



**HAL**  
open science

# Estimation of Systemic Shortfall Risk Measure using Stochastic Algorithms

Sarah Kaakai, Anis Matoussi, Achraf Tamtalini

► **To cite this version:**

Sarah Kaakai, Anis Matoussi, Achraf Tamtalini. Estimation of Systemic Shortfall Risk Measure using Stochastic Algorithms. 2023. hal-03871246v2

**HAL Id: hal-03871246**

**<https://hal.science/hal-03871246v2>**

Preprint submitted on 14 Nov 2023 (v2), last revised 19 Feb 2024 (v3)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Estimation of Systemic Shortfall Risk Measure using Stochastic Algorithms\*

Sarah Kaakäi <sup>†</sup>      Anis Matoussi <sup>‡</sup>      Achraf Tamtalini <sup>§</sup>

## Abstract

Systemic risk measures were introduced to capture the global risk and the corresponding contagion effects that is generated by an interconnected system of financial institutions. To this purpose, two approaches were suggested. In the first one, systemic risk measures can be interpreted as the minimal amount of cash needed to secure a system *after aggregating* individual risks. In the second approach, systemic risk measures can be interpreted as the minimal amount of cash that secures a system by allocating capital to each single institution *before aggregating* individual risks. Although the theory behind these risk measures has been well investigated by several authors, the numerical part has been neglected so far. In this paper, we use stochastic algorithms schemes in estimating Multivariate Shortfall Risk Measure (MSRM) and prove that the resulting estimators are consistent and asymptotically normal. We also test numerically the performance of these algorithms on several examples.

**Keywords:** Multivariate risk measures, shortfall risk, stochastic algorithms, stochastic root finding, risk allocations.

## Introduction

The axiomatic theory of risk measures, first initiated by the seminal paper of [Artzner et al. \(1999\)](#), has been widely studied during the last years. The Value-at-Risk (VaR) is one of

---

\*Acknowledgements: The authors research is part of the ANR project DREAMeS (ANR-21-CE46-0002) and benefited from the support of the "Chair Risques Emergents en Assurance" under the aegis of Fondation du Risque, a joint initiative by Le Mans University and Covéa.

<sup>†</sup>Laboratoire Manceau de Mathématiques & FR CNRS N° 2962, Institut du Risque et de l'Assurance, Le Mans University.

<sup>‡</sup>Laboratoire Manceau de Mathématiques & FR CNRS N° 2962, Institut du Risque et de l'Assurance, Le Mans University and Ecole Polytechnique.

<sup>§</sup>Laboratoire Manceau de Mathématiques & FR CNRS N° 2962, Institut du Risque et de l'Assurance, Le Mans University.

---

the most known and common risk measures used by practitioners and regulation authorities. However, the VaR lacks one important property: it does not take into account the diversification effect. To circumvent this problem, the VaR was replaced by the Conditional Value-at-Risk (CVaR) and a more general framework of improved risk measures has been introduced: *Utility-based Shortfall Risk* (SR). Nevertheless, when it comes to a system of financial institutions or portfolios, the question about how to assess the global risk, as well as individual risks, arises. Following the 2008 crisis, the traditional approach of measuring systemic risk that consists in considering each institution as a single entity isolated from other institutions, has shown some limits. Indeed, with this approach, the risk associated with a vector of positions  $X = (X_1, \dots, X_d)$  can be written as:

$$R(X) := \sum_{i=1}^d \eta_i(X_i),$$

where each  $\eta_i$  is a univariate risk measure. Then, [Chen et al. \(2013\)](#) proposed an approach that is very close in spirit to the axiomatic framework initiated by [Artzner et al. \(1999\)](#). They showed that any systemic risk measure verifying certain axioms is the composition of a univariate risk measure  $\eta$  and an aggregation  $\Lambda$ , i.e.,

$$R(X) = \eta(\Lambda(X)).$$

The previous representation is known as the “Aggregate then Add Cash” approach as it consists first in aggregating the positions  $X_1, \dots, X_d$  through the aggregation function  $\Lambda$  and then to apply a univariate risk measure. One of the most common ways to aggregate the outcomes  $X_i$  is to simply take the sum, that is to consider,  $\Lambda(x) = \sum_{i=1}^d x_i$ . It is worth noticing that, while summing up profit and losses might seem reasonable from the point of view of a portfolio manager because portfolios profits and losses compensate each other, this aggregation rule seems inadequate from the point of view of a regulator where cross-subsidization between institutions is not realistic since no institution will be willing to pay for the losses of another one.

Motivated by these considerations, [Biagini et al. \(2019\)](#) proposed another approach to measure the systemic risk. They first considered the systemic risk as the minimal capital that secures the system by injecting capital into the single institutions, *before aggregating the individual risks*:

$$R(X) := \inf \left\{ \sum_{i=1}^d m_i, \Lambda(X + m) \in \mathbb{A} \right\}, \quad (0.1)$$

where  $\mathbb{A}$  is an acceptance set. This approach, known as “Add Cash then Aggregate” consists in adding the amount  $m_i$  to the financial position  $X_i$  before the corresponding total loss  $\Lambda(X + m)$  is computed. The systemic risk is then measured as the minimal total amount  $\sum_{i=1}^d m_i$  injected into the institutions to make it acceptable. With this approach, a joint

---

measure of total risk as well as individuals risk contributions to systemic risk is obtained. If  $m^* = (m_1^*, \dots, m_d^*)$  is an optimum, that is  $R(X) = \sum_{i=1}^d m_i^*$  and  $\Lambda(X + m^*) \in \mathbb{A}$ , one could order the  $m_i^*$ 's and hence be able to say that institution/portfolio  $i$  requires more cash allocation or is riskier than  $j$  if  $m_i^* \geq m_j^*$ .

In this article, we are interested in the numerical computation of the multivariate shortfall risk measure (MSRM) that was introduced in [Armenti et al. \(2018\)](#). They are an extension of univariate SR measures, and can be obtained by taking the aggregation function  $\Lambda(x) = l(-x)$  where  $l$  is a multivariate loss function (see [Section 1](#)) and the acceptance set  $\mathbb{A} = \{X \in L^0(\mathbb{R}), E[X] \leq 0\}$ .

In [Armenti et al. \(2018\)](#), MSRMs and their optimal allocations are approximated using a two step procedure. The expectation in the aggregation function is first approximated using Monte Carlo, Fourier or Chebychev schemes. The minimum in [equation \(0.1\)](#) is then found by a built-in minimizing algorithm in python. However, this minimizing algorithm can be sensitive to the starting point of the algorithm, and there are no theoretical results on the convergence of this procedure. In the univariate case, SR can also be characterized as the unique root of a function  $g : \mathbb{R} \mapsto \mathbb{R}$  that is expressed as an expectation, which can be estimated by using a deterministic root finding algorithm combined with a Monte Carlo procedure that estimates  $g(s)$  for all  $s \in \mathbb{R}$ .

Another idea to compute the VaR and CVaR comes from the fact that they are solutions and the value of the same convex optimization problem as pointed out in [Rockafellar and Uryasev \(2000\)](#). Moreover, as they can be expressed as an expectation, this led [Bardou et al. \(2009\)](#) to define consistent and asymptotically normal estimators of both quantities using a classical Robbins-Monro (RM) procedure. For SR, a constrained RM algorithm has also been proposed by [Dunkel and Weber \(2010\)](#).

In this paper, we introduce stochastic algorithms for MSRM, and prove their consistency and asymptotic normality. Indeed, the optimal allocations of MSRM can also be characterized as the root of a function that is expressed as an expectation. However, the multivariate setting adds additional constraints: the optimal allocations are characterized as the solution of a first order condition of the Lagrangian associated with the MSRM. Hence, the new problem is to find a saddle point  $z^* = (m^*, \lambda^*) \in \mathbb{R}^d \times \mathbb{R}^+$ , which is the unique root of a function  $h(z) = E[H(X, z)]$  determined by the first order conditions.

RM stochastic algorithms have been the subject of a vast literature (see e.g. [Benveniste et al. \(1990\)](#), [Duflo \(1996\)](#)). The basic paradigm in its simplest form is to approximate  $z^*$  by the following stochastic difference equation:  $Z_{n+1} = Z_n + \gamma_n Y_n$ , where  $\gamma_n > 0$  is the step size that goes to zero as  $n \rightarrow \infty$ , and  $Y_n = H(X_{n+1}, Z_n)$ , where  $(X_n)$  is a sequence of i.i.d random variables with the same law as  $X$ . Moreover, if the random variable  $X$  is not directly simulatable, [Frikha \(2016\)](#) extended the scope of multi-level Monte Carlo to the framework of stochastic algorithms. In many cases, the analysis of these algorithms is based on the so-called Ordinary Differential Equation (ODE) method introduced by [Ljung \(1977\)](#). The main idea

---

is to show that, in the long run, the noise is eliminated so that, asymptotically, the behaviour of the algorithm is determined by that of the “mean” ODE:  $\dot{z} = h(z)$ .

To ensure the convergence of RM algorithms, the function  $h$  is assumed to have sub-linear growth of the function  $h$ . This assumption is often not verified for MSRM, due to the choice of loss functions. In order to avoid reducing drastically the scope of the applications, we introduce a constrained version of the RM algorithm for MSRM. This consists in using the projection into a compact  $K$  each time the sequence  $Z_n$  goes out of  $K$ . First introduced by [Kushner and Sanvicente \(1975\)](#), an excellent survey on projection techniques can be found in [Kushner and Yin \(2003\)](#). In the case of MSRM, the convergence of the constrained algorithm toward the optimal allocation  $z^*$  is then closely related to the study a projected multivariate ODE, whose solutions remain on  $K$ . We provide a rigorous analysis of the asymptotic stability of this ODE, based on the introduction of a Lyapunov function and on the extension of the Lasalle’s invariant principle to this ODE with discontinuous r.h.s. Under additional assumptions, we also establish the asymptotic normality of the estimator.

The analysis and estimation of the asymptotic covariance matrix is also more difficult due to the multivariate setting. By introducing an averaged “Polyak-Ruppert” version of the stochastic algorithm, we obtain a simpler expression of the asymptotic covariance matrix and propose an estimator allowing the computation of empirical confidence intervals.

Stochastic algorithms also offer a flexible framework that can be conveniently combined with variance reduction techniques such as Importance Sampling (IS). [Dunkel and Weber \(2010\)](#) have adapted the RM for unidimensional SR in order to include IS. However, in order to apply this non parametric approach, it is necessary to find beforehand an appropriate change of measure for each parameter  $z$ , which can be very difficult even in the one dimensional case. Inspired by the parametric approach of [Bardou et al. \(2009\)](#) for the VaR/CVaR, we also propose an adaptive RM algorithm based on recursive IS. This also provides a general variance reduction method for univariate SR, which doesn’t require prior knowledge of the change of measure.

The paper is organized as follows. Section 1, is dedicated to MSRM. In particular, we recall the main theorem characterizing the optimal allocations. The main results of consistency and asymptotic normality of the RM algorithm are presented in Section 2, along with the Polyak-Ruppert averaging and adaptative importance sampling. Finally, section 3 is devoted to some numerical experiments of our procedures. We present a first testing example with an exponential loss function, where we have a closed formula for optimal allocations. We also give a second example using a loss function with a mixture of positive part and quadratic functions.

# 1 About Multivariate Risk Measures

Let  $(\Omega, \mathcal{F}, P)$  be a probability space, and denote by  $L^0(\mathbb{R}^d)$  the space of  $\mathcal{F}$ -measurable  $d$ -variate random variables on this space with  $d \geq 2$ . For  $x, y \in \mathbb{R}^d$ , we say that that  $x \geq y$  ( $x > y$  resp.) if  $x_k \geq y_k$  ( $x_k > y_k$  resp.) for every  $1 \leq k \leq d$ . We denote by  $\|\cdot\|$  the Euclidean norm, and  $x \cdot y = \sum x_k y_k$ . For a function  $f : \mathbb{R}^d \mapsto [-\infty, \infty]$ , we denote by  $f^*(y) = \sup_x \{x \cdot y - f(x)\}$  the convex conjugate of  $f$ . The space  $L^0(\mathbb{R}^d)$  inherits the lattice structure of  $\mathbb{R}^d$  and therefore, we can use the classical notations in  $\mathbb{R}^d$  in a  $P$ -almost-sure sens. We say, for example, for  $X, Y \in L^0(\mathbb{R}^d)$ , that  $X \geq Y$  (or  $X > Y$  resp.) if  $P(X \geq Y) = 1$  (or  $P(X > Y) = 1$  resp.). To simplify the notation, we will simply write  $L^0$  instead of  $L^0(\mathbb{R}^d)$ . Now, let  $X = (X_1, \dots, X_d) \in L^0$  be a random vector of financial losses, i.e., negative values of  $X_k$  represents actually profits. We want to assess the systemic risk of the whole system and to determine a monetary risk measure, which will be denoted  $R(X)$ , as well as a risk allocation  $RA_k(X), k = 1, \dots, d$  among the  $d$  risk components. Inspired by the univariate case introduced in Föllmer and Schied (2002), Armenti et al. (2018) introduced a multivariate extension of shortfall risk measures by the means of loss functions and sets of acceptable monetary risk allocations.

**Definition 1.1.** A function  $l : \mathbb{R}^d \mapsto (-\infty, \infty]$  is called a loss function if:

(A1)  $l$  is increasing, that is  $l(x) \geq l(y)$  if  $x \geq y$ ;

(A2)  $l$  is convex and lower-semicontinuous with  $\inf l < 0$ ;

(A3)  $l(x) \geq \sum x_k - c$  for some constant  $c$ .

Furthermore, a loss function  $l$  is said to be permutation invariant if  $l(x) = l(\pi(x))$  for every permutation  $\pi$  of its components.

**Comment:** The property (A1) expresses the normative fact about the risk, that is, the more losses we have, the riskier is our system. As for (A2), it expresses the desired property of diversification. Finally, (A3) says that the loss function put more weight on high losses than a risk neutral evaluation.

**Example 1.** Let  $h : \mathbb{R} \mapsto \mathbb{R}$  be one dimensional loss function satisfying condition (A1), (A2) and (A3). We could build a multivariate loss function using this one dimensional loss function in the following way:

(C1)  $l(x) = h(\sum x_k)$ ;

(C2)  $l(x) = \sum h(x_k)$ ;

(C3)  $l(x) = \alpha h(\sum x_k) + (1 - \alpha) \sum h(x_k)$  for  $0 \leq \alpha \leq 1$ .

More specifically, in (C1), we are aggregating losses before evaluating the risk, whereas in (C2), we evaluate individual risks before aggregating. The loss function in (C3) is a convex

combination of those in (C1) and (C2).

One of the main examples we will be studying in this paper are the two following ones:

$$h(x) = \frac{1}{\alpha + 1} \left( \sum_i e^{\beta x_i} + \alpha e^{\beta \sum x_i} \right) - \frac{\alpha + d}{\alpha + 1}, \quad h(x) = \sum_i x_i + \frac{1}{2} \sum_i (x_i^+)^2 + \alpha \sum_{i < j} x_i^+ x_j^+,$$

where the coefficient  $\alpha > 0$  is called the systemic weight and  $\beta > 0$  is a risk aversion coefficient.

