Optimal Taxation with Multiple Incomes and Types Work in Progress - Comments welcom*

Etienne LEHMANN† Sander RENES‡ Kevin SPIRITUS§
and Floris T. ZOUTMAN¶

January 15, 2020

Abstract

We derive an optimal nonlinear income tax formula in the case where taxpayers have several incomes and can differ along multiple unobservable dimensions. We show that the tax perturbation approach of Golosov et al. (2014) and the mechanism design approach of Mirrlees (1976) lead to the same optimal tax formula. We also decompose the design of the optimal tax system in two steps: which taxpayers are assigned to the same tax liability (the design of isotax curves) and which tax liability is assigned to the each isotax curve. The solution to the second step is characterized by an ABC formula as Diamond (1998) and Saez (2001), in which welfare weights and behavioral elasticities are averaged among all taxpayers located on the same isotax curves. Applying our model to the optimal household tax problem, our numerical results display isotax curves that are almost linear and parallel, except close to the boundaries of the income domain.

I Introduction

We consider an optimal, nonlinear tax framework where taxes depend on many incomes and unobservable heterogeneity is also assumed to be multidimensional. This framework allows one to study optimal taxation of couples, or the optimal combination of taxes on labor and capital incomes without imposing ad-hoc restrictions on the form of the tax schedule.

Two approaches have been used in the literature to derive optimal tax formula. On the one hand, Mirrlees (1976, 1986) used the First-Order Mechanism Design approach. In this approach the social planner finds the best allocation, expressed as a mapping from types to choices or incomes, achievable under a resource constraint and the the first-order incentive constraints of the tax payers. On the other hand, Golosov et al. (2014) apply a tax perturbation approach to derive another optimal tax formula. The perturbation allows the authors to derive

*We thank participants to various conferences and seminars, incluing the Paris Tax workshop and IIPF 2019, with a special mention to Jonas Loebbing.

†CRED, Université Panthéon-Assas Paris II, 12 Place du Panthéon, 75 234 Paris Cedex 05, France. Email: elehmann@u-paris2.fr. Etienne LEHMANN is also member of CEPR, CESifo, IZA, IDEP and TEPP.

‡Erasmus Universiteit Rotterdam, Department of Business Economics, P.O. box 1738, 3000 DR Rotterdam, The Netherlands. Email srenes@ese.eur.nl

§Erasmus Universiteit Rotterdam, Burgemeester Oudlaan 50, 3062 PA Rotterdam, The Netherlands. Email: spiritus@ese.eur.nl

¶Norwegian School of Economics and Business Administration (NHH), Helleveien 30, 5045 Bergen, Norway. Email: fzoutman@gmail.com

an optimal tax equation expressed in terms of behavioral elasticities, welfare weights and the joint distribution of incomes. Our first contribution is to connect these two approaches by showing that they lead to the same optimal tax formulas expressed in terms of behavioral elasticities, welfare weights and the joint distribution of types. While Saez (2001) verified the consistency of these two approaches in the case with one income and one type, we bridge these two approaches in the multidimensional case. We also provide a sufficient condition that ensures the solution to the optimal tax problem is unique and maximizes social welfare.

We decompose the derivation of the multi-dimensional optimal tax system in two steps. The first step characterizes the taxpayers that are assigned to the same tax liability. In the income space, these taxpayers are located on the same "isotax curves". The second step assigns a tax liability to each isotax curve. We show that this second step is characterized by applying an ABC formula reminiscent to Diamond (1998) and Saez (2001), where welfare weights and behavioral elasticities are averaged among all taxpayers located in the same isotax curve. Therefore, the specific difficulty of the multidimensional tax problem appears to lie in the first step.

We show an application of our analytical results to the taxation of couples. In our simulations, where the couples are unitary actors and have quasilinear, additively separable and isoelastic preferences, our numerical results show that isotax curves are almost linear and parallel. The main departure from linearity appears close to the boundaries where the curvature of isotax curves are driven by the boundary conditions.

Mirrlees (1976) (see also Mirrlees (1986)) was the first to derive optimal necessary conditions for the optimal tax problems in the multidimensional case. His formula took the form of a second-order nonlinear Partial Differential Equation (PDE) expressed in terms of the type distribution and the derivatives of the individual utility functions and the social welfare function. Mirrlees was neither able to derive intuitive implications from his for formula, nor able to solve the PDE associated to his formula.

