming.* Willey, Chichester.

[14] Kan, Y.S., and Kibzun, A.I. (1996). *Stochastic Programming Problems with Probability and Quantile Functions*, John Wiley & Sons, 316.

[15] Kast, R., Luciano, E., and L. Peccati (1998) *VaR and Optimization.* 2nd International Workshop on Preferences and Decisions, Trento, July 1-3 1998.

[16] Kibzun, A.I., Malyshev, V.V., and D.E. Chernov (1988) Two Approaches to Solutions of Probabilistic Optimization Problems. *Soviet Journal of Automaton and Information Silences*, V.20, No.3, 20-25.

[17] Kibzun, A.I. and G.L. Tretyakov (1996) Probability Function Differentiability. *Doklady RAN,* (in Russian).

[18] Kibzun A.I. and S. Uryasev. (1998) Differentiability of Probability Functions. Stochastic Analysis and Applications. 16(6), 1101-1128.

[19] Konno, H. and H. Yamazaki (1991): Mean Absolute Deviation Portfolio Optimization Model and Its Application to Tokyo Stock Market. *Management Science.* **37**, 519-531.

[20] Litterman, R. (1997) Hot Spots and Hedges (I). *Risk.* 10 (3), 42-45.

[21] Litterman, R. (1997) Hot Spots and Hedges (II). Risk, 10 (5), 38-42.

[22] Lucas, A., and Klaassen, P. (1998) Extreme Returns, Downside Risk, and Optimal Asset Allocation. *Journal of Portfolio Management*, Vol. 25, No. 1, 71-79.

[23] Marti,K. (1994) Approximations and Derivatives of Probability Functions. *In Approximation, Probability and Related Fields,* edited by G. Anastassiou and S.T. Rachev, Plenum Press, New York., 367-377.

[24] Marti K. (1995) Differentiation Formulas for Probability Functions: The Transformation Method. *Mathematical Programming Journal*, Series B, Vol. 75, No. 2.

[25] Markowitz, H.M. (1952) Portfolio Selection. *Journal of finance.* Vol.7, 1, 77–91.

[26] Mausser, H. and D. Rosen (1998) Beyond VaR: From Measuring Risk to Managing Risk, *ALGO Research Quarterly*, Vol. 1, No. 2, 5–20.

[27] Mausser, H. and D. Rosen (1999) Applying Scenario Optimization to Portfolio Credit Risk, *ALGO Research Quarterly*, Vol. 2, No. 2, 19–33.

[28] Palmquist, J., Uryasev, S., and P. Krokhmal (2000) Portfolio Optimization with Conditional Value-At-Risk Objective and Constraints. *The Journal of Risk.* To appear.

[29] Pflug, G.Ch. (1996) *Optimization of Stochastic Models : The Interface Between Simulation and Optimization.* Kluwer Academic Publishers, Dordrecht, Boston.

[30] Pflug, G.Ch. (2000) *Some Remarks on the Value-at-Risk and the Conditional Value-at-Risk.* In."Probabilistic Constrained Optimization: Methodology and Applications", Ed. S. Uryasev, Kluwer Academic Publishers, 2000

[31] Prékopa, A. (1970) On Probabilistic Constrained Programming, in: *Proceedings of the Princeton Symposium on Mathematical Programming,* (Princeton University Press, Princeton, N.J.), 113–138.

[32] Prekopa, A. (1995) *Stochastic Programming*, Kluwer Academic Publishers.

[33] Pulkkinen, A. and S. Uryasev (1991) *Optimal Operational Strategies for an Inspected Component – Solution Techniques.* Collaborative Paper CP-91-13, International Institute for Applied Systems Analysis, Laxenburg, Austria.

[34] Prékopa, A. and T. Szántai (1978) A New Multivariate Gamma Distribution and Its Fitting to Empirical Streamflow Data. *Water Resources Research*, 14, 19-24.

[35] Pritsker, M. (1997) Evaluating Value at Risk Methodologies, *Journal of Financial Services Research*, 12:2/3, 201-242.

[36] Raik, E. (1975) The Differentiability in the Parameter of the Probability Function and Optimization of the Probability Function via the Stochastic Pseudogradient Method. *Eesti NSV Teaduste Akadeemia Toimetised. Füüsika-Matemaatika,* 24, 1, 3-6 (in Russian).

[37] RiskMetrics™ (1996) *Technical Document*, 4-th Edition, New York, NY, J.P.Morgan Inc., December.

[38] Rockafellar, R.T. (1970): *Convex Analysis.* Princeton Mathematics, Vol. 28, Princeton Univ. Press.

[39] Rockafellar, R.T. and S. Uryasev (2000) Optimization of Conditional Value-At-Risk. *The Journal of Risk*, Vol. 2, No. 3.

[40] Rockafellar, R.T. and R.J.-B. Wets (1982) On the Interchange of Subdifferentiation and Conditional Expectation for Convex Functionals. *Stochastics*, 7, 173-182.

[41] Roenko, N. (1983) *Stochastic Programming Problems with Integral Functionals over Multivalued Mappings*, (Ph.D. Thesis, Kiev, Ukraine) (in Russian).

[42] Rubinstein, R. (1992) Sensitivity Analysis of Discrete Event Systems by the "Push Out" Method. *Annals of Operations Research*, 39.

[43] Rubinstein, R. and A. Shapiro (1993) *Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization via the Score Function Method.* Willey, Chichester.

[44] Simon, J. (1989) Second Variation in Domain Optimization Problems. *In International Series of Numerical Mathematics,* ed.by F.Kappel, K.Kunish and W.Schappacher, Birkhauser Verlag, 91, 361-378.

[45] Simons, K. (1996) Value-at-Risk New Approaches to Risk Management. *New England Economic Review*, Sept/Oct, 3-13.

[46] Stambaugh, F. (1996) Risk and Value-at-Risk. *European Management Journal*, Vol. 14, No. 6, 612-621.

[47] Thompson, K.M., D.E. Burmasater and E.A.C. Crouch (1992) *Monte Carlo Techniques for Quantitative Uncertainty Analysis in Public Risk Assessments.* Risk Analysis, 12.

[48] Uryasev, S. (1987) *Adaptive Algorithms for Stochastic Optimization and Game Theory.* Nauka, Moscow (in Russian).

[49] Uryasev, S. (1989) A Differentiation Formula for Integrals over Sets Given by Inclusion. *Numerical Functional Analysis and Optimization,* 10 (7 & 8), 827-841.

[50] Uryasev, S. (1994) Derivatives of Probability Functions and Integrals over Sets Given by Inequalities. *J. Computational and Applied Mathematics,* 56, 197-223.

[51] Uryasev, S. (1995) Derivatives of Probability Functions and some Applications.*Annals of Operations Research*, 56, 287-311.

[52] Uryasev, S. (2000) Conditional Value-at-Risk: Optimization Algorithms and Applications. Financial Engineering News, No. 14, February.

[53] Uryasev, S. and A. Shlyakhter (1994) *A Procedure for Simultaneous Calculation of Sensitivities in Probabilistic Risk Analysis.* In: Abstracts Society for Risk Analysis Annual Conference and Exposition, Baltimore, Maryland, December 1994.

[54] Young, M.R. (1998): A Minimax Portfolio Selection Rule with Linear Programming Solution. *Management Science.* Vol.44, No. 5, 673–683.

[55] Ziemba, W.T. and J.M. Mulvey (Eds.) (1998): *Worldwide Asset and Liability Modeling*, Cambridge Univ. Pr.