In the following, we will consider multivariate risk measures defined on Orlicz spaces (see [Rao and Ren \(1991\)](#) for further details on the theory of Orlicz spaces). This has several advantages. From a mathematical point of view, it is a more general setting than  $L^\infty$ , and in the same time, it simplifies the analysis especially for utility maximization problems. Therefore, we will consider loss vectors in the following multivariate Orlicz heart:

$$M^\theta = \{X \in L^0 : E[\theta(\lambda X)] < \infty, \forall \lambda > 0\},$$

where  $\theta(x) = l(|x|), x \in \mathbb{R}^d$ . See Appendix in [Armenti et al. \(2018\)](#) for more details about Orlicz spaces.

Next, we give the definition of multivariate shortfall risk measures as it was introduced in [Armenti et al. \(2018\)](#).

**Definition 1.2.** *Let  $l$  be a multivariate loss function and  $X \in M^\theta$ , we define the acceptance set  $\mathcal{A}(X)$  by:*

$$\mathcal{A}(X) := \{m \in \mathbb{R}^d : E[l(X - m)] \leq 0\}.$$

*The multivariate shortfall risk of  $X \in M^\theta$  is defined as:*

$$R(X) := \inf \left\{ \sum m_k : m \in \mathcal{A}(X) \right\} = \inf \left\{ \sum m_k : E[l(X - m)] \leq 0 \right\}. \quad (1.1)$$

**Remark 1.3.** *When  $d = 1$ , the above definition corresponds exactly to the univariate shortfall risk measure in [Föllmer and Schied \(2002\)](#).*

The following theorem from [Armenti et al. \(2018\)](#) shows that the multivariate shortfall risk measure has the desired properties and admits a dual representation as in the case of univariate shortfall risk measure. We introduce  $Q^{\theta^*}$  the set of measure densities in  $L^{\theta^*}$ , the dual space of  $M^\theta$ :

$$Q^{\theta^*} := \left\{ \frac{dQ}{dP} := (Z_1, \dots, Z_d), Z \in L^{\theta^*}, Z_k \geq 0 \text{ and } E[Z_k] = 1 \text{ for every } k \right\}.$$

**Theorem 1.4.** *[Theorem 2.10 in [Armenti et al. \(2018\)](#)] The function*

$$R(X) := \inf \left\{ \sum m_k : m \in \mathcal{A}(X) \right\},$$

*is real-valued, convex, monotone and translation invariant. Moreover, it admits the dual*

representation:

$$R(X) = \max_{Q \in Q^{\theta^*}} \{E_Q[X] - \alpha(Q)\}, \quad X \in M^\theta,$$

where the penalty function is given by

$$\alpha(Q) = \inf_{\lambda > 0} E \left[ \lambda l^* \left( \frac{dQ}{\lambda dP} \right) \right], \quad Q \in Q^{\theta^*}.$$

Now, we address the question of existence and uniqueness of a risk allocation which are not straightforward in the multivariate case. [Armenti et al. \(2018\)](#) showed that if the loss function is permutation invariant, then risk allocations exist and they are characterized by Kuhn-Tucker conditions. We denote by  $Z = \{m \in \mathbb{R}^d, \sum m_i = 0\}$  the *zero-sum allocations* set.

**Definition 1.5.** *A risk allocation is an acceptable monetary risk allocation  $m \in \mathcal{A}(X)$  such that  $R(X) = \sum m_k$ . When a risk allocation is uniquely determined, we denote it by  $RA(X)$ .*

We make the following assumption on the loss function  $l$  and the vector of losses  $X \in M^\theta$ :

- ( $\mathcal{A}_1$ )    i. For every  $m_0$ ,  $m \mapsto l(X - m)$  is differentiable at  $m_0$  a.s.;  
           ii.  $l$  is permutation invariant.

**Theorem 1.6.** *[Theorem 3.4 in [Armenti et al. \(2018\)](#)] Let  $l$  be a loss function and  $X \in M^\theta$  such that assumption ( $\mathcal{A}_1$ ) holds. Then, risk allocations  $m^* \in \mathbb{R}^d$  exists and they are characterized by the first order conditions:*

$$1 = \lambda^* \mathbb{E}[\nabla l(X - m^*)], \quad \mathbb{E}[l(X - m^*)] = 0,$$

where  $\lambda^* \geq 0$  is a Lagrange multiplier. If moreover  $l(x + \cdot)$  is strictly convex along zero sum allocations for every  $x$  such that  $l(x) \geq 0$ , the risk allocation is unique.

**Comment:** Let  $f_0(m) = \sum_{i=1}^d m_i$  and  $f_1(m) := E[l(X - m)]$ , for  $m \in \mathbb{R}^d$  and  $X \in M^\theta$ . The assumption ( $\mathcal{A}_1$ )-i. together with the convexity of the function  $m \mapsto l(X - m)$ , we have that, by Theorem 7.46 in [Shapiro et al. \(2009\)](#),  $f_1$  is differentiable at every  $m \in \mathbb{R}^d$  and that,

$$\nabla f_1(m) = -E[\nabla l(X - m)], \quad m \in \mathbb{R}^d.$$

Therefore, the first order conditions given in the above theorem are equivalent to :

$$\begin{cases} \nabla f_0(m^*) + \lambda^* \nabla f_1(m^*) = 0, \\ \lambda^* f_1(m^*) = 0. \end{cases}$$

Furthermore, we also know, thanks to Theorem 28.3 in [Rockafellar \(1970\)](#), that the above conditions are equivalent to saying that  $(m^*, \lambda^*)$  is a saddle point of the Lagrangian associated



to the problem in (1.1), i.e.,

$$L(m, \lambda) := f_0(m) + \lambda f_1(m) = \sum_{i=1}^d m_i + \lambda E[l(X - m)]. \quad (1.2)$$

Under the assumptions of the above theorem  $z^* := (m^*, \lambda^*)$  is the unique solution of  $h(z) = 0$ , where:

$$h(z) := \begin{pmatrix} \lambda E[\nabla l(X - m)] - 1 \\ E[l(X - m)] \end{pmatrix}, \quad z = (m, \lambda) \in \mathbb{R}^d \times [0, \infty[. \quad (1.3)$$

Thus, in order to find the unique risk allocation  $m^*$ , we can look for the zeros of the function  $h$ . We introduce in the section a stochastic algorithm in order to approximate  $z^*$ , with the advantage of being incremental, less sensitive to dimension, and with offer a flexible framework that can be conveniently combined with features such as importance sampling, as discussed in Section 2.5.

## 2 Multivariate Systemic Risk Measures and Stochastic Algorithms

Let  $l$  be a loss function and a vector of losses  $X \in M^\theta$ . In the remainder of the paper, we assume that the hypotheses of Theorem 1.6 are verified by  $l$ , that is:

- ( $\mathcal{A}'_1$ )
- i. For every  $m_0$ ,  $m \mapsto l(X - m)$  is differentiable at  $m_0$  a.s.;
  - ii.  $l$  is permutation invariant;
  - iii.  $m \mapsto E[l(X - m)]$  is strictly convex.

Under ( $\mathcal{A}'_1$ ), Theorem 1.6 ensures that there exists a unique risk allocation  $m^*$  such that  $z^* = (m^*, \lambda^*)$  is the unique root of the function  $h(z) := E[H(X, z)]$ , where we set

$$H(X, z) = \begin{pmatrix} \lambda \nabla_m l(X - m) - 1 \\ l(X - m) \end{pmatrix}, \quad X \in M^\theta. \quad (2.1)$$

Note that the assumptions ( $\mathcal{A}'_1$ ) (i) and (iii) allow us to consider loss functions of interest, which are not necessarily differentiable and/or strictly convex, such as the second example in Section 3.2.

The aim of this section is to construct a stochastic algorithm that converges to the root  $z^* = (m^*, \lambda^*)$  under some suitable assumptions. The standard Robbins-Monro (RM) procedure approximates the function  $h(z) = E[H(X, z)]$  at the  $n + 1$ th step by  $Z_{n+1} = (m_{n+1}, \lambda_{n+1}) = Z_n + \gamma_n H(X_{n+1}, Z_n)$ , with  $X_{n+1}$  a random variable distributed as  $X$ . In order to obtain a.s. convergence of the algorithm toward the unique root  $z^*$ , one crucial condition is the sublinearity of  $h$ , which is very constraining and does not allow us to define

a framework general enough to cover a wide range of loss functions. Consequently, we propose a constrained version of the RM algorithm. The constrained algorithm requires the sequence  $(Z_n)_{n \geq 1}$  to remain in a compact set  $K$ , for non-explosion reasons. Each time  $Z_n$  goes out of  $K$ , it is replaced by its projection on  $K$  leading to the following projected RM algorithm:

$$\begin{aligned} Z_{n+1} &= \Pi_K[Z_n + \gamma_n H(X_{n+1}, Z_n)], \quad Z_0 = z_0 \in K \\ &= \Pi_K[Z_n + \gamma_n h(Z_n) + \gamma_n \delta M_n], \end{aligned} \tag{2.2}$$

where  $\delta M_n = H(X_{n+1}, Z_n) - h(Z_n)$ , and  $(X_n)_{n \geq 1}$  is an i.i.d sequence of random variables with the same distribution as  $X$ , independent of  $Z_0$ . An important point is that the Algorithm (2.2) approximates simultaneously the optimal allocation  $m^*$  and the Kuhn-Tucker coefficient  $\lambda^*$ , which differs from the univariate case of [Dunkel and Weber \(2010\)](#) where an algorithm which approximates  $m^*$  directly can be implemented. We refer to Chapter 4 in [Kushner and Yin \(2003\)](#) for a general introduction on constrained stochastic algorithms.

In the sequel, we denote  $\mathcal{F}_n = \sigma(Z_0, X_i, i \leq n)$ . Note that  $\delta M_n$  is a martingale difference sequence with respect to the filtration  $\mathcal{F} = (\mathcal{F}_n)$ .  $(\gamma_n)_{n \geq 1}$  is assumed to be a deterministic step sequence decreasing to zero and satisfying:

$$\sum_{n \geq 1} \gamma_n = +\infty \text{ and } \sum_{n \geq 1} \gamma_n^2 < +\infty. \tag{2.3}$$

In the following, we will take  $\gamma_n = \frac{c}{n^\gamma}$  where  $c$  is a positive constant and  $\gamma \in ]\frac{1}{2}, 1]$ .

Finally,  $K$  is an hyperrectangle such that  $z^*$  is in the interior of  $K$ :  $K = \{m \in \mathbb{R}^d, a_i \leq m_i \leq b_i\} \times [0, A]$ . This assumption can be difficult to verify in specific examples. However, the projection on the compact set  $K$  can be replaced by an adaptive projection procedure, the so-called projection “à la Chen”, introduced in [Chen and Hu \(1986\)](#), [Chen et al. \(1987\)](#). Formally, given an increasing sequence of compact sets  $(K_n)_n$  that covers  $\mathbb{R}^d$ , each time the sequence  $Z_n$  leaves the current compact set  $K_{\sigma(n)}$ , the compact set is replaced by  $K_{\sigma(n)+1}$  before the next step of the algorithm. [Lelong \(2013\)](#) proved Central Limit Theorems for this type of projection, that are similar to the ones with the projection over one single compact set (see also [Andrieu et al. \(2005\)](#)). For the sake of simplicity, we will use here the classical projection over a fixed compact set  $K$ .

## 2.1 Main Results

We start by stating the two main results regarding the consistency and asymptotic normality of Algorithm (2.2). The proofs of Theorem 2.1 and 2.2 can be found in Section 2.2 and 2.3.

Let  $\sigma^2(\cdot)$ ,  $\Sigma(\cdot)$  and  $m^{2+p}(\cdot)$ , for  $p > 0$ , be defined as follows:

$$\begin{cases} \sigma^2(z) = E[|H(X, z) - h(z)|^2]; \\ m^{2+p}(z) = E[|H(X, z) - h(z)|^{2+p}]; \\ \Sigma(z) = E[(H(X, z) - h(z))(H(X, z) - h(z))^\top]. \end{cases}$$

Recall that Assumption  $(\mathcal{A}'_1)$  ensures the existence of a unique risk allocation  $z^* = (m^*, \lambda^*)$ . We introduce the following classical assumptions for stochastic algorithms in order to obtain the almost sure convergence of  $(Z_n)_n$  to  $z^*$ :

- $(\mathcal{A}_{a.s.})$
- i.  $h$  is continuous on  $K$ ;
  - ii.  $\sup_{z \in K} \sigma^2(z) < \infty$ .

In particular the assumption (i) is satisfied for continuously differentiable loss functions  $l$ , by the Lebesgue dominating convergence theorem. In the case of univariate SR (see e.g. Dunkel and Weber (2010)), the first assumption is also trivial since  $h$  only depends on the continuous loss function  $l$  and not on  $\nabla l$ .

**Theorem 2.1** (Almost sure convergence). *Assume that the sequence  $(Z_n)$  is defined by the algorithm (2.2) and that assumptions  $(\mathcal{A}'_1)$  and  $(\mathcal{A}_{a.s.})$  hold. Then,  $Z_n \rightarrow z^*$   $P$ -almost surely as  $n \rightarrow \infty$ .*

We introduce the following assumptions for the central limit theorem:

- $(\mathcal{A}_{a.n.})$
- i.  $m \mapsto E[\nabla l(X - m)]$  is continuously differentiable. Let  $A := Dh(z^*)$  (Jacobian matrix of  $h$  at  $z^*$ );
  - ii.  $(H(X_{n+1}, Z_n)\mathbf{1}_{|Z_n - z^*| \leq \rho})$  is uniformly integrable for small  $\rho > 0$ ;
  - iii. For some  $p > 0$  and  $\rho > 0$ ,  $\sup_{|z - z^*| \leq \rho} m^{2+p}(z) < \infty$ ;
  - iv.  $\Sigma(\cdot)$  is continuous at  $z^*$ . Let  $\Sigma^* := \Sigma(z^*)$ .

In the particular case when  $X$  is bounded, the assumptions  $(\mathcal{A}_{a.n.})(ii)$ ,  $(\mathcal{A}_{a.n.})(iii)$  and  $(iv)$  are straightforward.

