



Exascale Quantification of Uncertainties for
Technology and Science Simulation

D6.2 Report on the calculation of stochastic sensitivities

Document information table

Contract number:	800898
Project acronym:	ExaQUte
Project Coordinator:	CIMNE
Document Responsible Partner:	EPFL
Deliverable Type:	Report
Dissemination Level:	PUBLIC
Related WP & Task:	WP6 task 6.2
Status:	Approved



Authoring

Prepared by EPFL				
Authors	Partner	Modified	Version	Comments
Sundar Ganesh	EPFL	All	1.0	Redaction
Quentin Ayoul-Guilnard				Expertise, re- view and ap- proval
Fabio Nobile				

Change Log

Versions	Modified Page/Sections	Comments
1.0	All	Submitted version

Approval

Approved by CIMNE and EPFL				
	Name	Partner	Date	OK
Task leader	Fabio Nobile	EPFL	2019-05-30	✓
Coordinator	Riccardo Rossi	CIMNE	2019-05-30	✓

Executive summary

This deliverable outlines the methodology behind the calculation of stochastic sensitivities. To this end, the work first formulates the general optimisation problem of interest to the ExaQUte project. It then presents the steps required for calculating the sensitivity of the objective function with respect to design variables. The objective function is assumed to depend on some statistics (risk measure) of an output quantity of interest and includes a design penalisation term. The outline of this document is as follows:

- formulation of the optimisation problem;
- derivation of expressions for the sensitivities for a general risk measure;
- discussion on possible methodologies to compute such sensitivities using available deterministic solvers;
- examples to illustrate the above for specific risk measures and output quantities.

Contents

Acronyms	5
1 Introduction	6
2 Formulation	6
2.1 Optimisation problem	7
2.2 Examples of interest	8
3 Sensitivity calculation	9
3.1 First order optimality conditions	9
3.2 Reduced formulation	11
3.3 Algorithmic considerations	12
3.4 Examples of risk measures and their sensitivities	13
3.4.1 Variance	13
3.4.2 Conditional value at risk	15
4 Summary and future directions	17

Acronyms

OOU optimisation under uncertainties

PDE partial differential equation

QoI quantity of interest

CVaR conditional value at risk

MLMC multi-level Monte Carlo

1 Introduction

One of the goals of the ExaQUte project is to design scalable algorithms for optimisation under uncertainties (OUU) towards application in shape design. In the target area of wind engineering, the stochastic nature of wind needs to be factored into the shape optimisation process. Structures designed for mean conditions are often not robust to perturbations. One example is the case of a transonic airfoil. The position of the shock on a mean-optimal airfoil may not be robust to fluctuations in inflow conditions, resulting in deteriorating airfoil performance for small perturbations from the mean conditions.

One way to account for this uncertainty is to optimise the design with respect to a risk measure associated to the (random) output quantity of interest. Optimising a design with respect to some statistics of a random output quantity is an active research area. A list of challenges exist that need to be solved: (I) a proper formulation of the optimisation problem, (II) an analysis of its properties and (III) the development of scalable and efficient algorithms that can be used to numerically solve the optimisation problem. Gradient-based optimisation techniques were shown in [13] to be effective in solving optimisation problems of this type.

In the following sections, the formulation of a possible stochastic optimisation problem relevant to the ExaQUte project is first presented. This formulation is based on the approach used in [8]. Expressions for the sensitivity of the objective function with respect to the design variables are then derived. It is shown how these sensitivities can be computed using existing deterministic adjoint solvers in combination with statistical estimation techniques, specifically using Monte Carlo sampling and the multi-level Monte Carlo (MLMC) algorithm. The challenges and issues faced when using MLMC-type algorithms for the estimation of different statistics, along with potential solutions, are described. Lastly, mathematical expressions for the sensitivities of some statistics of interest in optimal design are derived; namely, the variance and the conditional value at risk (CVAR) [9].

2 Formulation

In this work we denote by X^* the topological dual of a Banach space X and by $n' \in [1, +\infty[$ the conjugate exponent of $n \in [1, +\infty[$, such that $\frac{1}{n} + \frac{1}{n'} = 1$.

2.1 Optimisation problem

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let Z denote the space of design variables and $\mathcal{A} \subset Z$ the subset of feasible designs. For any $z \in \mathcal{A}$, we assume that the state u of the system to be optimised satisfies a model based on partial differential equation (PDE), possibly including some random effects here denoted by $\omega \in \Omega$. We write the mathematical model in abstract form as

$$F(u(\omega), z, \omega) = 0, \text{ for } \mathbb{P}\text{-a.e. } \omega \in \Omega \quad (1)$$

and assume that the (random) solution $u(\omega)$ belongs to a suitable Banach space U for (almost) every realisation $\omega \in \Omega$ and has up to $p \in [1, +\infty[$ finite moments, i.e. it belongs to the Bochner space $L^p(\Omega, \mathcal{F}, \mathbb{P}; U)$, henceforth noted $L^p(\Omega, U)$ for convenience. The goal is to optimise the performance of the system with respect to some statistics of one or more output quantity of interest (QoI), here defined as a function $u \mapsto Q(u)$ that maps the random solution to a random vector in \mathbb{R}^m , namely $Q : L^p(\Omega, U) \rightarrow L^q(\Omega, \mathbb{R}^m)$ for some $q \in [1, +\infty[$. To facilitate the analysis, we assume that the QoI Q has the form

$$Q(u)(\omega) = \tilde{Q}(\omega, u(\omega)), \text{ for } u \in L^p(\Omega, U), \mathbb{P}\text{-a.e. } \omega \in \Omega, \quad (2)$$

for some function $\tilde{Q} : \Omega \times U \rightarrow \mathbb{R}^m$, that is, the output Q can be evaluated ‘pointwise’ for each realisation $\omega \in \Omega$ and corresponding state $u(\omega)$.