The second contribution in this line of research appears forty years after Mirrlees (1976) when Kleven et al. (2006, 2007) consider the optimal household taxation problem.[1] As we do in our numerical application, they assumed quasilinear and additively separable preferences. Moreover, they restrict the type distribution by assuming both types are independently distributed. Under these very specific preferences, they pioneered the tax perturbation approach in the multidimensional context, and also connected this approach to the Mirrleesian mechanism design approach. They also provide a condition for the optimal tax to be characterized by *negative jointness*, i.e. optimal marginal tax on male income decrease in female income and vice-versa. We are able to numerically solve the optimal tax problem with the same preference

[1]The working papers Kleven et al. (2006, 2007) formed a second part to Kleven et al. (2009). Surprisingly, only the first part, which assume the husbands take intensive labor supply decisions and their wives only decide whether or not to work, was published in Kleven et al. (2009). The second part, which assume both spouses take intensive labor supply decisions, remains unpublished.

assumptions, but we do not need to impose independence between male and female productivity. This allows us to consider the realistic case of positive assortative matching.

Lastly, Golosov et al. (2014) derive an optimal tax formula in the multidimensional case by using a tax perturbation approach without any restriction on utility function or type distribution. They were able to provide some intuitions for the economics behind their optimal formula, however, they did not verify their formula was consistent with Mirrlees (1976, 1986). Finally, they did not solve, neither analytically nor numerically, the PDE associated to their formula. Our decomposition of the optimal formula allows us to take steps towards solving the PDE and we provide a numeric algorithm to solve the resulting system of equations.

The paper is organized as follow. The problem is described in Section II. Section III is devoted to the tax perturbation approach while Section IV is devoted the mechanism design approach. Finally, numerical results are described in Section V.

II The model

II.1 Households

The economy consists of a unit mass of taxpayers who differ in a p-dimensional vector of characteristics denoted $\mathbf{w} \stackrel{\text{def}}{\equiv} (w_1, ..., w_p)$. We refer to the complete vector of characteristics of an individual as their type. Types are drawn from the type space, which is denoted $\mathcal{W} \subset \mathbb{R}_+^p$ and is assumed to be closed and convex. Types are distributed over $\mathcal{W}$ according to a twice continuously differentiable density denoted $f(\cdot)$.

Taxpayers have $n \geq 2$ different observable income, actions or choices. As we will be interested in their tax consequences, we will denote their corresponding cash-flows $\mathbf{x} \stackrel{\text{def}}{\equiv} (x_1, ..., x_n) \in \mathbb{R}_+^n$ and call the choice incomes for brevity.[2] In most of the paper, we assume $n = p$ but we will briefly discuss both the case where $n < p$ and the case where $p > n$. Taxpayers pay a tax $T(\mathbf{x})$ that depends on all incomes in a potentially nonlinear way. Taxpayers who earn incomes $\mathbf{x}$ consume after-tax income $c = \sum_{i=1}^n x_i - T(x_1, ..., x_n)$ that we will refer to as consumption for short.

The preference of taxpayers of type $\mathbf{w}$ over consumption, and income choices is described by a twice continuously differentiable utility function $\mathcal{U}(c, \mathbf{x}; \mathbf{w})$ defined over $\mathbb{R}_+^{n+1} \times \mathcal{W}$. Taxpayers enjoy utility from consumption but endure disutility to obtain income so $\mathcal{U}_c > 0$ and $\mathcal{U}_{x_i} < 0$. All else being equal, taxpayers are better off when any exogenous characteristics is larger so $\mathcal{U}_{w_i} \geq 0$, with the equality strict for $\mathbf{x}$ $s.t.$ $\forall$ $x_i > 0$.

Let $\mathcal{C}(\cdot, \mathbf{x}; \mathbf{w})$ be the reciprocal of $\mathcal{U}(\cdot, \mathbf{x}; \mathbf{w})$. A taxpayer of type $\mathbf{w}$ earning incomes $\mathbf{x}$ should consume $\mathcal{C}(u, \mathbf{x}; \mathbf{w})$ to enjoy utility level u. We obviously get that $\mathcal{C}_u = 1/\mathcal{U}_c$ and $\mathcal{C}_{x_i} = -\mathcal{U}_{x_i}/\mathcal{U}_c$. We assume the utility function is weakly concave and indifference surfaces $\mathbf{x} \mapsto c = \mathcal{C}(u, \mathbf{x}; \mathbf{w})$ are strictly convex for all utility levels u and all types $\mathbf{w}$.