**Theorem 2.2** (Asymptotic Normality). *Recall that the step sequence  $(\gamma_n)_{n \geq 0}$  in (2.2) is defined by  $\gamma_n = \frac{c}{n^\gamma}$ . Assume that  $\gamma \in (\frac{1}{2}, 1)$  and that assumptions  $(\mathcal{A}'_1)$ ,  $(\mathcal{A}_{a.s.})$  and  $(\mathcal{A}_{a.n.})$  hold. Then,*

$$\sqrt{n^\gamma}(Z_n - z^*) \rightarrow \mathcal{N}\left(0, c^2 \int_0^\infty e^{cAt} \Sigma^* e^{cA^\top t} dt\right).$$

If furthermore,  $cA + \frac{I}{2}$  is a Hurwitz matrix and  $cI - P$  is positive definite with  $P$  solution to the Lyapunov's equation:  $A^\top P + PA = -I$ , then for  $\gamma = 1$ ,

$$\sqrt{n}(Z_n - z^*) \rightarrow \mathcal{N}\left(0, c^2 \int_0^\infty e^{(cA + \frac{I}{2})t} \Sigma^* e^{(cA^\top + \frac{I}{2})t} dt\right).$$

**Remark 2.3.**

1. Note that, for convex optimization problems, where the matrix  $A$  is symmetric negative definite, the two additional conditions reduce to the classical condition  $cA + \frac{I}{2}$  is negative definite. Indeed, in this case, the solution of the Lyapunov's equation  $A^\top P + PA = -I$  is simply  $P = -A^{-1}/2$  and the condition  $cI - P$  is positive definite, becomes equivalent to  $cA + \frac{I}{2}$  is negative definite.
2. From a formal point of view, the choice  $\gamma = 1$  gives the best rate of convergence. The asymptotic variance in this case depends on the constant  $c$ . We need to choose it such that  $cA + \frac{I}{2}$  is a Hurwitz matrix and  $cI - P$  is positive definite. Setting  $c$  too small may lead to no convergence at all, while setting it too large, may lead to slower convergence as the effects of large noises early in the procedure might be hard to overcome in a reasonable period of time.
3. The choice of the constant  $c$  is a burning issue. One way to bypass this problem is to premultiply  $A$  by a conditioning matrix  $\Gamma$ , nonsingular, that will make  $A$  close to a constant times the identity. This can be done by considering  $\gamma_n = \Gamma/n$  and we can draw the same conclusions as in Theorem 2.2 as soon as  $\Gamma A + \frac{I}{2}$  is a Hurwitz matrix. This will lead to the following asymptotic behaviour:

$$\sqrt{n}(Z_n - z^*) \rightarrow \mathcal{N}\left(0, \int_0^\infty e^{(\Gamma A + \frac{I}{2})t} \Gamma \Sigma^* \Gamma^\top e^{(A^\top \Gamma^\top + \frac{I}{2})t} dt\right).$$

The optimal choice of the conditioning matrix  $\Gamma$ , which is also called the gain matrix, is the one that will minimize the trace of asymptotic covariance:

$$\int_0^\infty e^{(\Gamma A + \frac{I}{2})t} \Gamma \Sigma^* \Gamma^\top e^{(A^\top \Gamma^\top + \frac{I}{2})t} dt.$$

This is done by taking  $\Gamma = -A^{-1}$  which yields the asymptomatic optimal covariance:  $A^{-1} \Sigma^* (A^{-1})^\top$ .

4. The optimal choice of  $\Gamma$  depends on the function  $h$  and the equilibrium point  $z^*$  which are unknown to us. Adaptive procedures that choose the matrix  $\Gamma$  dynamically by estimating  $Dh(z^*)$  adaptively have been suggested in the literature (see for example [Ruppert \(1991\)](#)), but are generally not as efficient as the Polyak-Ruppert averaging estimators discussed in Section 2.4.

## 2.2 Proof of Theorem 2.1

The proof of the almost sure convergence of  $(Z_n)$  to  $z^*$  rely on the so-called ODE method for the study of constrained RM algorithms. Introduced by Ljung and Kushner, the main idea of the method consists in replacing the analysis of the sequence  $(Z_n)$  by the analysis of continuous-time interpolations of the sequence of stochastic approximations. As detailed in

Theorem 4.8 of Appendix 4.2, these interpolations converge under suitable assumptions to a solution of the projected ODE defined by :

$$\dot{z} = h(z) + C(z), \quad C(z) \in -\mathcal{C}(z), \quad (2.4)$$

where  $\mathcal{C}(z)$  is the convex cone determined by the outer normals to the faces that need to be truncated at  $z$ , and  $C(z)$  is the minimum force needed to bring back  $z$  to  $K$ . In particular,  $C(z(t)) \neq 0$  only for  $z(t) \in \partial K$ . When  $z(t) \in \partial K$  with  $h(z(t))$  directed outside of  $K$ , then  $C(z(t))$  take the “smallest” value in  $-\mathcal{C}(z)$  to maintain  $z(\cdot)$  in  $K$ . More precisely, if  $z \in \mathbb{R}^{d+1}$  and for  $1 \leq i \leq d$ ,  $z_i = b_i$  (resp.  $a_i$ ) and  $h_i(z) > 0$  (resp.  $h_i(z) < 0$ ), then  $C(z) = -h(z_i)$ . We refer to Section 4.3 in Kushner and Yin (2003) for more details on the projected ODE, or to Section 2.2 of Nagurney and Zhang (1995) for a more general presentation of projected ODEs.

Theorem 5.2.1 of Kushner and Yin (2003) (Theorem 4.8) yields that for fixed  $\omega \in \Omega$ , the constrained sequence  $(Z_n(\omega))$  defined by (2.2) converges to a point in the limit set of the projected ODE (2.4). In order to prove Theorem 2.1, the main point is thus to show that the limit set of projected ODE is reduced to  $z^*$ , or in other words, that  $z^*$  is asymptotically stable in the sens of Lyapunov. This point, which is straightforward in the univariate case, is more complicated for MSRM due to the introduction of the Lagrangian function  $L$  from which  $h$  derives.

Since  $z^*$  is in the interior of  $K$  and  $h(z^*) = 0$ ,  $z^*$  is an equilibrium point for the projected ODE (2.4). In order to study the asymptotic stability of the equilibrium  $z^*$ , we define in the following proposition some convenient Lyapunov function  $V$ . A natural and classical choice for this type of problems is  $V(z) = \|z - z^*\|^2$ . Some recalls on Lyapunov stability and Lyapunov functions are presented in Section 4.1 (see also e.g. Slotine and Li (1991) for more details). The following proposition shows that its derivative along any state trajectory is negative semi-definite on  $K$ .

**Proposition 2.4.** *The function  $V(z) = \|z - z^*\|^2$  is such that  $z \rightarrow \dot{V}(z) = \langle \nabla V(z), h(z) + C(z) \rangle$  is negative semi-definite on  $K$  with the respect to the ODE in (2.4).*

Furthermore,

$$\dot{V}(z) \leq -W(z) := 2\langle z - z^*, h(z) \rangle \leq 0, \quad \forall z \in K. \quad (2.5)$$

*Proof.* First, let  $z = (m, \lambda) \in \text{int}(K)$  so that  $\dot{V}(z) = \langle \nabla V(z), h(z) \rangle = 2\langle z - z^*, h(z) \rangle$ , and define  $L$  the Lagrangian as defined in (1.2). We have:

$$\begin{aligned} \langle z - z^*, h(z) \rangle &= \langle m - m^*, \lambda E[\nabla l(X - m)] - \mathbf{1} \rangle + (\lambda - \lambda^*) E[l(X - m)] \\ &= -\langle m - m^*, \partial_m L(m, \lambda) \rangle + (\lambda - \lambda^*) \partial_\lambda L(m, \lambda). \end{aligned}$$

Now, thanks to the convexity of  $L$  with respect to  $m$ , we have:  $L(m^*, \lambda) \geq L(m, \lambda) + \langle m^* -$

$m, \partial_m L(m, \lambda)$ . This in turn implies that

$$L(m^*, \lambda) - L(m, \lambda) + (\lambda - \lambda^*) \partial_\lambda L(m, \lambda) \geq \langle m^* - m, \partial_m L(m, \lambda) \rangle + (\lambda - \lambda^*) \partial_\lambda L(m, \lambda).$$

But, we also have,

$$\begin{aligned} L(m, \lambda) &= \sum m_i + \lambda E[l(X - m)] = \sum m_i + \lambda^* E[l(X - m)] + (\lambda - \lambda^*) E[l(X - m)] \\ &= L(m, \lambda^*) + (\lambda - \lambda^*) \partial_\lambda L(m, \lambda). \end{aligned}$$

The previous inequality becomes then

$$\langle m - m^*, \partial_m L(m, \lambda) \rangle - (\lambda - \lambda^*) \partial_\lambda L(m, \lambda) \geq L(m, \lambda^*) - L(m^*, \lambda)$$

The RHS of the last inequality is non-negative, because,  $(m^*, \lambda^*)$  is a saddle point, that is  $L(m^*, \lambda) \leq L(m^*, \lambda^*) \leq L(m, \lambda^*)$ . Moreover, because  $L$  is strictly convex with respect to  $m$ , it is also negative if  $m \neq m^*$ . Therefore, we get that,

$$\langle z - z^*, h(z) \rangle \leq 0. \quad (2.6)$$

Note that this is true irrespective of whether  $z \in \text{int}(K)$  or not.

Let  $z \in \partial K$ . If  $z_i = b_i$  (resp.  $a_i$ ) and  $h_i(z) \geq 0$  (resp.  $h_i(z) \leq 0$ ) for some  $i$ , then  $C(z_i) = -h_i(z)$ . Since  $z^* \in \text{int}(K)$ ,  $a_i < z_i^* < b_i$  and thus  $(z_i - z_i^*)C(z_i) \leq 0$ . Hence,  $\dot{V}(z) = 2\langle z - z^*, h(z) + C(z) \rangle \leq \langle z - z^*, h(z) \rangle \leq 0$ , which achieves the proof.  $\square$

**Remark 2.5.** *We cannot conclude that  $\dot{V}$  is negative definite on  $K$ , because  $z \neq z^*$  does not imply that  $m \neq m^*$ . Besides, if  $z = (m^*, \lambda)$  such that  $\lambda \neq \lambda^*$ , we have  $\dot{V}(z) = 0$  and  $z \neq z^*$ .*

As stated in the previous remark,  $\dot{V}$  is not negative definite, and thus the standard Liapunov theorem (see e.g. Theorem 3.3 in Slotine and Li (1991)) cannot be applied. In order to prove the asymptotic stability of  $z^*$ , we apply LaSalle's invariant principle. The invariant set theorem characterizes the limit set of the ODE as the largest invariant set in  $\{z; \dot{V}(z) = 0\}$ . Note that, due to projection on  $K$ , the r.h.s of the ODE 2.4 is discontinuous. Versions of Lasalle's invariance principle for very general nonsmooth ODEs can be found in Fischer et al. (2013) (see also e.g. Filippov (1988), Shevitz and Paden (1994)). In the case of (2.4), the direct proof presented below is facilitated by the fact that  $K$  is a compact set.

**Proposition 2.6.** *The equilibrium point  $z^*$  of the ODE (2.4) is asymptotically stable.*

*Proof.* Let us first show that limit points of the ODE (2.4) are in the set  $R = \{z; W(z) = 0\}$ . This can be viewed as special case of Corollary 1 in Fischer et al. (2013). Let  $(z(t))$  be a solution to the projected ODE (2.4) with initial condition  $z(0) \in K$ .

The first step of the proof is proving that  $W(z(t)) \rightarrow 0$  when  $t \rightarrow \infty$ , with  $W$  the scalar function defined by  $W(z) = 2\langle z - z^*, h(z) \rangle$ , by applying the Barbalat's Lemma 4.6. By

Proposition 2.4,  $\dot{V} \leq 0$ . Hence,  $t \rightarrow V(z(t))$  is decreasing and

$$\int_0^t W(z(s))ds \leq - \int_0^t \dot{V}(z(s))ds = V(z(0)) - V(z(t)). \quad (2.7)$$

Since  $V(z(t)) \geq 0$ , the l.h.s of the previous inequality is increasing and bounded by  $V(z(0))$ , and thus  $\lim_{t \rightarrow \infty} \int_0^t W(z(s))ds < \infty$ . Since  $z$  is absolutely continuous and  $h$  is bounded on  $K$ ,  $z$  is uniformly continuous. Finally, since  $W$  is a continuous function, it is uniformly continuous on the compact  $K$  and thus  $t \rightarrow W(z(t))$  is uniformly continuous. By Barbalat's lemma, we can conclude that

$$\lim_{t \rightarrow \infty} W(z(t)) = 0.$$

Thus, limit points of the ODE (2.4) are in the set  $R = \{z; W(z) = 0\}$ .

Since solutions of (2.4) are in the bounded domain  $K$ , the limit set of the projected ODE is an invariant set (see e.g. Section 4.3 of Kushner and Yin (2003) or Lemma 12.5 in Filippov (1988)). It thus remains to show that  $\{z^*\}$  is the only invariant set in  $R$ . Let  $M$  be the largest invariant set in  $R$ , and  $z = (m, \lambda) \in M$ . As discussed in the proof of Proposition 2.4, if  $(m, \lambda) \in R$ , then necessarily  $m = m^*$ . Thus, since  $M$  is an invariant set in  $R$ , if  $z(0) = (m, \lambda) \in M$ , then  $z(t) = (m^*, \lambda(t)) \in R$  for all  $t \geq 0$ , with  $z(\cdot)$  solution of the following ODE,

$$\begin{cases} 0 = \lambda(t)E[\nabla l(X - m^*)] - \mathbf{1} + C(z(t))_m, & t \geq 0, \\ \frac{d\lambda(t)}{dt} = E[l(X - m^*)] + C(z(t))_\lambda, & t \geq 0. \end{cases} \quad (2.8)$$

Now, since  $z^* = (m^*, \lambda^*) \in \text{int}(K)$ , we get that,  $C(z(t))_m = C((m^*, \lambda(t)))_m = 0$ . We obtain that  $\lambda(t)E[\nabla_m l(X - m^*)] - \mathbf{1} = 0$ , and consequently  $t \rightarrow \lambda(t)$  is a constant function equal to  $\lambda^*$ .

We have then showed that the limit set is simply  $\{z^*\}$  and therefore  $z^*$  is asymptotically stable equilibrium for the ODE (2.4).  $\square$

*Proof of Theorem 2.1.* By Proposition 2.6,  $z^*$  is the only point in the limit set of ODE (2.4). Thus, if the assumptions 4.7 (1)-(3) are verified, we can conclude that  $Z_n \rightarrow z^*$  as  $n \rightarrow \infty$  by Theorem 4.8 in the Appendix. The first assumption (1) of the theorem is exactly the Assumption ( $\mathcal{A}_{a.s.}$ )-(ii). Assumption (2) corresponds to the continuity of  $h$ , i.e. Assumption ( $\mathcal{A}_{a.s.}$ )-(i). Finally, Assumptions (3) corresponds to the hypothesis (2.3) on the step sequence  $(\gamma_n)$ , which concludes the proof.  $\square$

## 2.3 Proof of the asymptotic normality

The proof of Theorem 2.2 is obtained as an application of Theorem 4.10 in the Appendix (Theorem 10.2.1 in Kushner and Yin (2003)), with  $g = ch$  and  $Y_n = cH(X_{n+1}, Z_n)$  and  $\epsilon_n = \frac{1}{n^\gamma}$ .

*Proof of Theorem 2.2.* By the proof of Theorem 2.1, Assumption 4.7 is satisfied. Let us prove that Assumption 4.9 holds.

First, let us start with the case  $\gamma \in (\frac{1}{2}, 1)$ . Then Assumption 4.9.1 is true since  $\mu_n = 1 + \frac{\gamma}{2n} + o(\frac{1}{n^2})$ . Assumption 4.9.2 corresponds to the assumption  $(\mathcal{A}_{a.n.})$ -ii..

Let  $(Z^n(\cdot))$  be the sequence of continuous time processes defined by (4.2). For all  $t \geq 0$  and  $n \in \mathbb{N}$ ,  $Z^n(t) \in \{Z_p; p \geq n\} \subset K$ . Hence, the sequence  $(Z^n(\cdot))$  is uniformly bounded. Furthermore, since  $(Z_n)$  converges a.s. to  $z^*$  by Theorem 2.1, Aldous' criterion for tightness (see e.g. Theorem VI.4.5 in Jacod and Shiryaev (2013)) is satisfied and hence the family processes  $(Z^n(\cdot))$  is tight. Finally, for all  $t \geq 0$ ,  $Z^n(t) = Z^0(t_n + t)$  converges a.s. to the constant  $z^*$  by definition of the interpolation process  $Z^0$ . Thus,  $(Z^n(\cdot))$  converges weakly to  $z^*$ , and Assumption 4.9.3 holds.