The statistics of the output QoI with respect to which the optimisation is performed is called hereafter the *risk measure* and is defined as a map $L^q(\Omega, \mathbb{R}^m) \ni Q \mapsto \mathcal{R}(Q) \in \mathbb{R}$. It could be, for instance, a centred or uncentred moment (mean, variance, etc.), a quantile or a CVaR of Q .

The final optimisation problem that will be studied in this work has therefore the following form:

$$\begin{cases} \min_{\substack{z \in \mathcal{A} \\ u \in L^p(\Omega, U)}} \{J(u, z) := \mathcal{R}(Q(u)) + P(z)\} \\ \text{s.t. } F(u(\omega), z, \omega) = 0, \text{ for } \mathbb{P}\text{-a.e. } \omega \in \Omega \end{cases} \quad (3)$$

where $P : Z \rightarrow \mathbb{R}$ is a penalisation term on the design variables and the PDE operator $F : U \times Z \times \Omega \rightarrow Y^*$ takes values in the dual of some Banach space Y , i.e. the residual of the PDE is in Y^* . We also assume that for any $z \in Z$ and $y \in L^p(\Omega, U)$, the PDE residual $F(y(\cdot), z, \cdot)$ has $n' \in [1, +\infty[$ finite moments, i.e. $F(y(\cdot), z, \cdot) \in L^{n'}(\Omega, Y^*)$.

2.2 Examples of interest

In this section, an example of each of the above quantities is used to demonstrate the formulation with the help of a problem of interest in wind engineering. Consider the problem of fluid flow around a structure with uncertainties in the inlet conditions. The design variables are the design parameters defining the shape of the structure and uncertainties enter the problem in the form of stochastic wind conditions.

The governing PDE in (4) is typically a fluid flow model such as the Navier–Stokes equations. In the target application, the design variables affect the solution by modifying the shape of the domain boundary. Uncertainties in the wind are modelled as stochastic inlet conditions. The system is then written as follows

$$\begin{cases} F(u(\cdot), z, \cdot) = 0 & \text{in } D_z, \\ g(u) = 0 & \text{on } \Gamma_z, \\ u = u_\Omega & \text{on } \Gamma_\Omega. \end{cases} \quad (4)$$

The domain boundary is partitioned as $\partial D_z = \Gamma_z \cup \Gamma_\Omega$ where Γ_z is the shape-dependent part of the boundary on which the boundary condition defined by g is prescribed and Γ_Ω is the portion on which the uncertain value u_Ω is applied.

One possibility for the QoI Q are force coefficients of the form

$$Q(u) = \int_{\partial\Gamma_z} f(u) \cdot \vec{n} \, dA$$

where \vec{n} is the direction along which the force component is to be calculated and f is the force per unit area in that direction.

It is often the case in structural engineering that a preferred design that is based on constructions costs and architectural considerations is supplied as a starting point for the optimisation algorithm, and that the optimal design not significantly differs from this design. One way to penalise this difference is through a penalisation term such as $P(z) = \|z - z_0\|^2$, where z_0 are the preferred design parameters and $\|\cdot\|$ a suitable norm.

The risk measure \mathcal{R} needs to be chosen such that the optimal design is robust to uncertainties. One such example of this measure is the mean-variance risk measure given by

$$\mathcal{R}(Q) = \mathbb{E}[Q] + \alpha \text{Var}(Q),$$

where $\alpha \in \mathbb{R}$ controls the relative importance of the two terms. Another

example is the CVaR, given by the following expression.

$$\mathcal{R}(Q, \beta) = \inf_{t \in \mathbb{R}} f_\beta(t, Q)$$

where $f_\beta(t, Q) := t + \frac{1}{1 - \beta} \mathbb{E}[(Q - t)^+]$ and $X^+ := \max(X, 0)$.

The CVaR has been shown to have favourable mathematical properties for use in optimisation algorithms [9]. Note that the CVaR is a parametric expectation, dependent on an extra parameter β .

3 Sensitivity calculation

In this section, we outline the calculation of sensitivities of the objective function J defined in (3) with respect to the design parameters z . We present a proof of the relation between these sensitivities and gradients available from deterministic adjoint solvers. We then demonstrate the calculations with some examples. We introduce the following notation that is used extensively in the remainder of the document. For a function of several variables $f : X_1 \times X_2 \times \dots \times X_K \rightarrow Y$, its partial differential with respect to the k -th variable is denoted $D_k f$ and is a linear operator between the spaces X_k and Y , that is $D_k f \in \mathfrak{L}(X_k, Y)$, where we denote by $\mathfrak{L}(X, Y)$ the space of linear operators from X to Y .