[2] Our model can also include observable actions like private expenditures in education. Such expenses can be defined as negative cash-flows and have a corresponding income equal to $-x_i$.

A taxpayer of type $\mathbf{w}$ solves:

$$U(\mathbf{w}) \overset{\text{def}}{\equiv} \max_{x_1,...,x_n} \quad \mathcal{U}\left(\sum_{i=1}^{n} x_i - T(x_1,...,x_n), x_1,...,x_n; \mathbf{w}\right) \tag{1}$$

Let $\mathbf{X}(\mathbf{w}) = (X_1(\mathbf{w}),...,X_n(\mathbf{w}))$ be the solution to this program and let $C(\mathbf{w}) = \sum_{i=1}^{n} X_i(\mathbf{w})) - T(\mathbf{X}(\mathbf{w}))$ be the consumption of taxpayers of type $\mathbf{w}$. Let the marginal rate of substitution between the i^{th} income and consumption at any bundle $(c, \mathbf{x})$ be defined as:

$$\mathcal{S}^i(c, \mathbf{x}; \mathbf{w}) \overset{\text{def}}{\equiv} -\frac{\mathcal{U}_{x_i}(c, \mathbf{x}; \mathbf{w})}{\mathcal{U}_c(c, \mathbf{x}; \mathbf{w})} > 0 \tag{2}$$

We can then write the first-order conditions for taxpayers of type $\mathbf{w}$ as:

$$\forall i \in \{1,...,n\} \qquad 1 - T_{x_i}(\mathbf{X}(\mathbf{w})) = \mathcal{S}^i(C(\mathbf{w}), \mathbf{X}(\mathbf{w}); \mathbf{w}) \tag{3}$$

II.2 Government

The government's resource constraint is given by:[3]

$$\iint_{\mathbf{w}\in\mathcal{W}} T(\mathbf{X}(\mathbf{w}))\, f(\mathbf{w})\mathrm{d}\mathbf{w} \geq E \tag{4}$$

where $E \geq 0$ is an exogenous amount of public expenditure. The government's objective is a social welfare function Φ that aggregates the utility of the individuals in the economy:

$$\mathcal{O} \overset{\text{def}}{\equiv} \iint_{\mathbf{w}\in\mathcal{W}} \Phi\left(U(\mathbf{w}); \mathbf{w}\right)\, f(\mathbf{w})\mathrm{d}\mathbf{w} \tag{5}$$

where the transformation $(u; \mathbf{w}) \mapsto \Phi(u; \mathbf{w})$ is twice continuously differentiable in $(u, \mathbf{w})$, increasing in u, weakly concave in U and potentially type-dependent. The government's problem consists of finding the tax function, $T(\cdot)$, that maximizes the social welfare function (5) subject to resource constraint (4), while taking into account individuals' optimization in (1) as a second constraint. Using Lagrangian methods, the budget constraint can be attached to the social welfare function with multipliers to form a Lagrangean:

$$\mathcal{L} = \iint_{\mathbf{w}\in\mathcal{W}} \left(T(\mathbf{X}(\mathbf{w})) + \frac{\Phi\left(U(\mathbf{w}); \mathbf{w}\right)}{\lambda}\right) f(\mathbf{w})\mathrm{d}\mathbf{w} - E \tag{6}$$

Were λ denotes the Lagrange multiplier associated to the resource constraint (4). Foloowing Saez (2001), we define the welfare weights of taxpayers of type $\mathbf{w}$ as the social marginal utility of consumption expressed in monetary terms:

$$g(\mathbf{w}) \overset{\text{def}}{\equiv} \frac{\Phi_u\left(U(\mathbf{w}); \mathbf{w}\right)\, \mathcal{U}_c\left(C(\mathbf{w}), \mathbf{X}(\mathbf{w}); \mathbf{w}\right)}{\lambda} \tag{7}$$

III The Tax Perturbation Approach

In this section we use the tax perturbation approach pioneered by Golosov et al. (2014) in a multidimensional context to characterize the optimal tax schedule.

[3] The sign $\iint$ is used instead of $\int$ to indicate the integral is over several dimensions when $n \geq 2$.