Assumption 4.9.4 corresponds to  $(\mathcal{A}_{a.n.})$ -i.. The first part of 4.9.6 is  $(\mathcal{A}_{a.n.})$ -iii.. By  $(\mathcal{A}_{a.n.})$ -iv. (continuity of  $\Sigma$  at  $z^*$ ) and since  $Z_n \rightarrow z^*$  a.s.,

$$E[(Y_n - g(Z_n))(Y_n - g(Z_n))^{\top} \mathbf{1}_{|Z_n - z^*| \leq \rho}] \rightarrow \Sigma_1 := c^2 \Sigma^*.$$

Finally, 4.9.7 holds for  $\epsilon_n = \frac{1}{n^\gamma}$  (see e.g. p.341 in Kushner and Yin (2003)).

It remains to show that 4.9.5 holds, that is the matrix  $A' = cA$  is a Hurwitz matrix. We have:

$$A = \left( \begin{array}{c|c} \lambda^* DE[\nabla l(X - m^*)] & E[\nabla l(X - m^*)] \\ \hline -E[\nabla l(X - m^*)] & 0 \end{array} \right) = - \left( \begin{array}{c|c} \hat{A} & -\frac{1}{\lambda^*} \\ \hline \frac{1}{\lambda^*} & 0 \end{array} \right),$$

where  $\hat{A} := -\lambda^* DE[\nabla l(X - m^*)]$  corresponds to the second derivative of the Lagrangian  $L$  with respect to  $m$ . Note that  $L$  is strictly convex with respect to  $m$  due to the strict convexity of  $m \mapsto E[l(X - m)]$ . This implies that  $\hat{A}$  is a positive definite matrix. Thanks to Theorem 3.6 in Benzi et al. (2005), we deduce that  $A$  (and thus  $A'$ ) is a Hurwitz matrix.

For the case  $\gamma = 1$ , it remains to prove Assumption 4.9.7, ie that  $\lambda > 1$  where  $\lambda$  is the largest real number such that  $I \geq \lambda P'$ , with  $P'$  the solution of the Lyapunov equation  $(A')^{\top} P' + P' A' = -I$ . Recalling  $A' = cA$ ,  $P'$  is also solution of  $A^{\top} cP' + cP' A = -I$ . By uniqueness of the Lyapunov equation's solution, we obtain that  $P = cP'$ , with  $P$  as defined in Theorem 2.2. Since  $cI - P = c(I - P')$  is positive definite by assumption,  $I - P'$  is positive definite and thus  $\lambda > 1$ , which achieves the proof.  $\square$

## 2.4 Polyak-Ruppert Averaging

In order to ease the tuning of the step parameter which is known to monitor the numerical efficiency of RM algorithms, we are led to modify our algorithm and to use an averaging procedure. Averaging algorithms were introduced by Ruppert (see Ruppert (1991)) and Polyak (see Polyak and Juditsky (1992)) and then widely investigated by many authors. Kushner



and Yin (2003) and Kushner and Yang (1995) studied these algorithms in combination with projection and proved a Central Limit Theorem (CLT) for averaging constrained algorithms. The following theorem describes the Polyak-Ruppert algorithm for MSRM and states its asymptotic normality. It is a direct consequence of theorem 11.1.1 in Kushner and Yin (2003).

**Theorem 2.7.** *Assume  $\gamma \in (\frac{1}{2}, 1)$  and that assumptions  $(\mathcal{A}'_l)$ ,  $(\mathcal{A}_{a.s.})$  and  $(\mathcal{A}_{a.n.})$  hold. For any arbitrary  $t > 0$ , we define*

$$\bar{Z}_n = \frac{\gamma_n}{t} \sum_{i=n}^{n+t/\gamma_n-1} Z_i, \quad (2.9)$$

where any upper summation index  $u \in \mathbb{R}^+$  is interpreted as its integer part. If  $\Sigma^*$  is positive definite, then we have the following CLT:

$$\sqrt{\frac{t}{\gamma_n}} (\bar{Z}_n - z^*) \xrightarrow[n \rightarrow \infty]{} \mathcal{N}\left(0, V + O\left(\frac{1}{t}\right)\right), \quad (2.10)$$

where  $V = A^{-1}\Sigma^*(A^{-1})^\top$ .

**Remark 2.8.**

1. In (2.9), the window of averaging is  $t/\gamma_n$  for any arbitrary real  $t > 0$ . Equivalently,  $\gamma_n \times$  (size of window) does not go to infinity as  $n \rightarrow \infty$ , hence the name “minimal window” of averaging. In contrast, the “maximal window” of averaging corresponds to a window size  $q_n t$  such that  $\gamma_n q_n \rightarrow \infty$ . A natural and a classical choice is taking  $\gamma_n = c/n^\gamma$  and  $q_n = n$ . In the case of maximal window of averaging, under some extra conditions, we are able to achieve the optimal asymptotic variance without an extra term  $O(1/t)$  (see Theorem 11.3.1 in Kushner and Yin (2003)).
2. Two sided averages can also be used instead of the one-sided average in (2.9).

**Estimator of asymptotic variance** The previous CLT theorems assert that, under some suitable conditions, our RM and PR algorithms converge to the root  $z^*$  with a corresponding rate. More specifically, in Theorem 2.7, the asymptotic variance  $V$  depends on  $\Sigma^*$  and  $A$ . In practice, these two quantities are unknown and need to be approximated in order to derive confidence intervals for our estimators. In Theorem 2.2, in both cases,  $\gamma = 1$  and  $\gamma \in (\frac{1}{2}, 1)$ , the asymptotic variance is expressed as an infinite integral that involves  $\Sigma^*$  and  $A$ . The numerical evaluation of these integrals is a non-trivial exercise even when  $\Sigma^*$  and  $A$  are known. In Hsieh and Glynn (2002), they described an approach that produces confidence regions and that avoids the necessity of having to explicitly estimate these integrals.

In the following proposition, we provide consistent estimators of these two quantities. The proof relies mainly on the Martingale Convergence Theorem.

**Proposition 2.9.** *Assume  $(\mathcal{A}'_I)$ ,  $(\mathcal{A}_{a.s.})$  and  $(\mathcal{A}_{a.n.})$  hold. If  $z \rightarrow E[\|H(X, z)\|^4]$  is locally bounded around  $z^*$ , then,*

$$\Sigma_n := \frac{1}{n} \sum_{k=1}^n H(X_k, Z_{k-1})H(X_k, Z_{k-1})^\top \rightarrow \Sigma^* \text{ a.s.} \quad (2.11)$$

Let  $A_n^\epsilon$  the matrix whose elements  $A_n^\epsilon(i, j)$  for  $i, j \in \{1, \dots, d+1\}$  are defined as follows:

$$A_n^\epsilon(i, j) := \frac{1}{\epsilon n} \sum_{k=1}^n H_i(X_k, Z_{k-1} + \epsilon e_j) - H_i(X_k, Z_{k-1}),$$

then,

$$\lim_{\epsilon \rightarrow 0} \lim_{n \rightarrow \infty} A_n^\epsilon = A \text{ a.s.} \quad (2.12)$$

*Proof.* Let  $(S_n)_{n \in \mathbb{N}^*}$  be the sequence defined as:

$$S_n = H(X_n, Z_{n-1})H(X_n, Z_{n-1})^\top - \Sigma(Z_{n-1}) - h(Z_{n-1})h(Z_{n-1})^\top, \quad n \geq 1$$

$(S_n)_{n \in \mathbb{N}^*}$  is a martingale difference sequence adapted to  $\mathcal{F}$  and consequently the following sequence  $(M_n)_{n \in \mathbb{N}^*}$  defined as:

$$M_n = \sum_{i=1}^n \frac{S_i}{i}, \quad n \geq 1,$$

is a  $\mathcal{F}$ -martingale. Moreover, since  $Z_n \rightarrow z^*$  a.s.,  $h$  and  $\Sigma$  are continuous in  $z^*$  (Assumptions  $(\mathcal{A}_{a.s.})$ -i. and  $(\mathcal{A}_{a.n.})$ -iv.), and by the boundedness of  $z \rightarrow E[\|H(X, z)\|^4]$  around  $z^*$ , we have that:

$$\sup_{n \geq 1} E[\|S_n\|^2 | \mathcal{F}_{n-1}] < \infty \text{ a.s.}$$

Thus, the martingale convergence theorem ensures the existence of a finite random variable  $M_\infty$  such that  $M_n \rightarrow M_\infty$  a.s. Kronecker's lemma then guarantees that  $\frac{1}{n} \sum_{i=1}^n S_i \rightarrow 0$ . Now, since,

$$\Sigma_n = \frac{1}{n} \sum_{i=1}^n S_i + \frac{1}{n} \sum_{i=1}^n \Sigma(Z_{i-1}) + \frac{1}{n} \sum_{i=1}^n h(Z_{i-1})h(Z_{i-1})^\top,$$

we deduce that  $\Sigma_n \rightarrow \Sigma^*$ .

The proof of (2.12) follows using the same arguments above.  $\square$

**Remark 2.10.**

1. *Instead of averaging on all observations, one could modify the estimators above and average only on recent ones. This might improve the behaviour of these estimators.*
2. *If we denote  $V_n := (A_n^\epsilon)^{-1} \Sigma_n ((A_n^\epsilon)^{-1})^\top$ , then we obtain an approximate confidence interval for PR estimator with a confidence of  $1 - \alpha$  in the following form:*

$$\left[ \bar{Z}_{j,n} - \sqrt{\frac{V_{jj,n}}{tn^\gamma}} q_\alpha, \bar{Z}_{j,n} + \sqrt{\frac{V_{jj,n}}{tn^\gamma}} q_\alpha \right], \quad j \in \{1 \dots d\}, \quad \gamma \in (0, 1), \quad (2.13)$$

where  $q_\alpha$  is the  $1 - \frac{\alpha}{2}$  quantile of a standard normal random variable. Note that this confidence interval has the advantage of being obtained with one simulation run. For RM estimators, confidence intervals could be estimated empirically.

## 2.5 Variance reduction with adaptive importance sampling (IS)

The convergence of the stochastic algorithms presented in this section can be improved by implementing variance reduction techniques such as important sampling (IS). The original idea of IS for the Monte-Carlo computation of an expectation  $E[G(X)]$ , is to find a change of probability measure  $P'$ , such that  $E[G(X)] = E'[G'(X)]$  and  $\text{Var}_{P'}(G'(X)) < \text{Var}(G(X))$ . The main issue is then to find the right change of measure in order to reduce the variance. A possible parametric approach is to look for the optimal parameter  $\theta^*$  minimizing the variance, among a family of distribution  $(P_\theta)$ . Under some assumptions, the minimization problem can be reinterpreted as finding the root of a function written as an expectation. This is the main idea of recursive IS by stochastic algorithm, where the parameter  $\theta^*$  is estimated by a RM algorithm.

The Esscher transform (or exponential tilting) is a natural tool for parametric IS, used extensively in actuarial science and for variance reduction. This transform is a natural change of measure since it corresponds to a translation in the gaussian case, and a change of intensity for Poisson processes. Let us first recall the main idea behind the Esscher transform and recursive IS in the standard case. For more details, see e.g. see e.g. [Arouna \(2004\)](#), [Kawai \(2009\)](#), or [Lemaire and Pagès \(2010\)](#).

**Esscher transform and parametric IS** In the following, we assume that the  $\mathbb{R}^d$ -valued random variable  $X$  has a density  $f$ . For  $\theta \in \mathbb{R}^d$ , let  $\psi$  be the log-Laplace transform of  $X$ , assumed to be finite:

$$\psi(\theta) = \log (E[\exp(\langle \theta, X \rangle)]) < \infty.$$

Then,  $\psi$  is convex and infinitely differentiable. The Esscher transform is defined by the following change of measure introducing the density  $g_\theta$ :

$$g_\theta(x) = \exp(\langle \theta, x \rangle - \psi(\theta)) f(x). \tag{2.14}$$

In the following we denote by  $X^{(\theta)}$  the r.v. of density  $g_\theta$ . The expectation can be replaced by

$$E[G(X)] = E[G(X^{(\theta)}) \exp(-\langle X^{(\theta)}, \theta \rangle + \psi(\theta))].$$

The main idea of the recursive IS procedure is to choose  $\theta$  in order to minimize the variance of  $G(X^{(\theta)}) \exp(-\langle X^{(\theta)}, \theta \rangle + \psi(\theta))$ , which boils down to minimizing the function

$$E[G^2(X^{(\theta)}) \exp(-2\langle X^{(\theta)}, \theta \rangle + 2\psi(\theta))] = E[G^2(X) \exp(-\langle X, \theta \rangle + \psi(\theta))].$$

In other words, the optimal parameter is

$$\theta^* = \operatorname{argmin} E[G^2(X, z) \exp(-\langle X, \theta \rangle + \psi(\theta))]. \quad (2.15)$$

Under some regularity assumptions, the minimization problem can be solved using stochastic algorithms.

**Adaptative RM algorithm with IS** Our setting of multivariate systemic risk measures is more complex. For  $1 \leq i \leq d + 1$ , let  $H_i$  be the  $i$  component of the function  $H$  defined by (2.1). Then, for  $z = (m, \lambda) \in \mathbb{R}^d \times [0, \infty[$  and  $\theta \in \mathbb{R}^d$ ,

$$h_i(z) = E[H_i(X, z)] = E[H_i(X^{(\theta)}, z) \exp(-\langle X^{(\theta)}, \theta \rangle + \psi(\theta))], \quad (2.16)$$

Let  $V = (V_i)_{1 \leq i \leq d}$ , with  $V_i$  the function defined for  $z \in \mathbb{R}^d \times [0, \infty[$  and  $\theta \in \mathbb{R}^d$  by

$$\begin{aligned} V_i(\theta, z) &= E[H_i^2(X^{(\theta)}, z) \exp(-2\langle X^{(\theta)}, \theta \rangle + 2\psi(\theta))] \\ &= E[H_i^2(X, z) \exp(-\langle X, \theta \rangle + \psi(\theta))]. \end{aligned} \quad (2.17)$$

In order to reduce the asymptotic variance obtained in Theorem 2.7, it is natural to take  $\theta^i = \theta^{i,*}$  in (2.16) such that

$$\theta^{i,*} = \operatorname{argmin} V_i(\theta, z^*) = \operatorname{argmin} E[H_i^2(X, z^*) \exp(-\langle X, \theta \rangle + \psi(\theta))]. \quad (2.18)$$

Assume that for each  $1 \leq i \leq d$ ,

$$E[|X| H_i^2(X, z) \exp(-\langle X, \theta \rangle + \psi(\theta))] < \infty, \quad \forall z \in \mathbb{R}^d \times [0, \infty[, \theta \in \mathbb{R}^d.$$

Then,  $V(\cdot, z)$  is differentiable and for  $1 \leq i \leq d$ ,

$$\nabla V_i(\theta, z^*) = E[(\nabla \psi(\theta) - X) H_i^2(X, z^*) \exp(-\langle \theta, X \rangle + \psi(\theta))]. \quad (2.19)$$

Lemaire and Pagès (2010) (Proposition 3, see also Bardou et al. (2009)) give sufficient conditions for  $V_i(\cdot, z^*)$  to be convex and for the set  $\{\theta \in \mathbb{R}^d; \nabla V_i(\theta, z^*) = 0\}$  to be non empty. We also refer to Kawai (2009) in the special case when  $X = \xi_T$  with  $\xi$  a Levy process. A RM algorithm as in (2.2) can be implemented to find the root of each gradient  $\nabla V_i$ , replacing  $H_i$  with

$$F_i(X, z^*, \theta) = (\nabla \psi(\theta) - X) H_i^2(X, z^*) \exp(-\langle \theta, X \rangle + \psi(\theta)). \quad (2.20)$$

However such an algorithm cannot be implemented in practice, since  $z^*$  is unknown. The same issue appears in the case of the VaR/CVaR estimation in [Bardou et al. \(2009\)](#). Hence, the authors introduced an adaptive procedure combining the RM algorithm and recursive Importance Sampling, by substituting the unknown parameters at the  $n$ th step the by their estimation at the previous step. This idea can be adapted to our framework of multivariate systemic risk measures.