3.1 First order optimality conditions

We begin by setting up the Lagrangian functional $\mathcal{L} : L^p(\Omega, U) \times Z \times L^n(\Omega, Y) \rightarrow \mathbb{R}$ of the problem as follows.

$$\mathcal{L}(u, z, \lambda) := \mathcal{R}(Q(u)) + P(z) - \mathbb{E}[\langle F(u(\cdot), z, \cdot), \lambda(\cdot) \rangle_{Y^*, Y}]$$

Here, $\langle \cdot, \cdot \rangle_{Y^*, Y}$ is the duality pairing corresponding to the pair (Y^*, Y) . In the following, we assume sufficient regularity of the objects involved.

The derivative in λ of the Lagrangian, in the direction $\delta\lambda \in L^n(\Omega, Y)$ reads:

$$D_3 \mathcal{L}(u, z, \lambda)(\delta\lambda) = - \mathbb{E}[\langle F(u(\cdot), z, \cdot), \delta\lambda(\cdot) \rangle_{Y^*, Y}]. \tag{5}$$

By setting (5) equal to 0 and taking variations of the form $\delta\lambda = v^\lambda w^\lambda$ with $v^\lambda \in Y$ and $w^\lambda \in L^n(\Omega)$, we obtain

$$\mathbb{E}[\langle F(u(\cdot), z, \cdot), v^\lambda \rangle_{Y^*, Y} w^\lambda(\cdot)] = 0, \quad \forall v^\lambda \in Y, w^\lambda \in L^n(\Omega)$$

which simply corresponds to the primal problem being satisfied for almost every $\omega \in \Omega$

$$\begin{aligned} \langle F(u(\omega), z, \omega), v^\lambda \rangle_{Y^*, Y} &= 0, \quad \forall v^\lambda \in Y \text{ and } \mathbb{P}\text{-a.e. } \omega \in \Omega \\ \iff F(u(\omega), z, \omega) &= 0, \text{ in } Y^* \text{ and for } \mathbb{P}\text{-a.e. } \omega \in \Omega. \end{aligned} \quad (6)$$

To compute the derivative of the Lagrangian in u , we require the following. Since $\mathcal{R} : L^q(\Omega, \mathbb{R}^m) \rightarrow \mathbb{R}$, its derivative $D_1 \mathcal{R}(Q) \in \mathfrak{L}(L^q(\Omega, \mathbb{R}^m), \mathbb{R})$. Hence, $D_1 \mathcal{R}(Q)$ can be identified with an element $\nabla_Q \mathcal{R}(Q)$ of the dual space $L^{q'}(\Omega, \mathbb{R}^m)$ [16, theorem 6.10], such that

$$D_1 \mathcal{R}(Q)(\delta Q) = \mathbb{E}[\nabla_Q \mathcal{R}(Q) \cdot \delta Q].$$

In addition, we also have that $D_1 Q(u) \in \mathfrak{L}(L^p(\Omega, U), L^q(\Omega, \mathbb{R}^m))$. As a result, the derivative in u of the Lagrangian, in the direction $\delta u \in L^p(\Omega, U)$, reads

$$\begin{aligned} D_1 \mathcal{L}(u, z, \lambda)(\delta u) &= \mathbb{E}[\nabla_Q \mathcal{R}(Q(u)) \cdot D_1 Q(u)(\delta u)] \\ &\quad - \mathbb{E}[\langle D_1 F(u(\cdot), z, \cdot)(\delta u(\cdot)), \lambda(\cdot) \rangle_{Y^*, Y}]. \end{aligned} \quad (7)$$

From (2) and the limit definition of the derivative, it is possible to show that the following holds:

$$D_1 Q(u)(\delta u)(\omega) = \langle D_2 \tilde{Q}(\omega, u(\omega)), \delta u(\omega) \rangle_{U^*, U} \text{ for } \mathbb{P}\text{-a.e. } \omega \in \Omega$$

where $D_2 \tilde{Q}(\omega, y)$ is the element of $(U^*)^m$ for all $\omega \in \Omega$, since $\tilde{Q}(\omega, \cdot) : U \rightarrow \mathbb{R}^m$. Using the above, we obtain from (7)

$$\begin{aligned} D_1 \mathcal{L}(u, z, \lambda)(\delta u) &= \mathbb{E} \left[\nabla_Q \mathcal{R}(Q(u))(\cdot) \cdot \langle D_2 \tilde{Q}(\cdot, u(\cdot)), \delta u(\cdot) \rangle_{U^*, U} \right] \\ &\quad - \mathbb{E} \left[\langle D_1 F(u(\cdot), z, \cdot)(\delta u(\cdot)), \lambda(\cdot) \rangle_{Y^*, Y} \right] \end{aligned} \quad (8)$$