III.1 Responses to tax perturbations

We first investigate the responses to a tax perturbation. A tax perturbation is a twice continuously differentiable function $(\mathbf{x}, t) \mapsto \widetilde{T}(\mathbf{x}, t)$ such that, for all $\mathbf{x}$, we get $\widetilde{T}(\mathbf{x}, 0) = T(\mathbf{x})$. The parameter $t \in \mathbb{R}$ is a tax shifter that describes how far away the perturbed tax schedule $\mathbf{x} \mapsto \widetilde{T}(\mathbf{x}, t)$ is from the original tax schedule $\mathbf{x} \mapsto T(\mathbf{x})$.

Three types of perturbations are of particular interest. The first one is the *lump sum* perturbation which decreases tax liability by a uniform amount:

$$\widetilde{T}(\mathbf{x}, \rho) \stackrel{\text{def}}{\equiv} T(\mathbf{x}) - \rho \tag{8a}$$

where the tax shifter for such perturbation is denoted as ρ. The second type of tax perturbation considers a uniform change in the j^{th} marginal tax rate. This is the the *uncompensated perturbation of the j^{th} marginal tax rate*:

$$\widetilde{T}(\mathbf{x}, \tau_j) \stackrel{\text{def}}{\equiv} T(\mathbf{x}) - \tau_j \, x_j \tag{8b}$$

where the tax shifter for such perturbation is denoted as τ_j. The last perturbation, the *compensated perturbation of the j^{th} marginal tax rate for taxpayers of type* $\mathbf{w}$ is defined as:

$$\widetilde{T}(\mathbf{x}, \tau_j) \stackrel{\text{def}}{\equiv} T(\mathbf{x}) - \tau_j \left(x_j - X_j(\mathbf{w}) \right) \tag{8c}$$

where the tax shifter for such perturbation is denoted as τ_j. This perturbation is said to be "compensated for taxpayers of type $\mathbf{w}$" because it leaves the tax liability at incomes $\mathbf{x} = \mathbf{X}(\mathbf{w})$ unchanged.

Given a tax perturbation $(\mathbf{x}, t) \mapsto \widetilde{T}(\mathbf{x}, t)$, the utility of taxpayers of type $\mathbf{w}$ becomes a function of the tax shifter t through:

$$\widetilde{U}(\mathbf{w}, t) \stackrel{\text{def}}{\equiv} \max_{x_1, \ldots, x_n} \qquad \mathcal{U}\left(\sum_{i=1}^{n} x_i - \widetilde{T}(x_1, \ldots, x_n, t), x_1, \ldots, x_n; \mathbf{w} \right) \tag{9}$$

Let $\widetilde{\mathbf{X}}(\mathbf{w}, t)$ being the solution associated to (9). We obviously have $\widetilde{U}(\mathbf{w}, 0) = U(\mathbf{w})$ and $\widetilde{\mathbf{X}}(\mathbf{w}, 0) = \mathbf{X}(\mathbf{w})$ for all types $\mathbf{w} \in \mathcal{W}$. The first-order conditions associated to (9) are:

$$\forall i \in \{1, \ldots, n\} \qquad 1 - \widetilde{T}_{x_i}\left(\widetilde{\mathbf{X}}(\mathbf{w}, t), t \right) = \mathcal{S}^i \left(\sum_{i=1}^{n} \widetilde{X}_i(\mathbf{w}, t) - \widetilde{T}\left(\widetilde{\mathbf{X}}(\mathbf{w}, t), t \right), \widetilde{\mathbf{X}}(\mathbf{w}, t); \mathbf{w} \right) \tag{10}$$

We determine the behavioral responses to the tax perturbation $\widetilde{T}(\cdot, \cdot)$ by applying the implicit function theorem to first-order conditions (10). For this purpose, we need to impose the following assumption on the unperturbed tax schedule $T(\cdot)$.

Assumption 1. *Tax schedule $T(\cdot)$ verifies the following assumptions:*

i) The tax schedule $\mathbf{x} \mapsto T(\mathbf{x})$ is twice continuously differentiable.

ii) For each $\mathbf{w} \in \mathcal{W}$, the second-order conditions associated to (1) are strictly verified, i.e. the matrix $[\mathcal{S}^i_{x_j} + \mathcal{S}^j \mathcal{S}^i_c + T_{x_i x_j}]_{i,j}$ is positive definite at $c = C(\mathbf{w})$ and $\mathbf{x} = \mathbf{X}(\mathbf{w})$.[4]

iii) For each $\mathbf{w} \in \mathcal{W}$, function $\mathbf{x} \mapsto \mathcal{U}\left(\sum_{i=1}^n x_i - T(\mathbf{x}), \mathbf{x}; \mathbf{w}\right)$ admits a single global maximum.