Let us first introduce some notations. For  $\theta = (\theta^1, \dots, \theta^d) \in \mathbb{R}^{d(d+1)}$ , let  $\tilde{H}$  be the function replacing  $H$  in (2.16), defined by

$$\tilde{H}(X, z, \theta) = (H_i(X, z) \exp(-\langle X, \theta^i \rangle + \psi(\theta^i)))_{\{1 \leq i \leq d+1\}},$$

and

$$F(X, z, \theta) = (F_i(X, \theta^i, z))_{\{1 \leq i \leq d+1\}},$$

where  $F_i$  has been defined in (2.20). Finally, we denote by  $X^{(\theta)} = (X^{(\theta^1)}, \dots, X^{(\theta^d)})$  the  $\mathbb{R}^{d^2}$ -valued random vector, where each  $\mathbb{R}^d$  component  $X^{(\theta^i)}$  has the density  $g_{\theta^i}$  and is independent of the others. Finally, let  $\Sigma(z, \theta)$  be the covariance matrix of  $\tilde{H}(X, z, \theta)$ .

The new stochastic algorithm is defined as follow:

$$\begin{aligned} Z_{n+1} &= \Pi_K[Z_n + \gamma_n \tilde{H}(X_{n+1}^{(\theta_n)}, Z_n, \theta_n)], \\ \theta_{n+1} &= \Pi_{K'}[\theta_n + \gamma_n F(X_{n+1}, Z_n, \theta_n)], \end{aligned} \tag{2.21}$$

where  $K'$  is an hyper rectangle containing 0 and  $\theta^*$ . In the specific case of the VaR-CVaR approximation, [Bardou et al. \(2009\)](#) propose an unconstrained version of the algorithm, based on [Lemaire and Pagès \(2010\)](#). However, the procedure cannot be applied here since  $H$  does not verify the sublinearity condition, as discussed in Section 2. However, as for  $K$ ,  $K'$  can also be replaced by an increasing sequence of compact  $(K'_{\sigma(n)})_{n \in \mathbb{N}}$ .

The proof of Theorem 2.1 and 2.2 can be adapted to obtain the a.s. convergence and asymptotic normality of the approximations  $(Z_n, \theta_n)_n$  introduced in (2.21), as well as its Polyak-Ruppert counterpart. In particular, the covariance matrix  $\Sigma^* = \Sigma(z^*, 0)$  will be replaced by  $\Sigma(z^*, \theta^*)$ , minimizing the variance of the vector's  $\tilde{H}(X, z^*, \cdot)$  components.

The asymptotic variance for the Polyak-Ruppert estimator with adaptative IS is actually  $V(\theta^*) = A^{-1} \Sigma(z^*, \theta^*) (A^{-1})^\top$ , with  $A$  the Jacobian matrix of  $h$  at  $z^*$ . The matrix  $A$  does not depend on the  $\theta^*$ , but the diagonal elements of the covariance matrix  $V$  might not minimize the function  $V(\theta) = A^{-1} \Sigma(z^*, \theta^*) (A^{-1})^\top$ . In the case of the joint computation of the VaR/CVaR, the expression of the matrix  $A^{-1}$  is explicit, and this might be the case for MSRM for specific loss functions. Another potential drawback of the adaptative RM with IS (2.21) is the number  $d(d+1)$  of additional parameters introduced by the variance reduction procedure. Once again, the number of parameters could be significantly reduced depending

on the loss function  $l$ .

**Remark 2.11.** *In dimension 1, it is straightforward to show that  $\theta^*$  minimizes  $V(\theta)$ . Hence the adaptative algorithm (2.21) also provides a general variance reduction method for univariate shortfall risk measures. This is very useful when the optimal change of measure is not known beforehand, which is the case most of the time.*

### 3 Numerical examples

In this section, we test the performance of the proposed stochastic algorithms schemes for MSRM. We provide numerical examples showing the consistency properties of the different estimators and then their normal asymptotic behaviour, with (PR algorithm) and without averaging (RM algorithm). Two examples are considered. In the first one, we consider a loss function of an exponential type coupled with a normal distribution. This example is relevant for our numerical analysis as we can explicitly express the optimal allocations in a closed form. In the second example, we consider a loss function that involves positive part function with a Gaussian and a compound Poisson distributions. All the algorithms have been implemented on a standard computer using the programming language Python, and are publicly available at <https://github.com/AchrafTamtalini/SRM/>.

In the following,  $n$  will denote the number of steps in one simulation run and  $N$  the number of simulations. We introduce the following sequences:

$$\bar{D}_n := \sqrt{tn^\gamma}(\bar{Z}_n - z^*) \quad \gamma \in \left(\frac{1}{2}, 1\right), \quad (3.1)$$

$$D_n := \sqrt{n^\gamma}(Z_n - z^*), \quad \gamma \in \left(\frac{1}{2}, 1\right]. \quad (3.2)$$

#### 3.1 Toy example

As a first simple example, we will consider a exponential loss function of the following form:

$$l(x_1, \dots, x_d) = \frac{1}{1 + \alpha} \left[ \sum_{i=1}^d e^{\beta x_i} + \alpha e^{\beta \sum_{i=1}^d x_i} \right] - \frac{\alpha + d}{\alpha + 1} \quad (3.3)$$

We will set  $d = 2$  and consider a bivariate normal vector  $X = (X_1, X_2) \sim \mathcal{N}(0, M)$  with  $M = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$ .  $\alpha$  is a systemic weight parameter taken to be non negative and  $\beta > 0$  is the risk aversion coefficient. In this case, we can explicitly solve the first order conditions and derive closed formulas for optimal allocations. This will be useful to test our algorithms.

**Lemma 3.1.** *Let  $l$  be the loss function defined by (3.3), and  $X$  the vector of random losses*

as above. Then, the optimal risk allocation is:

$$m_i^* = \begin{cases} \frac{\beta\sigma_i^2}{2}, & \text{if } \alpha = 0, \\ \frac{\beta\sigma_i^2}{2} + \frac{1}{\beta} \text{SRC}(\rho, \sigma_1, \sigma_2, \alpha, \beta), & \text{if } \alpha > 0. \end{cases} \quad (3.4)$$

With

$$\text{SRC}(\rho, \sigma_1, \sigma_2, \alpha, \beta) = \ln \left( \frac{\alpha e^{\rho\beta^2\sigma_1\sigma_2}}{-1 + \sqrt{1 + \alpha(\alpha + 2)e^{\rho\beta^2\sigma_1\sigma_2}}} \right).$$

This shows that, in the case  $\alpha > 0$ , the risk allocations are disentangled into two components: an individual contribution  $\frac{\beta\sigma_i^2}{2}$  and the *Systemic Risk Contribution* term (SRC). Note that taking  $\alpha \rightarrow 0$  makes the SRC null since the systemic weight  $\alpha$  is responsible for the systemic contribution in the loss function  $l$ . One can also show, by easy calculations, that the SRC is increasing with respect to  $\rho$ : the higher the correlation is, the more costly the acceptable monetary allocations are. This could be explained by the fact that, with a higher correlation between the two components, the losses of one will induce the loss of the other and consequently the system will become riskier. Note also that we could also express in a closed form the Jacobian matrix  $A$  and  $\Sigma^*$ .

*Proof.* The optimal risk allocations  $m_i^*$  are characterized by the first order conditions given in Theorem 1.6, i.e.

$$\begin{cases} \frac{\lambda^*}{1 + \alpha} E \left[ e^{\beta(X_i - m_i^*)} + \alpha e^{\beta(X_1 + X_2 - m_1^* - m_2^*)} \right] = 1, & i = 1, 2, \\ \frac{1}{1 + \alpha} E \left[ e^{\beta(X_1 - m_1^*)} + e^{\beta(X_2 - m_2^*)} + \alpha e^{\beta(X_1 + X_2 - m_1^* - m_2^*)} \right] - \frac{\alpha + 2}{\alpha + 1} = 0. \end{cases}$$

The two first equations imply that  $E[e^{\beta(X_1 - m_1^*)}] = E[e^{\beta(X_2 - m_2^*)}]$ , which in turn gives that,

$$\frac{\beta\sigma_1^2}{2} - m_1^* = \frac{\beta\sigma_2^2}{2} - m_2^*. \quad (3.5)$$

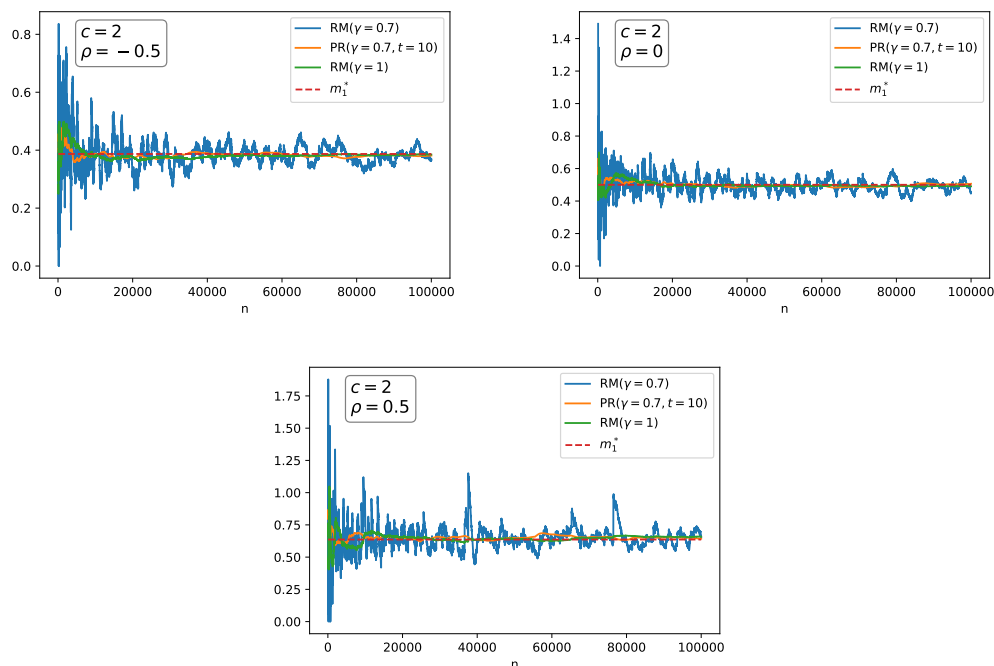
The third equation gives  $e^{\frac{\beta^2\sigma_1^2}{2} - \beta m_1^*} + e^{\frac{\beta^2\sigma_2^2}{2} - \beta m_2^*} + \alpha e^{\rho\beta^2\sigma_1\sigma_2 + \frac{\beta^2}{2}(\sigma_1^2 + \sigma_2^2) - \beta(m_1^* + m_2^*)} = 2 + \alpha$ . Thanks to (3.5) and denoting  $Q = e^{\frac{\beta^2\sigma_1^2}{2} - \beta m_1^*}$ , we get that,  $\alpha e^{\rho\beta^2\sigma_1\sigma_2} Q^2 + 2Q - (2 + \alpha) = 0$ . Taking the positive solution of the last equation gives  $Q = \frac{-1 + \sqrt{1 + \alpha(\alpha + 2)e^{\rho\beta^2\sigma_1\sigma_2}}}{\alpha e^{\rho\beta^2\sigma_1\sigma_2}}$ . Now, denoting by  $\text{SRC} = -\log(Q)$ , we obtain that  $m_i^* = \frac{\beta\sigma_i^2}{2} + \frac{1}{\beta} \text{SRC}$ .  $\square$

In all this example, we fix  $\alpha = 1$ ,  $\beta = 1$  and  $\sigma_1 = \sigma_2 = 1$ . With  $\rho \in \{-0.5, 0, 0.5\}$ , we obtain the exact values in the table below. Note that since we have  $X_1 \sim X_2 \sim \mathcal{N}(0, 1)$  and  $l$  is permutation invariant, it follows that  $m_1^* = m_2^*$ .

Table 1: Exact optimal risk allocations.

| $\rho$ | $m_1^* = m_2^*$ |
|--------|-----------------|
| -0.5   | 0.3868          |
| 0      | 0.5             |
| 0.5    | 0.6364          |

For RM/PR algorithms, we used a number of steps  $n = 10^5$ . As for the compact  $K$ , we took  $K = [0, 2]^3$  and  $Z_0$  was taken uniformly on  $K$ . We run the different algorithms for  $\gamma = 1$  and  $\gamma = 0.7$ . We chose an averaging parameter  $t = 10$  and we set  $c = 2$  in a first step. Figure 1 shows that, for different values of  $\rho \in \{-0.5, 0, 0.5\}$ , our RM algorithm with  $\gamma = 1$  converges relatively quickly to the optimal allocations, whereas when  $\gamma = 0.7$ , noise is still persisting. This is due to the step parameter  $c$  as discussed in the previous section. In order to get a smoother numerical behaviour, two solutions are available to us: either we use PR averaging (c.f. Figure 1), or we reduce the value of the parameter  $c$  (Figure 2).

Figure 1: Consistency of RM/PR estimators with for different values of  $\rho$ .



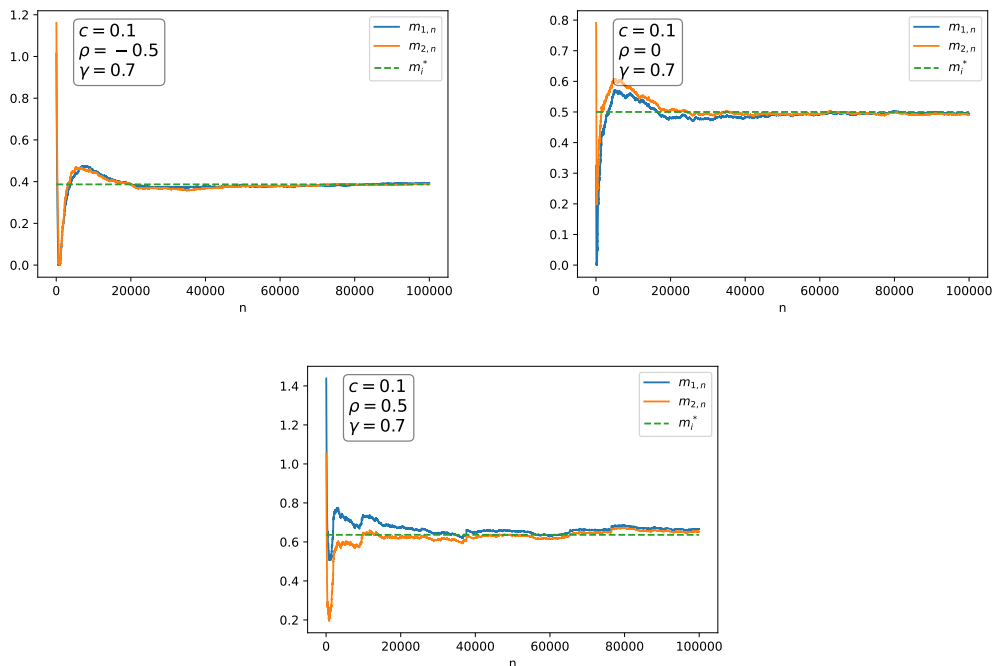


Figure 2: Consistency of RM estimators with  $c = 0.1$  for different values of  $\rho$ .