Taking again variations of the form $\delta u = v^u w^u$ with $v^u \in U$ and $w^u \in L^p(\Omega)$, and setting (8) equal to 0, we obtain

$$\begin{aligned} \mathbb{E} \left[\nabla_Q \mathcal{R}(Q(u))(\cdot) \cdot \langle D_2 \tilde{Q}(\cdot, u(\cdot)), v^u \rangle_{U^*, U} w^u(\cdot) \right] \\ - \mathbb{E} \left[\langle D_1 F^*(u(\cdot), z, \cdot)(\lambda(\cdot)), v^u \rangle_{U^*, U} w^u(\cdot) \right] = 0, \quad \forall v^u \in U, w^u \in L^p(\Omega) \end{aligned}$$

which corresponds to the adjoint equation for λ being satisfied for almost

every $\omega \in \Omega$:

$$\begin{aligned} & \langle D_1 F^*(u(\omega), z, \omega)\lambda(\omega), v^u \rangle_{U^*, U} \\ &= \nabla_Q \mathcal{R}(Q(u))(\omega) \cdot \langle D_2 \tilde{Q}(\omega, u(\omega)), v^u \rangle_{U^*, U}, \quad \forall v^u \in U, \mathbb{P}\text{-a.e. } \omega \in \Omega \\ \\ & \iff D_1 F^*(u(\omega), z, \omega)\lambda(\omega) \\ &= \nabla_Q \mathcal{R}(Q(u))(\omega) \cdot D_2 \tilde{Q}(\omega, u(\omega)), \quad \text{in } U^* \text{ for } \mathbb{P}\text{-a.e. } \omega \in \Omega. \end{aligned} \quad (9)$$

Finally, the derivative in z of the Lagrangian, in the direction $\delta z \in Z$, reads

$$\begin{aligned} D_2 \mathcal{L}(u, z, \lambda)(\delta z) &= D_1 P(z)(\delta z) - \mathbb{E} \left[\langle D_2 F(u(\cdot), z, \cdot)(\delta z), \lambda(\cdot) \rangle_{Y^*, Y} \right] \\ &= \langle D_1 P(z) - \mathbb{E}[D_2 F^*(u(\cdot), z, \cdot)(\lambda(\cdot))], \delta z \rangle_{Z^*, Z} \end{aligned}$$

where again we notice that $D_1 P(z) \in Z^*$ since $P : Z \rightarrow \mathbb{R}$. Since the optimal point can lie on the boundary of the feasible region, the following variational inequality is required to be satisfied by $D_2 \mathcal{L}$ at the optimum \bar{z} .

$$\begin{aligned} & \langle D_2 \mathcal{L}(u, z, \lambda), z - \bar{z} \rangle_{Z^*, Z} \geq 0, \quad \forall z \in \mathcal{A}, \\ \iff & \langle D_1 P(z) - \mathbb{E}[D_2 F^*(u(\cdot), z, \cdot)(\lambda(\cdot))], z - \bar{z} \rangle_{Z^*, Z} \geq 0, \quad \forall z \in \mathcal{A}. \end{aligned} \quad (10)$$

The three conditions given by (6), (9) and (10) represent the first order optimality conditions corresponding to the optimisation problem given by (3).

3.2 Reduced formulation

To implement a gradient-based algorithm for the optimisation problem, several assumptions must be made on the problem in addition to those listed in the previous sections.

We assume that, for every $z \in \mathcal{A}$, there exist $u_z \in L^p(\Omega, U)$ and $\lambda_z \in L^n(\Omega, Y)$ such that u_z uniquely solves (6) and λ_z uniquely solves the corresponding adjoint equation (9). Under this assumption, we define the reduced objective function $\mathcal{J} : Z \rightarrow \mathbb{R}$ as follows.

$$\mathcal{J}(z) := J(u_z, z) = \mathcal{R}(Q(u_z)) + P(z). \quad (11)$$

Under certain conditions [8, p. 14], it is possible to show that the following holds for the derivative of the reduced objective function $D_1 \mathcal{J}(z) \in \mathfrak{L}(Z, \mathbb{R})$.

$$\begin{aligned} D_1 \mathcal{J}(z)(\delta z) &= D_2 \mathcal{L}(u_z, z, \lambda_z)(\delta z) \\ &= \langle D_1 P(z) - \mathbb{E}[D_2 F^*(u_z(\cdot), z, \cdot)(\lambda_z(\cdot))], \delta z \rangle_{Z^*, Z} \end{aligned} \quad (12)$$

Hence, the optimality conditions for the reduced problem are the same as the optimality conditions for the original problem.