Assumption 1.*i*) implies that marginal tax rates vary smoothly with income, as is the case in the German Income tax system. Assumption 1.*i*) rules out kinks like those in piecewise linear tax schedules. However, in reality we observe very little bunching at convex kinks in the tax schedule. This suggests that observed behavior can be described by assuming taxpayers base their decision on a smoothed approximation of the actual tax schedule, and smoothed approximations of piecewise linear schedules are twice-differentiable as assumed in 1.*i*). Assumption 1.*ii*) implies that second-order conditions associated to (1) are strictly verified. Parts *i*) and *ii*) of Assumption 1 together enable one to apply the implicit function theorem to determine how a local maximum of $\mathbf{x} \mapsto \mathcal{U}\left(\sum_{i=1}^n x_i - T(\mathbf{x}), \mathbf{x}; \mathbf{w}\right)$ is affected by a small tax perturbation or a small change in types. Assumption 1.iii rules out the existence of two different global maxima. This prevents an incremental tax perturbation from causing the taxpayer's choice to "jump" from one maximum to another. With such jumps, the derivative of $\mathbf{x}$ toward the perturbation t is undetermined, such that the perturbation approach could not be used.

Assumption 1 amounts to assuming that the tax function is at least as convex as utility. As the indifference surfaces $\mathbf{x} \mapsto c = \mathcal{C}(u, \mathbf{x}; \mathbf{w})$ are assumed convex, this is automatic if the tax function is linear, convex, or not "too" concave (See Appendix A.1). Geometrically, it implies that for each type, the indifference surface $\mathbf{x} \mapsto c = \mathcal{C}(U(\mathbf{w}), \mathbf{x}; \mathbf{w})$ associated to the utility level $U(\mathbf{w})$ admits a single tangency point with the budget set $c = \sum_{i=1}^n x_i - T(\mathbf{x})$ and lies strictly above the budget set otherwise.

Under Assumption 1, one can apply the implicit function theorem to determine the behavioral responses to small changes in taxation. Let us respectively denote by $\frac{\partial X_i(\mathbf{w})}{\partial \rho}$, $\frac{\partial X_i{}^u(\mathbf{w})}{\partial \tau_j}$ and $\frac{\partial X_i(\mathbf{w})}{\partial \tau_j}$ the response for taxpayers of type $\mathbf{w}$ of their i^{th} income to, respectively, the lump sum perturbation (8a), the uncompensated perturbation (8b) of the j^{th} marginal tax rate, and to the compensated perturbation (8c) of j^{th} marginal tax rate for taxpayers of type $\mathbf{w}$.[5] The tax perturbation $(\mathbf{x}, t) \mapsto \widetilde{T}(\mathbf{x}, t)$ affects the first-order conditions through the changes in marginal tax rates on the left-hand side of (10) and through the changes in tax liabilities that determine the marginal rates of substitutions on the right hand side of (10). Consequently, to a first-order approximation, the tax perturbation $(\mathbf{x}, t) \mapsto \widetilde{T}(\mathbf{x}, t)$ induces the same responses as the combination of a lump-sum perturbation (8a) times the perturbation $-\left.\frac{\partial \widetilde{T}(\mathbf{X}(\mathbf{w}), t)}{\partial t}\right|_{t=0}$ of tax liability with the n compensated perturbations of each of the j^{th} marginal tax rates (8c) times the per-

[4]We denote $[a(k)]_k$ a column vector whose k^{th} row is $a(k)$, $[A(k, \ell)]_{k,\ell}$ a square matrix of size n whose k^{th} row and ℓ^{th} column is $A(k, \ell)$ and $\cdot$ stands for the matrix product. The transpose operator is denoted with superscript T and the inverse operator is denoted with superscript -1.