Figure 3 shows the increase of computational time when the error become smaller. We have also compared in Table 2 the computational time (CT) of the RM algorithms (with  $c = 2$  and  $\gamma = 1$ ) with the Monte Carlo method used in Armenti et al. (2018). As it is often the case, the RM algorithm is slower than the Monte Carlo method in Armenti et al. (2018). However, as stated above, the latter does not provide any convergence results, and might not converge in practical applications.

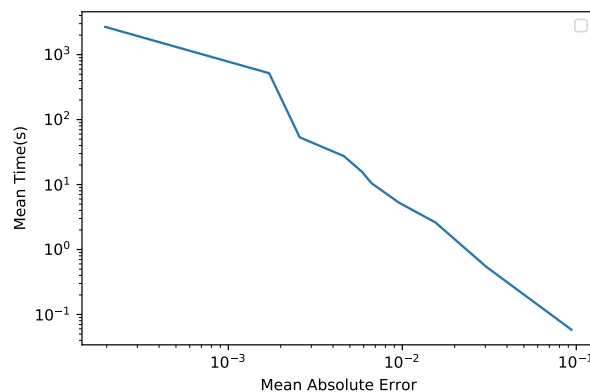


Figure 3: Computational time (in seconds) as a function of the (average) absolute error (log log scale), for the RM algorithm ( $\gamma = 1$ ,  $c = 2$  and  $\rho = 0.5$ ).

| Absolute error     | CT RM (s) | CT Monte Carlo (s) |
|--------------------|-----------|--------------------|
| $2 \times 10^{-3}$ | 50        | 0.4                |
| $2 \times 10^{-4}$ | 2673      | 54                 |

Table 2: Comparison of computation times for RM algorithms ( $\gamma = 1$  and  $c = 2$ ) and Monte Carlo Method in [Armenti et al. \(2018\)](#) ( $m_1^* = m_2^* = 0.6364$ ).

Note that we can easily verify that all conditions in  $(\mathcal{A}_{a.s.})$  and  $(\mathcal{A}_{a.n.})$  hold. We can also verify thanks to the exact formula of  $\Sigma^*$ , that this matrix is positive definite for the different values of  $\rho$  used. This is a condition needed in [Theorem 2.7](#).

For any random estimator, constructing confidence intervals is important to assess the error in the estimation. For PR estimator, confidence interval can be obtained in one simulation run after estimating matrices  $\Sigma^*$  and  $A$  and hence the asymptotic variance matrix  $V$ . [Figure 4](#) shows the convergence, in the case  $\rho = 0$ , of the estimator of  $V_n = A_n^{-1}\Sigma_n(A_n^{-1})^\top$  where  $A_n$  and  $\Sigma_n$  are as introduced in [Proposition 2.9](#).

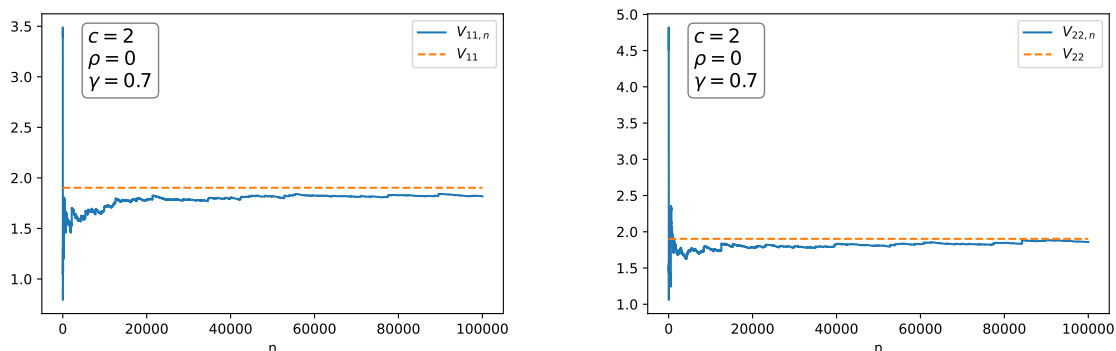


Figure 4: Convergence of the estimator  $V_n$ .

In the following table, we give the estimated confidence interval for PR estimator with a confidence coefficient of 95%:

Table 3: Confidence intervals for PR estimators.

| $\rho$ | CI for $m_1^*$   | CI for $m_2^*$   |
|--------|------------------|------------------|
| -0.5   | [0.3772, 0.4047] | [0.3679, 0.3949] |
| 0      | [0.4962, 0.5259] | [0.4912, 0.5213] |
| 0.5    | [0.6194, 0.6629] | [0.6203, 0.6665] |

As for RM estimators, it is difficult to estimate the asymptotic covariance matrix due to its complexity. In order to visualize the normal behaviour of these estimators, we give the

empirical probability density function (EPDF) in both cases  $\gamma = 1$  and  $\gamma = 0.7$ . To this end, we use again a number of steps  $n = 100000$  and we repeat the procedure  $N = 10000$  times. We restrict our attention to the case  $\rho = 0$ . Figure 5 shows that  $D_{n,i} = \sqrt{n^\gamma}(m_{n,i} - m_i^*)$ ,  $i \in \{1, 2\}$  are very close to a normal distribution.

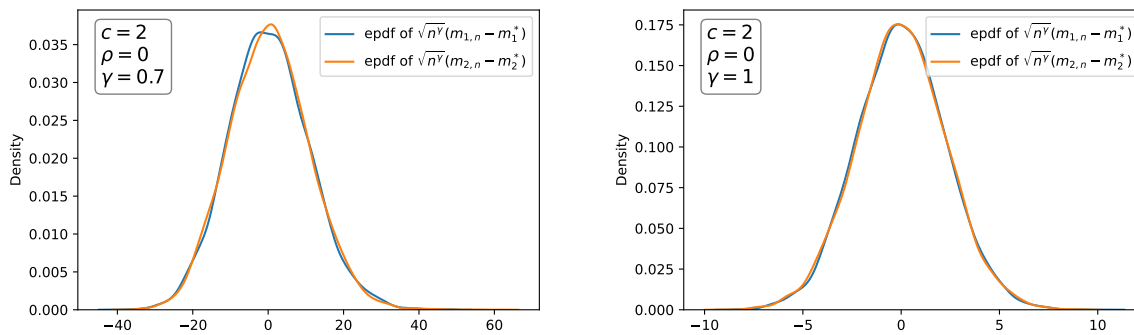


Figure 5: Empirical cumulative density function of  $m_n - m^*$ .

In order to appreciate the quality of convergence of RM estimators, we also give the empirical cumulative density function (ECDF) of the error  $m_n - m^*$ .

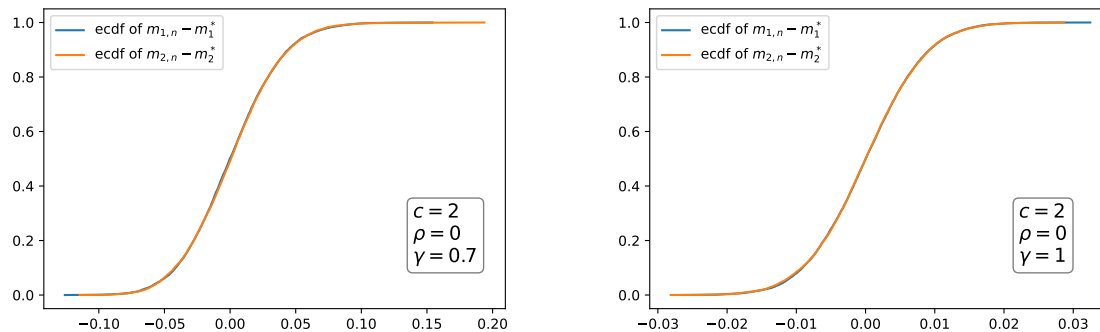


Figure 6: Empirical cumulative density function of  $m_n - m^*$ .

From the two figures above, the width of the 90% confidence interval of the RM estimator for the case  $\gamma = 0.7$  is approximately 8% and for the case  $\gamma = 1$  is roughly 2%.

### 3.2 Second example

As a second example, we will consider consider the following loss function used in [Armenti et al. \(2018\)](#):

$$l(x_1, \dots, x_d) = \sum_{i=1}^d x_i + \frac{1}{2} \sum_{i=1}^d (x_i^+)^2 + \alpha \sum_{i < j} x_i^+ x_j^+.$$

### 3.2.1 First case: Gaussian distribution and $d = 2$

We start by a simple case where we fix  $d = 2$  and use standard two dimensional Gaussian distribution for the loss vector  $X$ . We take  $K = [0, 2]^3$ ,  $n = 10^5$ ,  $t = 10$ ,  $\alpha = 1$  and  $c = 6$ . Again, we compare RM and PR estimators for different values of  $\rho$ . Note that all assumptions can be easily verified except the one that requires the matrix  $cA + I/2$  to be a Hurwitz matrix. However, the PR estimator solve this issue.

The following figure 7 allows us to draw the same conclusions as in the previous example: RM estimator with  $\gamma = 1$  and PR estimator are better than RM estimator with  $\gamma = 0.7$ . RM estimator with  $\gamma = 0.7$  is noisy and one can remediate to this by choosing a smaller value of  $c$  as we did in the first example.

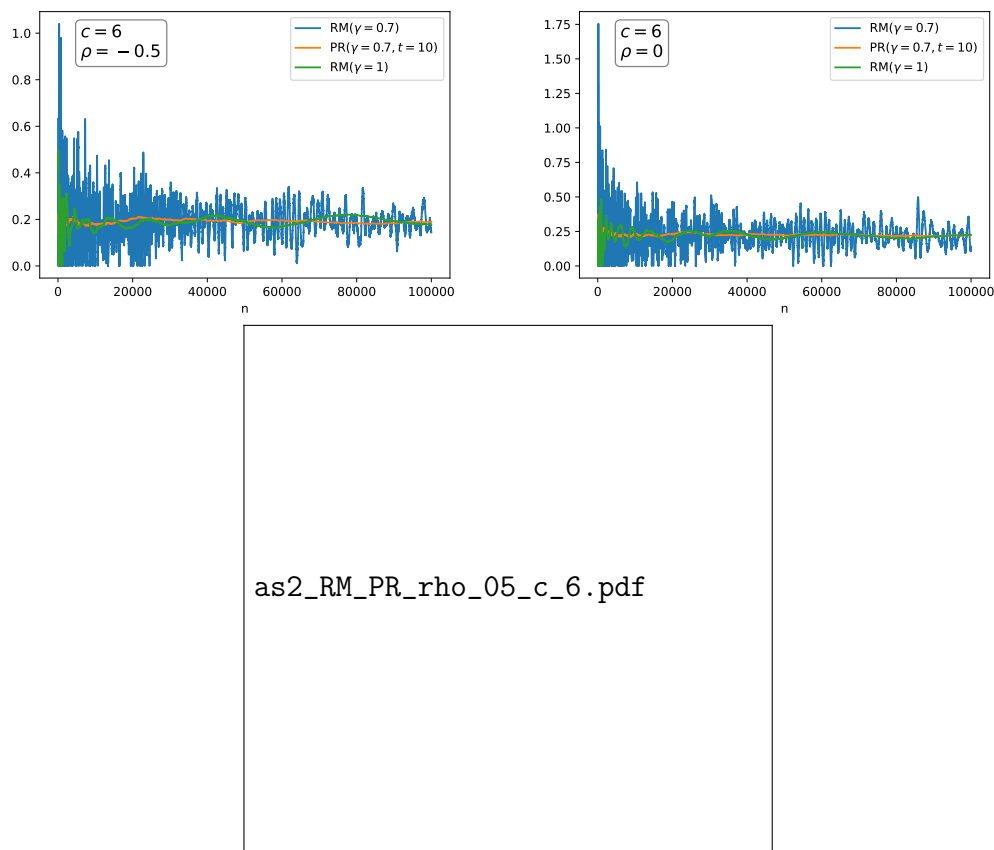


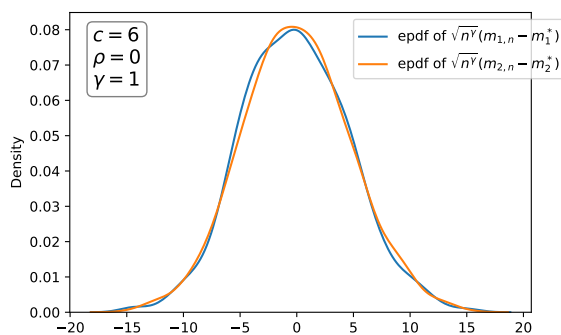
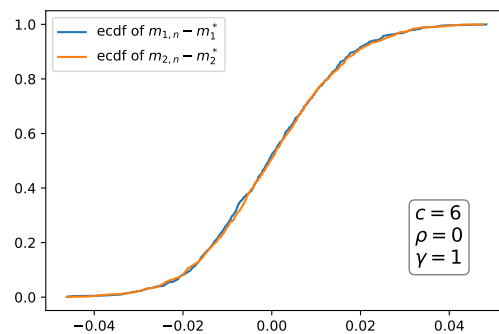
Figure 7: Consistency of RM/PR estimators with for different values of  $\rho$ .

In order to assess the accuracy of our PR estimator, we give the confidence interval with a 95% confidence coefficient (CI), using the estimators of Proposition 2.9. Since  $m_1^* = m_2^*$  in this example, we only present in Table 4 the results for the estimator  $\hat{m}_1^*$  of  $m_1^*$ .

Table 4: Confidence intervals for PR estimators.

| $\rho$ | $\hat{m}_1^*$ | CI               |
|--------|---------------|------------------|
| -0.5   | 0.188         | [0.1790, 0.2089] |
| 0      | 0.21          | [0.1963, 0.2303] |
| 0.5    | 0.25          | [0.2415, 0.2769] |

For RM estimator, we plotted the EPDF of  $D_{n,i}, i \in \{1, 2\}$  as well as the ECDF of the error  $m_n - m^*$  for the case  $\rho = 0$  and  $\gamma = 1$ . These figures shows that the length of the confidence interval of 90% in the case  $\gamma = 0.7$  is much higher that in the case  $\gamma = 1$  (approx 0.2 against 0.04).

Figure 8: Empirical cumulative density function of  $m_n - m^*$ .Figure 9: Empirical cumulative density function of  $m_n - m^*$ .

### 3.2.2 Second case: Compound Poisson Distribution and higher dimensions

In this section, we propose to use compound Poisson processes to model the loss vector  $X$ . The scope of application of compound Poisson processes is very wide. It ranges from

statistical physics and biology to financial mathematics. In biology, they are used to study dynamics of populations. In the modern financial modeling, compound Poisson processes are used to describe dynamics of risk factors such as interest rates (see for instance [Li et al. \(2017\)](#)), foreign exchange rates and option pricing (see [Jaimungal and Wang \(2006\)](#)). In actuarial science, compound processes are extensively used to model claims sizes and to compute the ruin probability, i.e. the probability that the initial reserves increased by premiums received from clients and decreased by their claims, drops below zero.