Under the additional assumption that Z is a Hilbert space equipped with the inner product $\langle \cdot, \cdot \rangle_Z$, we can also use the Riesz representation theorem to show that for $f : Z \rightarrow \mathbb{R}$ Fréchet-differentiable at $z \in Z$, $D_1 f(z) \in Z^*$ admits a representation $\nabla_z f(z) \in Z$. Hence, introducing $\nabla_z \mathcal{J}(z) \in Z$ such that

$$\langle \nabla_z \mathcal{J}(z), \delta z \rangle_Z = \langle D_1 \mathcal{J}(z), \delta z \rangle_{Z^*, Z}$$

in combination with (12) allows us to write

$$\langle \nabla_z \mathcal{J}(z), \delta z \rangle_Z = \langle D_1 P(z) - \mathbb{E}[D_2 F^*(u_z(\cdot), z, \cdot)(\lambda_z(\cdot))], \delta z \rangle_{Z^*, Z} \quad (13)$$

3.3 Algorithmic considerations

A typical gradient-based minimisation method would compute iteratively a sequence of approximations $(z_k)_{k \in \mathbb{N}}$ where, for any $k \in \mathbb{N}$,

$$z_{k+1} := \Pi_{\mathcal{A}}(z_k - \gamma_k \nabla_z \mathcal{J}(z_k))$$

where $\Pi_{\mathcal{A}} : Z \rightarrow \mathcal{A}$ is the projection operator onto the feasible set \mathcal{A} and $\gamma_k \in \mathbb{R}$ is the ‘step size’. The steps to compute $\nabla_z \mathcal{J}(z_k)$ in this setting are outlined in algorithm 1.

Algorithm 1 Computation of $\nabla_z \mathcal{J}(z_k)$ given z_k

- 1: Solve (6) for $u_{z_k}(\omega)$ for all $\omega \in \Omega$
 - 2: Solve (9) for $\lambda_{z_k}(\omega)$ for all $\omega \in \Omega$
 - 3: Compute the expectation term in (13) using u_{z_k} and λ_{z_k} .
 - 4: Compute $\nabla_z \mathcal{J}(z_k)$ using (13).
-

Step 3 of Algorithm 1 requires in theory the solutions $u_{z_k}(\omega)$ and $\lambda_{z_k}(\omega)$ for all $\omega \in \Omega$. However, since only a finite number of samples can be generated from Ω in practice, the expectation needs to be estimated. We propose the use of Monte Carlo techniques in this work, where only a finite number of samples $(\omega_j)_{j=1}^N \subset \Omega$ are generated independently from the same probability law \mathbb{P} . The expectation in (13) is then estimated as follows.

$$\mathbb{E}[D_2 F^*(u_z(\cdot), z, \cdot)(\lambda_z(\cdot))] \approx \frac{1}{N} \sum_{j=1}^N D_2 F^*(u_z(\omega_j), z, \omega_j)(\lambda_z(\omega_j))$$

To accelerate the convergence of the standard Monte Carlo estimator, the MLMC method [3, 4, 15] uses correlated samples generated on a sequence of

meshes of the primal and adjoint problems. MLMC methods have also been used with success in combination with stochastic optimisation techniques, wherein a ‘fully accurate’ estimation of the gradient is not required at each iteration of the optimisation algorithm [13].

The use of MLMC algorithms to estimate the gradient of the optimisation problem poses a number of challenges. A major challenge here is to have efficient multi-level estimators to obtain estimations in reasonable time that will enable the sequence of designs z_k to convergence as swiftly as possible. Although unbiased multi-level estimators for moments have been successfully demonstrated [14], unbiased multi-level estimators for parametric expectations such as the CVaR and their sensitivities is still an active area of research [10].

It is also possible that, depending on the risk measure used, the object $\nabla_Q \mathcal{R}$ required to solve the adjoint problem can itself contain statistics that need to be estimated (see next section for an example). As a result, the solution of the adjoint equation for one particular point in Ω requires, in theory, the solution of the primal problem for all $\omega \in \Omega$. One possible way to solve this issue is to have a separate MLMC estimation of these expectations, which are then used in the MLMC algorithm that estimates the expectation in (13). Another possible way to solve the issue is to decouple the primal and adjoint problems by parametrising the expectations within $\nabla_Q \mathcal{R}$ and iteratively updating them to convergence.

3.4 Examples of risk measures and their sensitivities

We illustrate here the above calculations for two different risk measures. Specific problems arising for each of the examples are highlighted and possible solutions are proposed. We assume in the following that the random quantity of interest depends *linearly* on the solution of the PDE, i.e. $\exists q \in U^*$ such that $\tilde{Q}(\cdot, x) = \langle q, x \rangle_{U^*, U}$. To lighten the notation, hereafter we omit the subscript in the duality pairing when no ambiguity arises. Since there is no explicit dependence of \tilde{Q} on $\omega \in \Omega$ with this definition we have simply $Q(u)(\omega) = \langle q, u(\omega) \rangle$.

3.4.1 Variance

Let us consider the case $\mathcal{R}(Q) := \text{Var}(Q) = \mathbb{E}[Q^2] - \mathbb{E}[Q]^2$. We assume that $p \geq 2$ for $u \in L^p(\Omega, U)$ so that $q = p$ for $Q(u) \in L^q(\Omega, \mathbb{R})$, since Q is linear in u and the variance of Q is well defined. Solving the adjoint problem (9) for λ requires the computation of the objects $\nabla_Q \mathcal{R}$ and $D_2 \tilde{Q}$.