[5]Strictly speaking, these responses do not just depend on types $\mathbf{w}$, but also on consumption $c = C(\mathbf{w})$ and incomes $\mathbf{x} = \mathbf{X}(\mathbf{w})$ of the evaluated types, as well as on the Hessian of the tax function.

turbation $-\left.\frac{\partial \widetilde{T}_{x_j}(\mathbf{X}(\mathbf{w}),t)}{\partial t}\right|_{t=0}$ of the j^{th} marginal tax rate. We thus get (See Appendix A.2):

$$\left.\frac{\partial \widetilde{X}_i(\mathbf{w},t)}{\partial t}\right|_{t=0} = \underbrace{-\frac{\partial X_i(\mathbf{w})}{\partial \rho}\left.\frac{\partial \widetilde{T}\left(\mathbf{X}(\mathbf{w}),t\right)}{\partial t}\right|_{t=0}}_{\text{Income responses}} \underbrace{-\sum_{j=1}^{n}\frac{\partial X_i(\mathbf{w})}{\partial \tau_j}\left.\frac{\partial \widetilde{T}_{x_j}\left(\mathbf{X}(\mathbf{w}),t\right)}{\partial t}\right|_{t=0}}_{\text{Compensated responses}} \tag{11}$$

In particular, we get the Slutsky Equation by applying (11) to the uncompensated perturbation (8b) of the j^{th} marginal tax rate.

$$\frac{\partial X_i{}^u(\mathbf{w})}{\partial \tau_j} = \frac{\partial X_i(\mathbf{w})}{\partial \tau_j} + X_j(\mathbf{w})\,\frac{\partial X_i(\mathbf{w})}{\partial \rho} \tag{12}$$

When the tax function is nonlinear, the responses to a tax reform generate changes in the marginal tax rates, which further induce compensated responses to these changes in marginal tax rates, etc (Saez, 2001). By applying the implicit function theorem, the income responses $\frac{\partial X_i(\mathbf{w})}{\partial \rho}$, uncompensated responses $\frac{\partial X_i{}^u(\mathbf{w})}{\partial \tau_j}$ and compensated responses $\frac{\partial X_i(\mathbf{w})}{\partial \tau_j}$ encapsulate this "circular process" through the endogeneity of marginal tax rates. We therefore refer to these responses are *total responses*. Since Equation (11) consists of total responses, it takes into account that marginal tax rates are endogenous. Conversely, the empirical literature typically estimates *direct* responses by assuming the tax schedule is linear, thus ignoring the circularity of the process (i.e. this circularity has no effect under the assumption that the tax schedule is linear). We discuss the relation between direct and total responses in Appendix A.3.

To analyze the effect of a tax perturbation on the government's budget constraint (4), we now compute the response of tax liabilities $\widetilde{T}(\widetilde{\mathbf{X}}(\mathbf{w},t),t)$ to tax shifter t. For each type of taxpayer, tax liabilities are modified in two ways. First, independently of any behavioral change, tax revenue is directly affected by the *mechanical* effect: $\left.\frac{\partial \widetilde{T}(\mathbf{X}(\mathbf{w}),t)}{\partial t}\right|_{t=0}$. Second, taxpayers of type $\mathbf{w}$ respond to the tax perturbation by changing all of their incomes $\mathbf{X}(\mathbf{w})$ through the n behavioral responses $\left.\frac{\partial \widetilde{X}_i(\mathbf{w})}{\partial t}\right|_{t=0}$, for $i = 1, ..., n$. For each type of income x_i, these responses modify tax liability by the product of its behavioral response to its marginal tax rate: $T_{x_i}(\mathbf{X}(\mathbf{w}))\left.\frac{\partial \widetilde{X}_i(\mathbf{w})}{\partial t}\right|_{t=0}$. The total change in tax liability due to the perturbation thus equals:

$$\left.\frac{\partial \widetilde{T}\left(\widetilde{\mathbf{X}}(\mathbf{w},t),t\right)}{\partial t}\right|_{t=0} = \underbrace{\left.\frac{\partial \widetilde{T}\left(\mathbf{X}(\mathbf{w}),t\right)}{\partial t}\right|_{t=0}}_{\text{Mechanical effects}} + \sum_{i=1}^{n}\underbrace{T_{x_i}(\mathbf{X}(\mathbf{w}))\left.\frac{\partial \widetilde{X}_i(\mathbf{w},t)}{\partial t}\right|_{t=0}}_{\text{Behavioral effects}} \tag{13}$$