More precisely, given a final time  $T$ , we consider a multivariate Poisson random vector  $N^T = (N_1^T, \dots, N_d^T)$ , where each  $N_i^T \sim \mathcal{P}(\lambda_i T)$  and the loss corresponding to the  $i^{\text{th}}$  component is  $X_i = \sum_{k=1}^{N_i^T} G_i^k$  and  $(G_i^k)_k$  is an i.i.d sequence representing the jump sizes and independent of  $N_i^T$ . We will take two examples for the distribution of the jumps sizes: One with a Gaussian distribution and another one with an exponential one. The correlation between the different components of  $X$  will be done through the correlations between components of  $N^T$ . In what follows, we detail the method of generating a multivariate Poisson random vector,  $N = (N_1, \dots, N_d)$  with a vector of corresponding intensities  $(\lambda_1, \dots, \lambda_d)$ . To do so, we will use a method that is based on the Gaussian vectors. More precisely, denote  $\eta = (\eta_1, \dots, \eta_d)$  to be a Gaussian random vector having a centered normal distribution with correlation matrix  $R = (\rho_{kl})$  and  $\Phi$  to be the standard normal cdf. Then, the random vector  $\xi = (\Phi(\eta_1), \dots, \Phi(\eta_d))$  has a multivariate distribution with standard uniform marginal distributions. Let  $P_\lambda(x) = \sum_{j=0}^{[x]} (\lambda^j / j!) e^{-\lambda}$  be the cdf of the Poisson distribution with parameter  $\lambda$ . Now, consider the vector  $\zeta = (\zeta_1, \dots, \zeta_d)$  where  $\zeta_k = P_{\lambda_k}^{-1}(\Phi(\eta_k))$ ,  $k = 1, \dots, d$ .  $\zeta$  has therefore Poisson marginal distributions with intensities  $(\lambda_1, \dots, \lambda_d)$ . We can express the correlation coefficient  $\rho_{kl}^* = \text{corr}(\zeta_k, \zeta_l)$  as a function of  $\rho_{kl} = \text{corr}(\eta_k, \eta_l)$ :

$$\begin{aligned} \rho_{kl}^* &= \frac{E(\zeta_k \zeta_l) - E(\zeta_k)E(\zeta_l)}{\sigma(\zeta_k)\sigma(\zeta_l)} \\ &= \frac{E(\zeta_k \zeta_l) - \lambda_k \lambda_l}{\sqrt{\lambda_k \lambda_l}}. \end{aligned}$$

We need to express the expectation  $E(\zeta_k \zeta_l)$  as a function of  $\rho_{kl}$ . We have:

$$\begin{aligned} E(\zeta_k \zeta_l) &= E \left[ P_{\lambda_k}^{-1}(\Phi(\eta_k)) P_{\lambda_l}^{-1}(\Phi(\eta_l)) \right] \\ &= \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} mn P(\zeta_k = m, \zeta_l = n) \\ &= \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} mn P(u_{m-1}^k \leq \Phi(\eta_k) \leq u_m^k, u_{n-1}^l \leq \Phi(\eta_l) \leq u_n^l), \end{aligned}$$

where  $u_j^i = P_{\lambda_i}(j)$ ,  $i, j = 1, \dots, d$ . It remains to explicit the probabilities in the last equality. If we denote  $\Phi_2(\cdot, \cdot, \rho_{kl})$  the bivariate Normal distribution function, we get finally that,

$$\begin{aligned} Z_{mn}(\rho_{kl}) &:= P(u_{m-1}^k \leq \Phi(\eta_k) \leq u_m^k, u_{n-1}^l \leq \Phi(\eta_l) \leq u_n^l) \\ &= \Phi_2(A_m^k, B_n^l, \rho_{kl}) - \Phi_2(A_{m-1}^k, B_n^l, \rho_{kl}) - \Phi_2(A_m^k, B_{n-1}^l, \rho_{kl}) + \Phi_2(A_{m-1}^k, B_{n-1}^l, \rho_{kl}) \end{aligned}$$

where  $A_m^k = \Phi^{-1}(P_{\lambda_k}(m))$  and  $B_n^l = \Phi^{-1}(P_{\lambda_l}(n))$ . As a conclusion, we obtain,

$$\sum_{m=1}^{\infty} \sum_{n=1}^{\infty} mn Z_{mn}(\rho_{kl}) = \lambda_k \lambda_l + \rho_{kl}^* \sqrt{\lambda_k \lambda_l}. \quad (3.6)$$

The equation (3.6) gives an implicit relation between  $\rho_{kl}^*$  and  $\rho_{kl}$ . It also involves two infinite sums which makes it hard to solve. In practice, one needs to truncate this sum and choose some appropriate upper-limits  $M^*$  and  $N^*$ . We are then able to compute the elements of the correlation matrix  $\rho$  of the Gaussian vector given the correlation matrix  $\rho^*$  of the vector  $N$ . However, there is a problem of sufficient conditions for a given positive semi-definite matrix to be a correlation matrix of a multivariate Poisson random vector. This issue is tackled in [Griffiths et al. \(1979\)](#) where it is shown that each  $\rho_{kl}^*$  has to be in a certain range,

$$-1 < \rho_{kl}^{\min} \leq \rho_{kl}^* \leq \rho_{kl}^{\max} \leq 1. \quad (3.7)$$

---

**Algorithm 1:** Algorithm for generating a sample of  $X$  with Compound Poisson Distribution

---

**Input:**  $(\lambda_i)_{i=1, \dots, d}$ ,  $(\sigma_i)_{i=1, \dots, d}$ ,  $T$ , and  $\rho^*$  correlation matrix of  $N_T$ ;

**Ensure:** For each  $k > l$ ,  $\rho_{kl}^*$  verifies the inequality in (3.7);

- 1 Solve the equation (3.6) to find  $\rho$  the correlation of the Gaussian vector  $\eta = (\eta_1, \dots, \eta_d)$ ;
- 2 Generate a sample of Gaussian vector  $\eta$  with correlation matrix  $\rho$  and for  $i = 1, \dots, d$ ;
- 3 **for**  $i = 1, \dots, d$  **do**
- 4     Set  $N_i^T = P_{\lambda_i T}^{-1}(\Phi(\eta_i))$ ;
- 5     Generate a i.i.d sample of  $G_i^k$  of size  $N_i^T$ ;
- 6     Set  $X_i = \sum_{k=1} G_i^k$ .

**Output:**  $X$ ;

---

The following figure shows the covariance matrix for the loss vector  $X$  of dimension  $d = 10$  obtained by generating a random correlation matrix  $(\rho_{kl})$  and using a Gaussian distribution for the jump sizes, i.e.  $G_i^k \sim \mathcal{N}(1, 1)$ . The intensity vector was taken uniformly in  $[1, 3]^{10}$ .

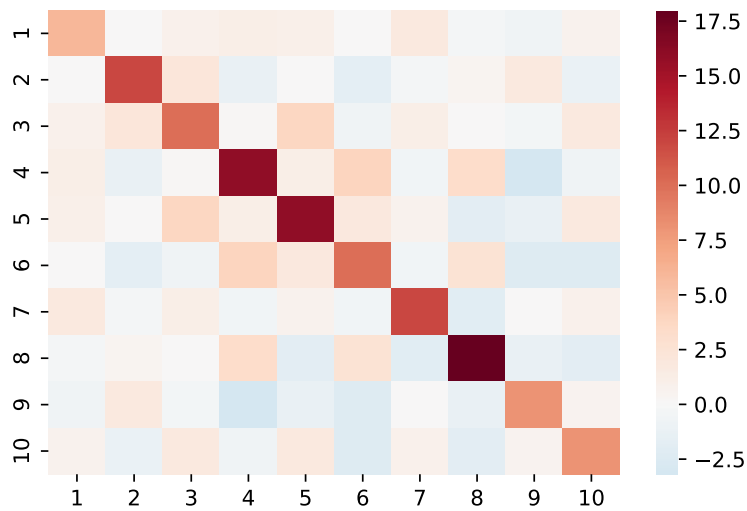


Figure 10: Correlation matrix of the vector loss  $X$  with Gaussian jump sizes with means and variances equal to 1.

Setting  $K = [-20, 20]^{10} \times [0, 20]$ , the averaging parameter  $t = 10$  and  $c = 6$ ,  $\gamma = 0.7$  and the number of steps  $n = 100000$ , we obtain the following optimal allocations for both cases  $\alpha = 0$  and  $\alpha = 1$ .

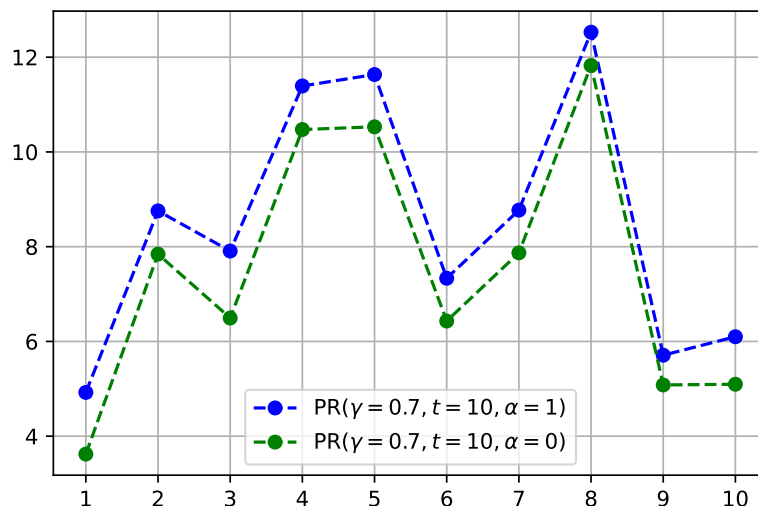


Figure 11: PR estimators for optimal allocations.

The above figure shows that there are components with the same optimal allocations for the case  $\alpha = 0$ . This is something we expect to see, since with  $\alpha = 0$ , correlations between components are not involved, so components with the same variance should have



the same optimal allocations. This is the case for components 4 – 5 and components 9 – 10. However, once  $\alpha$  is taken non null, we see that the same components have no longer the same optimal allocations. For instance, component 10 has higher optimal allocation than 9 when  $\alpha = 1$ . This could be explained by the fact that component 10 is more correlated with other components that have high variances, such as components 4 and 5, than component 9. We now consider an exponential distribution for jump sizes as a second example, i.e.  $G_i^k \sim \mathcal{E}(a_i)$ . The parameters  $a_i$  were generated randomly in  $[0.2, 1.2]$ . As for the other parameters in this example, we took again  $K = [-20, 20]^{10} \times [0, 20]$ ,  $c = 6$ ,  $t = 10$  and  $\gamma = 0.7$ . Covariance matrix of the loss vector  $X$  in this case and estimators of the optimal allocations obtained through PR algorithms with a number of steps  $n = 100000$  together with corresponding confidence intervals are given in the following figures.

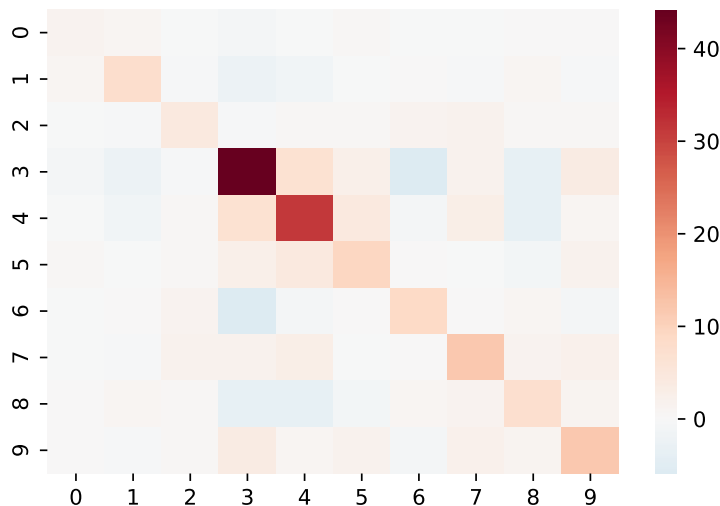


Figure 12: Correlation matrix of the vector loss  $X$  with exponential distribution for the law of jump sizes.

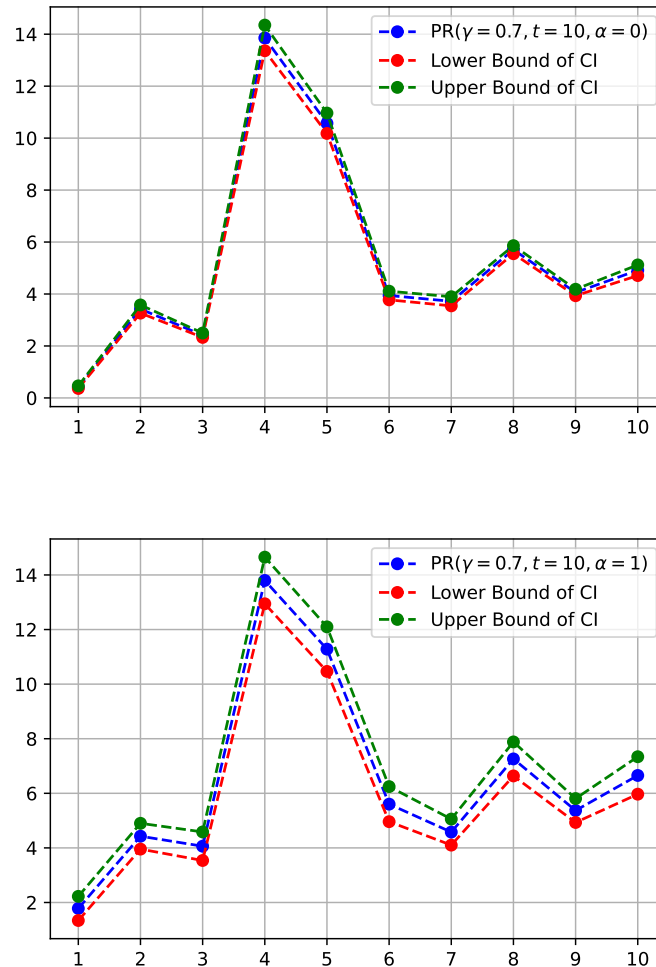


Figure 13: PR estimators for optimal allocations together with bounds of CI.

## 4 Appendix

### 4.1 ODE method and related concepts

In this section, we recall some key concepts of the stability of an ODE  $\dot{z} = h(z)$ .

#### 4.1.1 Concepts of stability of an ODE

A state  $z^*$  is an equilibrium of the ODE if  $h(z^*) = 0$ . To describe the behaviour of the system around the equilibrium, a number of stability concepts are needed. Let us first introduce the basic concepts of stability. To alleviate the notations, we will take 0 as an equilibrium state.

**Definition 4.1.** *The equilibrium  $z^* = 0$  is said to be stable, if for any  $R > 0$ , there exists  $r > 0$  such that if  $\|z(0)\| < r$ , then  $\|z(t)\| < R$  for all  $t \geq 0$ . Otherwise, the equilibrium is unstable.*

Essentially, this means, the system can be kept arbitrarily close to the origin by starting sufficiently close to it. This is also known as *Lyapunov stability*. In some applications, Lyapunov stability is not enough and the concept of asymptotic stability is needed.

**Definition 4.2.** *An equilibrium point  $z^* = 0$  is asymptotically stable if it is stable, and if in addition, there exists some  $r > 0$  such that  $\|z(0)\| < r$  implies that  $z(t) \rightarrow 0$  as  $t \rightarrow \infty$ . The ball  $B_r$  is called a domain of attraction of the equilibrium point.*

*If asymptotic stability holds for all initial states, the equilibrium is said to be globally asymptotically stable.*

#### 4.1.2 Lyapunov function and invariant set

The first property that needs to be verified by a Lyapunov function is *positive definiteness*.