The differential of \mathcal{R} evaluated at $Q \in L^q(\Omega, \mathbb{R}^m)$ in any direction $\delta Q \in L^q(\Omega, \mathbb{R}^m)$ can be written as

$$\begin{aligned} D_1 \mathcal{R}(Q)(\delta Q) &= \mathbb{E}[2Q\delta Q] - 2\mathbb{E}[Q] \mathbb{E}[\delta Q] \\ &= \mathbb{E}[2(Q - \mathbb{E}[Q]) \cdot \delta Q] \\ \implies \nabla_Q \mathcal{R} &= 2(Q - \mathbb{E}[Q]) \end{aligned}$$

The differential of Q evaluated at $u \in L^p(\Omega, U)$ can be written as $D_1 Q(u) = q$. Combining the above two, we can write down the following for a perturbation of the form $\delta u = v^u w^u$, with $v^u \in U$ and $w^u \in L^p(\Omega)$.

$$\begin{aligned} \mathbb{E}[\nabla_Q \mathcal{R}(Q(u)) \cdot D_1 Q(u)(\delta u)] &= \mathbb{E}[2(Q(u) - \mathbb{E}[Q(u)]) \cdot \langle q, \delta u \rangle] \\ &= \mathbb{E}[2(Q(u) - \mathbb{E}[Q(u)]) \cdot \langle q, v^u \rangle w^u] \\ &= \mathbb{E}[2\langle q, u - \mathbb{E}[u] \rangle \langle q, v^u \rangle w^u] \\ &=: \mathbb{E}[\langle G_u, v^u \rangle w^u] \end{aligned}$$

Note that we have defined $G_u \in L^{p'}(\Omega, U^*)$ where $p' \in [1, 2]$ is the conjugate of p with

$$G_u := 2\langle q, u - \mathbb{E}[u] \rangle q. \tag{14}$$

As a matter of fact, in this case, G_u belongs to the stronger space $L^p(\Omega, U^*)$. With G_u defined in this way, this object can be identified with the right hand side of the adjoint equation.

$$\langle G_u(\cdot), v^u \rangle = \nabla_Q \mathcal{R}(Q(u))(\cdot) \cdot \langle D_2 \tilde{Q}(\cdot, u(\cdot)), v^u \rangle.$$

We can then use expectation-approximation techniques, such as a Monte Carlo method, to estimate the sensitivities of the risk measure. However, the term $\mathbb{E}[u]$ in (14) has to be approximated before $G_u(\omega)$ can be estimated, which in theory requires the primal problem to be solved for \mathbb{P} -a.e. $\omega \in \Omega$ and in practice is likely to incur a sizeable cost.

One can work around this issue by parametrising the estimation of $\mathbb{E}[u]$: rewrite the optimisation problem (11) as

$$\left\{ \begin{array}{l} \min_{\substack{z \in \mathcal{A} \\ u \in L^p(\Omega, U) \\ \eta \in U}} \left\{ J(u, z, \eta) := \mathbb{E}[(Q(u) - \eta)^2] + P(z) \right\} \\ \text{s.t. } F(u(\omega), z, \omega) = 0 \text{ for } \mathbb{P}\text{-a.e. } \omega \in \Omega \\ \text{and } \eta = \mathbb{E}[Q(u)]. \end{array} \right. \tag{15}$$

From (15), we define the Lagrangian as

$$\begin{aligned} \mathcal{L}(u, z, \lambda, \eta, \mu) := & \mathbb{E}[(Q(u) - \eta)^2] + P(z) \\ & - \mathbb{E}[\langle F(u, z, \cdot), \lambda(\cdot) \rangle] - \mu(\eta - \mathbb{E}[Q(u)]) \end{aligned}$$

This adds the following equations to the list of optimality conditions.

$$\begin{aligned} D_5 \mathcal{L} &= -(\eta - \mathbb{E}[Q(u)]) = 0 \\ D_4 \mathcal{L} &= 2 \mathbb{E}[(\eta - Q(u))] - \mu = 0 \end{aligned}$$

Although this frees the solution of the adjoint problem for one $\omega \in \Omega$ from requiring the solution of the primal problem for all ω , a separate procedure for the update of η must be devised at an additional cost. Other approaches are proposed in [18], at least in the case of a linear PDE.

3.4.2 Conditional value at risk

The second example that we use to illustrate the calculations is that of the CVaR. The CVaR is expressed as follows

$$\text{CVaR}_\beta := \inf_{t \in \mathbb{R}} t + \frac{1}{1 - \beta} \mathbb{E}[(Q - t)^+]$$

To avoid the issue of solving both the above minimisation problem and problem (3), we reformulate the optimisation problem as follows.

$$\left\{ \begin{array}{l} \min_{\substack{z \in \mathcal{A}, t \in \mathbb{R} \\ u \in L^p(\Omega, U)}} \{J(u, z, t) := \mathcal{R}(Q(u), t) + P(z)\} \\ \text{s.t. } F(u(\omega), z, \omega) = 0 \text{ for } \mathbb{P}\text{-a.e. } \omega \in \Omega \\ \text{where } \mathcal{R}(Q, t) := t + \frac{1}{1 - \beta} \mathbb{E}[(Q - t)^+] \end{array} \right.$$

Since the objective function J is dependent on t as well, the Lagrangian formulation changes accordingly. The Lagrangian \mathcal{L} now has the form

$$\mathcal{L}(u, z, \lambda, t) = \mathcal{R}(Q(u), t) + P(z) - \mathbb{E}[\langle F(u(\cdot), z, \cdot), \lambda(\cdot) \rangle]$$

This adds one additional condition to the optimality conditions, namely the following.

$$D_4 \mathcal{L}(u, z, \lambda, t) = D_2 \mathcal{R}(Q(u), t) = 0$$

Following similar steps to § 3.4.1, the differential of \mathcal{R} evaluated at $Q \in L^q(\Omega, \mathbb{R})$ in any direction $\delta Q \in L^q(\Omega, \mathbb{R})$ is written as follows. For this case, we have that $q = p$.

$$\begin{aligned} D_1 \mathcal{R}(Q, t)(\delta Q) &= \frac{1}{1 - \beta} \mathbb{E}[\mathbb{1}_{Q \geq t} \delta Q] \\ \implies \nabla_Q \mathcal{R}(Q) &= \frac{\mathbb{1}_{Q \geq t}}{1 - \beta} \end{aligned}$$

The differential of Q evaluated at $u \in L^p(\Omega, U)$ can be written as $D_1 Q(u) = q$. Combining the above two, we can write down the following.

$$\begin{aligned} \mathbb{E}[\nabla_Q \mathcal{R}(Q(u)) \cdot D_1 Q(u)(\delta u)] &= \frac{1}{1 - \beta} \mathbb{E}[\mathbb{1}_{Q \geq t} \langle q, \delta u \rangle] \\ &= \frac{1}{1 - \beta} \mathbb{E}[\mathbb{1}_{Q \geq t} \langle q, v^u \rangle w^u(\cdot)] \\ &=: \mathbb{E}[\langle G_u, v^u \rangle w^u(\cdot)], \end{aligned}$$

where we have made similar assumptions to § 3.4.1 and have defined $G_u \in L^\infty(\Omega, U^*)$ with

$$G_u := \frac{\mathbb{1}_{\langle q, u \rangle \geq t}}{1 - \beta} q$$

We also report here the quantity $D_2 \mathcal{R}$ for use in the redefined optimisation problem.

$$D_2 \mathcal{R}(Q, t) = -\frac{1}{1 - \beta} \mathbb{E}[\mathbb{1}_{Q \geq t}]$$

The computation of the optimal design that minimises a given CVaR and the estimation of the gradient of the CVaR using MLMC algorithms provide several challenges to tackle. Firstly, it is possible that since $\mathbb{1}_{Q \geq t}$ is a discontinuous function, correlated MLMC samples could be created such that the samples lie on either side of the discontinuity. This leads to non-optimal variance reduction. Potential fixes for this problem such as regularisation of the discontinuity [9] as well as adaptive sampling [5] have been proposed in literature.

Additionally, it is possible that the corresponding objective function is such that gradients are discontinuous and not Lipschitzian. This property affects important parameters of gradient-based optimisation algorithms such as the step size. Adaptive step-size-selection algorithms such as line-search methods, as well as sub-gradient methods, tackle this difficulty; we refer the interested reader to [6, 7] for an extensive discussion.

4 Summary and future directions

The optimisation problem of interest to the ExaQUte project was formulated for a general risk measure in § 2. The primal equation (6) and adjoint equation (9) of the underlying PDE (1), as well as expressions for sensitivities of this risk measure with respect to design variables were derived. It was also shown that these expressions involved estimating expectations of sensitivities related to said PDE, for which deterministic primal and adjoint solvers could be used in combination with sampling techniques such as MLMC.

Practical gradient-based approaches to solve the OUU problem (3) were discussed in § 3.3. A number of challenging mathematical and algorithmic issues were discussed, such as nested expectations causing coupling between the primal and adjoint problems and the need for unbiased efficient estimators. Solutions that have shown promise in remedying these issues were also described. These challenges were illustrated using specific examples in § 3.4. In addition, nuances related to each particular example risk measure were enumerated in their corresponding sections and potential solutions described. In summary it was demonstrated that, although sensitivity calculation for the optimisation problem poses multiple challenges, promising solutions exist for each of the issues that will be implemented and studied in the following stages of the project.

Given the high computational cost of these calculations, we aim at curbing them by leveraging modern exascale systems with computational concurrency to exploit the latent parallelism in MLMC methods. To this end, the ExaQUte project is involved in the development of the PyCOMPSs [2, 11, 17] scheduling tool and the ExaQUte XMC library [1]. Further parallelisation is envisioned, e.g. through stochastic gradient methods [see 12], possibly in asynchronous algorithms.