**Definition 4.3.** *A scalar continuous function  $V(z)$  is said to be locally positive definite if  $V(0) = 0$  and in around 0, we have,  $z \neq 0 \Rightarrow V(z) > 0$ .*

*If the above property holds over the whole state space, then  $V(z)$  is said to be globally positive definite.*

The above definition implies that the function  $V$  has a unique minimum at the origin 0. Next, we define the “derivative of  $V$ ” with respect to time along the system trajectory. Assuming that  $V$  is differentiable, this derivative is defined as:

$$\dot{V}(z) = \frac{dV(z)}{dt} = \nabla V \cdot \dot{z} = \nabla V \cdot h(z).$$

**Definition 4.4** (Lyapunov function). *Let  $V$  be a positive definite function and continuously differentiable. If its time derivative along any state trajectory is negative semi-definite, i.e.,*

$$\dot{V}(z) = \nabla V_z \cdot h(z) \leq 0, \quad \forall z,$$

*then  $V$  is said to be a Lyapunov function for the system.*

It is possible to draw conclusions on asymptotic stability, with the help of the invariant set theorems introduced by LaSalle. The central concept in these theorems is that of invariant set, a generalization of the concept of equilibrium point.

**Definition 4.5.** *Let  $z(\cdot)$  be a solution of some ODE. A set  $G$  is said to be an invariant set for this ODE if  $z(0) \in G$  implies that  $z(t) \in G, \forall t \geq 0$ .*

For instance, the singleton  $\{z^*\}$  where  $z^*$  is an equilibrium point is an invariant set. Its domain of attraction is also an invariant set. One other trivial invariant set is the whole state-space,  $\cup_{z_0} \{z(t), t > 0, z(0) = z_0\}$ .

Informally, the LaSalle invariance theorem allows the identification of an ODE’s limit set as a subset of the largest invariant set contained in the set  $\{z; W(z) = 0\}$ , with  $W$  some

well-chosen function (most of the time  $W = \dot{V}$ ). We recall below Barbalat's Lemma (see e.g. Lemma 3 in [Fischer et al. \(2013\)](#)), which is central to proving LaSalle's invariance theorem.

**Lemma 4.6** (Barbalat's Lemma). *Let  $\phi : \mathbb{R}^+ \rightarrow \mathbb{R}$  be a uniformly continuous function. Suppose that  $\lim_{t \rightarrow \infty} \int_0^t \phi(s) ds < \infty$ . Then,  $\lim_{t \rightarrow \infty} \phi(t) = 0$ .*

## 4.2 Strong and weak convergence of constrained stochastic algorithms

Let  $g : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be a continuous function, and let  $z^* \in \mathbb{R}^n$  be an equilibrium point of  $h$ , i.e.  $h(z^*) = 0$ .

The constrained stochastic approximation algorithm is defined by

$$Z_{n+1} = \Pi_K[Z_n + \epsilon Y_n], \quad (4.1)$$

with  $(Y_n)$  a sequence of  $\mathbb{R}^n$  random variables such that  $E[Y_n | \mathcal{F}_n] = g(Z_n)$ ,  $\Pi_K$  the projection operator on the constraint set  $K = \{z; a_i \leq z_i \leq b_i\} \subset \mathbb{R}^n$  and  $(\epsilon_n)$  a sequence decreasing to zero, such that  $\sum_{n=0}^{\infty} \epsilon_n = \infty$ .

**Continuous time interpolation** let  $t_0 = 0$  and for  $n \geq 1$ ,  $t_n = \sum_{i=0}^{n-1} \epsilon_i$ . The piecewise constant continuous-time interpolation of the sequence  $(Z_n)$  is defined by

$$Z^0(t) = Z_n, \quad \forall t \in [t_n, t_{n+1}[.$$

The shifted process is defined by

$$Z^n(t) = Z^0(t_n + t), \quad \forall t \geq 0. \quad (4.2)$$

Similarly, let

$$U^0(t) = \frac{Z_n - z^*}{\sqrt{\epsilon_n}}, \quad \forall t \in [t_n, t_{n+1}[,$$

and

$$U^n(t) = U^0(t_n + t), \quad \forall t \geq 0. \quad (4.3)$$

We recall below the assumption and statement of Theorem 5.2.1 (p.127) and 10.2.1 (p.329) in [Kushner and Yin \(2003\)](#). Note that the framework presented here is restricted to the case where  $E[Y_n | \mathcal{F}_n] = g(Z_n)$  ( $g$  does not depend on  $n$ ). Hence, some assumptions are either trivial and not rewritten here or simplified.

**Strong convergence** The assumption below can be found in Section 5.2 of [Kushner and Yin \(2003\)](#).

**Assumption 4.7.** 1. (A2.1)  $\sup E[|Y_n^2|] < \infty$ .

2. (A2.3)  $g$  is continuous.

3. (A2.4)  $\sum_i \epsilon_i^2 < \infty$ .

**Theorem 4.8** (Theorem 5.2.1 in Kushner and Yin (2003) (partial)). Let  $\{Z^n(\cdot)\}$  be the sequence of continuous-time processes defined by (4.2) and let Assumption (4.7) holds.

There is a set  $N$  of probability zero such that for  $\omega \notin N$ , the family of function  $\{Z^n(\omega, \cdot)\}$  is equicontinuous. Furthermore, if  $z(\cdot)$  is a limit of a convergent subsequence, then the function is solution of the projected ODE

$$\dot{z} = g(z) + C(z), \quad C(z(t)) \in -\mathcal{C}(z(t)),$$

and  $\{Z_n(\omega)\}$  converge to some limit set of the ODE in  $K$ .

**Weak convergence** The assumptions below can be found in Section 10.2 and 10.4 of Kushner and Yin (2003):

**Assumption 4.9.**

1. (A2.0) Either  $\epsilon_n = \frac{1}{n}$  or  $\mu_n = o(\epsilon_n)$ , for  $\sqrt{\frac{\epsilon_n}{\epsilon_{n+1}}} = 1 + \mu_n$ .
2. (A2.1) and (A4.2)  $(Y_n \mathbf{1}_{|Z_n - z^*| \leq \rho})$  is uniformly integrable for some  $\rho > 0$ .
3. (A2.2) Let  $(Z^n(\cdot))$  be the sequence of continuous time processes defined by (4.2). Then  $(Z^n(\cdot))$  converges weakly to  $z^*$ .
4. (A2.4)  $g$  is continuously differentiable.
5. (A2.6 and A4.4) The Jacobian matrix  $A'$  of  $g$  at  $z^*$  is a Hurwitz matrix. If  $\epsilon_n = \frac{1}{n}$ , then  $A' + \frac{I}{2}$  is also a Hurwitz matrix.
6. (A2.7) For some  $p > 0$  and  $\rho > 0$ ,  $\sup_{|z - z^*| \leq \rho} m^{2+p}(z) < \infty$ . Furthermore,

$$E[(Y_n - g(Z_n))(Y_n - g(Z_n))^T \mathbf{1}_{|Z_n - z^*| \leq \rho}] \rightarrow \Sigma_1$$

in probability, with  $\Sigma_1$  some non-negative definite matrix.

7. (A4.5)

- Case  $\epsilon_n = \frac{1}{n}$ : let  $P'$  be the unique symmetric and positive definite solution of the Lyapunov equation  $(A')^T P' + P' A' = -I$ , and  $\lambda$  the largest positive number such that  $I \geq \lambda P'$ . Then  $\lambda > 1$ .
- Case  $\mu_n = o(\epsilon_n)$ : Let  $m(t)$  be the smallest integer  $n$  such that  $t_n \leq t < t_{n+1}$ , with  $m(t) = 0$  for  $t \leq 0$ . Then for all  $T > 0$ ,  $\liminf_n \min_{m(t_n - T) \leq i \leq n} \frac{\epsilon_n}{\epsilon_i} = 1$ .

**Theorem 4.10** (Theorem 10.2.1 in Kushner and Yin (2003) (partial)). Assume that the assumptions 4.7 and 4.9 hold, and let  $U^n(\cdot)$  be the process defined by (4.3).

If  $\mu_n = o(\epsilon_n)$ , then the sequence of processes  $(U^n(\cdot))$  converges weakly in the Skorohod space  $D^n(\mathbb{R}^+)$  to the stationary process  $U$  defined for all  $t \geq 0$  by:

$$U(t) = \int_{-\infty}^t e^{A'(t-s)} dW_s. \quad (4.4)$$

If  $\epsilon_n = \frac{1}{n}$ , then the sequence of processes  $(U^n(\cdot))$  converges weakly to the stationary process  $U$  defined for all  $t \geq 0$  by:

$$U(t) = \int_{-\infty}^t e^{(A' + \frac{I}{2})(t-s)} dW_s, \quad (4.5)$$

with in both cases  $W : \mathbb{R} \rightarrow \mathbb{R}^n$  a Wiener process of covariance matrix  $\Sigma_1$ .

*Proof.* By Assumption 4.7 and 4.9, the assumption (A2.0)-(A2.7) p.329 of Theorem 10.2.1 are straightforward, at the exception of assumption (A2.3). The latter condition is also satisfied as a consequence of Theorem 10.4.1 p.341 in Kushner and Yin (2003) since the assumptions (A4.1)-(A4.5) of this Theorem are also satisfied by Assumption 4.7 (1), Theorem 4.8, and Assumption 4.9 (5) and (7).  $\square$

## References

- Andrieu, C., Moulines, É., and Priouret, P. (2005). Stability of stochastic approximation under verifiable conditions. SIAM Journal on control and optimization, 44(1):283–312.
- Armenti, Y., Crépey, S., Drapeau, S., and Papapantoleon, A. (2018). Multivariate shortfall risk allocation and systemic risk. SIAM J. FINANCIAL MATH., 9(1):90–126.
- Arouna, B. (2004). Adaptative monte carlo method, a variance reduction technique. Monte Carlo Methods and Applications, 10(1):1–24.
- Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. Mathematical Finance, 9(3):203–228.
- Bardou, O., Frikha, N., and Pagès, G. (2009). Computing VaR and CVaR using stochastic approximation and adaptive unconstrained importance sampling. Monte Carlo Methods and Applications, 15(3):173–210.
- Benveniste, A., Metivier, M., and Priouret, P. (1990). Adaptive Algorithms and Stochastic Approximations. Springer-Verlag, Berlin, 1 edition.
- Benzi, M., Golub, G. H., and Liesen, J. (2005). Numerical solution of saddle point problems. Acta Numerica, 14:1–137.
- Biagini, F., Fouque, J., Frittelli, M., and Meyer-Brandis, T. (2019). A unified approach to systemic risk measures via acceptance sets. Mathematical Finance, 29(1):329–367.

- 
- Chen, C., Iyengar, G., and Moallemi, C. C. (2013). An axiomatic approach to systemic risk. Management Science, 59(6):1373–1388.
- Chen, H. and Hu, Y. (1986). Stochastic approximation procedures with randomly varying truncations. Science in China, Ser. A.
- Chen, H.-F., Guo, L., and Gao, A.-J. (1987). Convergence and robustness of the robbins-monro algorithm truncated at randomly varying bounds. Stochastic Processes and their Applications, 27:217–231.
- Duflo, M. (1996). Algorithmes stochastiques. Springer Berlin Heidelberg.
- Dunkel, J. and Weber, S. (2010). Stochastic root finding and efficient estimation of convex risk measures. Operations Research, 58(5):1505–1521.
- Filippov, A. (1988). Differential Equations with Discontinuous Righthand Sides. Springer Dordrecht.
- Fischer, N., Kamalapurkar, R., and Dixon, W. E. (2013). Lasalle-yoshizawa corollaries for nonsmooth systems. IEEE Transactions on Automatic Control, 58(9):2333–2338.
- Föllmer, H. and Schied, A. (2002). Convex measures of risk and trading constraints. Finance and stochastics, 6(4):429–447.
- Frikha, N. (2016). Multi-level stochastic approximation algorithms. The Annals of Applied Probability, 26(2):933–985.
- Föllmer, H. and Schied, A. (2002). Convex measures of risk and trading constraints. Finance and Stochastics, 6:429–447.
- Griffiths, R., Milne, R., and Wood, R. (1979). Aspects of correlation in bivariate poisson distributions and processes. Australian Journal of Statistics, 21(3):238–255.
- Hsieh, M.-h. and Glynn, P. W. (2002). Confidence regions for stochastic approximation algorithms. In Proceedings of the Winter Simulation Conference, volume 1, pages 370–376. IEEE.
- Jacod, J. and Shiryaev, A. (2013). Limit theorems for stochastic processes, volume 288. Springer Science & Business Media.
- Jaimungal, S. and Wang, T. (2006). Catastrophe options with stochastic interest rates and compound poisson losses. Insurance: Mathematics and Economics, 38(3):469–483.
- Kawai, R. (2009). Optimal importance sampling parameter search for lévy processes via stochastic approximation. SIAM Journal on Numerical Analysis, 47:293–307.
-

- 
- Kushner, H. J. and Sanvicente, E. (1975). Stochastic approximation for constrained systems with observation noise on the system and constraint. Automatica, 11:375–380.
- Kushner, H. J. and Yang, J. (1995). Stochastic approximation with averaging and feedback: Faster convergence. In Aström, K., Goodwin, G., and Kumar, P., editors, Adaptive Control, Filtering, and Signal Processing, pages 205–228. Springer, New York.
- Kushner, H. J. and Yin, G. (2003). Stochastic approximation and recursive algorithms and applications. Springer, New York, NY, 2 edition.
- Lelong, J. (2013). Asymptotic normality of randomly truncated stochastic algorithms. ESAIM. Probability and Statistics, 17:105–119.
- Lemaire, V. and Pagès, G. (2010). Unconstrained recursive importance sampling. The Annals of Applied Probability, 20(3):1029 – 1067.
- Li, S., Yin, C., Zhao, X., and Dai, H. (2017). Stochastic interest model based on compound poisson process and applications in actuarial science. Mathematical Problems in Engineering, 2017.
- Ljung, L. (1977). Analysis of recursive stochastic algorithms. IEEE Transactions on Automatic Control, 22(4):551–575.
- Nagurney, A. and Zhang, D. (1995). Projected dynamical systems and variational inequalities with applications, volume 2. Springer Science & Business Media.
- Polyak, B. T. and Juditsky, A. B. (1992). Acceleration of stochastic approximation by averaging. SIAM Journal on Control and Optimization, 30(4):838–855.
- Rao, M. M. and Ren, Z. D. (1991). Theory of Orlicz Spaces. Marcel Dekker Inc. N.Y.
- Rockafellar, R. T. (1970). Convex Analysis. Princeton University Press, Princeton, NJ, princeton math. ser.28 edition.
- Rockafellar, R. T. and Uryasev, S. (2000). Optimization of conditional value-at-risk. Journal of Risk, 2(3):21–41.
- Ruppert, D. (1991). Stochastic approximation. In Ghosh, B. K. and Sen, P., editors, Handbook of Sequential Analysis, pages 503–529. Dekker, New York.
- Shapiro, A., Dentcheva, D., and Ruszczyński, A. (2009). Lectures on Stochastic Programming: Modeling and Theory. SIAM.
- Shevitz, D. and Paden, B. (1994). Lyapunov stability theory of nonsmooth systems. IEEE Transactions on automatic control, 39(9):1910–1914.
-



Slotine, J.-J. E. and Li, W. (1991). Applied Nonlinear Control, volume 199. Prentice hall  
Englewood Cliffs, NJ.