References

- [1] Ramon Amela, Quentin Ayoul-Guilnard, Rosa M Badia, Sundar Ganesh, Fabio Nobile, Riccardo Rossi and Riccardo Tosi. *ExaQUTE XMC*. Comp. software. ExaQUTE consortium, May 2019. DOI: [10.5281/zenodo.3235833](https://doi.org/10.5281/zenodo.3235833).
- [2] Rosa M. Badia, J. Conejero, C. Diaz, J. Ejarque, D. Lezzi, F. Lordan, C. Ramon-Cortes and R. Sirvent. ‘COMP Superscalar, an interoperable programming framework’. In: *SoftwareX* 3–4 (Dec. 2015). DOI: [10.1016/j.softx.2015.10.004](https://doi.org/10.1016/j.softx.2015.10.004).
- [3] Nathan Collier, Abdul-Lateef Haji-Ali, Fabio Nobile, Erik Von Schwerin and Raúl Tempone. ‘A continuation multilevel Monte Carlo algorithm’. In: *BIT Numerical Mathematics* 55.2 (2015), pp. 399–432.
- [4] Michael B Giles. ‘Multilevel monte carlo path simulation’. In: *Operations Research* 56.3 (2008), pp. 607–617.
- [5] Michael B Giles and Abdul-Lateef Haji-Ali. ‘Multilevel nested simulation for efficient risk estimation’. In: *SIAM/ASA Journal on Uncertainty Quantification* 7.2 (2019), pp. 497–525.
- [6] Claude Lemaréchal Jean-Baptiste Hiriart-Urruty. *Convex Analysis and Minimization Algorithms I: Fundamentals*. 1st ed. Grundlehren der mathematischen Wissenschaften 305. Springer-Verlag Berlin Heidelberg, 1993. ISBN: 978-3-642-08161-3,978-3-662-02796-7.
- [7] Claude Lemaréchal Jean-Baptiste Hiriart-Urruty. *Convex Analysis and Minimization Algorithms II: Advanced Theory and Bundle Methods*. 1st ed. Grundlehren der mathematischen Wissenschaften 306. Springer-Verlag Berlin Heidelberg, 1993. ISBN: 978-3-642-08162-0,978-3-662-06409-2.
- [8] Drew P. Kouri and Alexander Shapiro. ‘Optimization of PDEs with uncertain inputs’. In: *Frontiers in PDE-Constrained Optimization*. Ed. by Harbir Antil, Drew P. Kouri, Martin-D. Lacasse and Denis Ridzal. Vol. 163.1. The IMA Volumes in Mathematics and its applications. Springer New York, 2018, pp. 41–82. DOI: [10.1007/978-1-4939-8636-1](https://doi.org/10.1007/978-1-4939-8636-1).
- [9] Drew P Kouri and Thomas M Surowiec. ‘Risk-averse PDE-constrained optimization using the conditional value-at-risk’. In: *SIAM J. Optim.* 26.1 (2016), pp. 365–396. DOI: [10.1137/140954556](https://doi.org/10.1137/140954556).

- [10] Sebastian Krumscheid and Fabio Nobile. ‘Multilevel Monte Carlo approximation of functions’. In: *SIAM/ASA Journal on Uncertainty Quantification* 6.3 (2018), pp. 1256–1293.
- [11] F. Lordan et al. ‘ServiceSs: an interoperable programming framework for the Cloud’. In: *Journal of Grid Computing* 12 (1 Mar. 2014). DOI: [10.1007/s10723-013-9272-5](https://doi.org/10.1007/s10723-013-9272-5).
- [12] Matthieu Claude Martin. ‘Stochastic approximation methods for PDE-constrained optimal control problems with uncertain parameters’. PhD thesis. Lausanne: EPFL, 2019. Chap. 4, pp. 93–130. DOI: [10.5075/epfl-thesis-7233](https://doi.org/10.5075/epfl-thesis-7233).
- [13] Matthieu Claude Martin, Sebastian Krumscheid and Fabio Nobile. *Analysis of stochastic gradient methods for PDE-constrained optimal Control Problems with uncertain parameters*. Tech. rep. 04.2018. MATHICSE, 2018. DOI: [10.5075/epfl-MATHICSE-263568](https://doi.org/10.5075/epfl-MATHICSE-263568).
- [14] Michele Pisaroni, Sebastian Krumscheid and Fabio Nobile. *Quantifying uncertain system outputs via the multilevel Monte Carlo method — Part I: central moment estimation*. Tech. rep. 23.2017. MATHICSE, 2017.
- [15] Michele Pisaroni, Fabio Nobile and Pénélope Leyland. ‘A Continuation Multi Level Monte Carlo (C-MLMC) method for uncertainty quantification in compressible inviscid aerodynamics’. In: *Computer Methods in Applied Mechanics and Engineering* 326 (2017), pp. 20–50.
- [16] Alexander Shapiro, Darinka Dentcheva and Andrzej Ruszczyski. *Lectures on stochastic programming: modeling and theory*. SIAM, 2009.
- [17] Enric Tejedor, Yolanda Becerra, Guillem Alomar, Anna Queralt, Rosa M. Badia, Jordi Torres, Toni Cortes and Jesús Labarta. ‘PyCOMPSs: Parallel computational workflows in Python’. In: *International Journal of High Performance Computing Applications* 31 (1 2017). DOI: [10.1177/1094342015594678](https://doi.org/10.1177/1094342015594678).
- [18] A. Van Barel and S. Vandewalle. ‘Robust Optimization of PDEs with Random Coefficients Using a Multilevel Monte Carlo Method’. In: *SIAM/ASA Journal on Uncertainty Quantification* 7.1 (2019), pp. 174–202. DOI: [10.1137/17M1155892](https://doi.org/10.1137/17M1155892).