Combining Equations (11) and (13) leads to:

$$\begin{aligned}\left.\frac{\partial \widetilde{T}\left(\widetilde{\mathbf{X}}(\mathbf{w},t),t\right)}{\partial t}\right|_{t=0} &= \left[1-\sum_{i=1}^{n}T_{x_i}(\mathbf{X}(\mathbf{w}))\frac{\partial X_i(\mathbf{w})}{\partial \rho}\right]\left.\frac{\partial \widetilde{T}\left(\mathbf{X}(\mathbf{w}),t\right)}{\partial t}\right|_{t=0} \\ &\quad - \sum_{1\le i,j\le n} T_{x_i}(\mathbf{X}(\mathbf{w}))\frac{\partial X_i(\mathbf{w})}{\partial \tau_j}\left.\frac{\partial \widetilde{T}_{x_j}\left(\mathbf{X}(\mathbf{w}),t\right)}{\partial t}\right|_{t=0}\end{aligned} \tag{14}$$

The tax perturbation affects the social objective only through the mechanical effect. This is because under Assumption 3 behavioral responses only induce second-order effects on taxpayers' utility (Saez, 2001). Applying the envelope theorem to social welfare $\Phi(U)$ after inserting (9) and using (7) leads to:

$$\frac{1}{\lambda} \left. \frac{\partial \Phi\left(\widetilde{U}(\mathbf{w},t);\mathbf{w}\right)}{\partial t} \right|_{t=0} = -g(\mathbf{w}) \left. \frac{\partial \widetilde{T}\left(\mathbf{X}(\mathbf{w}),t\right)}{\partial t} \right|_{t=0} \tag{15}$$

Let the perturbed Lagrangian be defined as:

$$\widetilde{\mathcal{L}}(t) \stackrel{\text{def}}{\equiv} \iint_{\mathbf{w}\in\mathcal{W}} \left\{ \widetilde{T}\left(\widetilde{\mathbf{X}}(\mathbf{w},t),t\right) + \frac{\Phi\left(\widetilde{U}(\mathbf{w},t);\mathbf{w}\right)}{\lambda} \right\} f(\mathbf{w})\mathrm{d}\mathbf{w} - E$$

To determine the value of λ, consider the effects of the lump-sum perturbation (8a) on the Lagrangian. At the optimum, the effect of the perturbation on the budget constraint is offset by its effects on the social objective. Using (7), this leads to:

$$\frac{\partial \widetilde{\mathcal{L}}(t)}{\partial t} = 0 \qquad \Leftrightarrow \qquad 0 = \iint_{\mathbf{w}\in\mathcal{W}} \left[1 - g(\mathbf{w}) - \sum_{i=1}^{n} T_{x_i}(\mathbf{X}(\mathbf{w})) \frac{\partial X_i(\mathbf{w})}{\partial \rho} \right] f(\mathbf{w})\mathrm{d}\mathbf{w} \tag{16}$$

A tax perturbation is not necessarily budget-balanced. To ensure the government budget is balanced we have to evaluate a tax perturbation in combination with a lump-sum rebate of the budget surplus induced by the tax perturbation. In Appendix A.4, we show the following result:

Proposition 1. *Under Assumption 1, combining a tax perturbation with $t > 0$ $\left(\text{respectively } t < 0\right)$ with a lump-sum rebate of the net budget surplus is welfare improving if and only if $\left.\frac{\partial \widetilde{\mathcal{L}}(t)}{\partial t}\right|_{t=0} > 0$ $\left(\text{resp. } \left.\frac{\partial \widetilde{\mathcal{L}}(t)}{\partial t}\right|_{t=0} < 0\right)$, where:*

$$\begin{aligned} \left. \frac{\partial \widetilde{\mathcal{L}}(t)}{\partial t} \right|_{t=0} = \iint_{\mathbf{w}\in\mathcal{W}} \Bigg\{ & \left[1 - g(\mathbf{w}) - \sum_{i=1}^{n} T_{x_i}(\mathbf{X}(\mathbf{w})) \frac{\partial X_i(\mathbf{w})}{\partial \rho} \right] \left. \frac{\partial \widetilde{T}\left(\mathbf{X}(\mathbf{w}),t\right)}{\partial t} \right|_{t=0} \\ & - \sum_{1\le i,j\le n} T_{x_i}(\mathbf{X}(\mathbf{w})) \frac{\partial X_i(\mathbf{w})}{\partial \tau_j} \left. \frac{\partial \widetilde{T}_{x_j}\left(\mathbf{X}(\mathbf{w}),t\right)}{\partial t} \right|_{t=0} \Bigg\} f(\mathbf{w})\mathrm{d}\mathbf{w} \end{aligned} \tag{17}$$

This proposition provides a condition for the desirability of tax reform in terms of empirically meaningful sufficient statistics.

III.2 Optimal tax formula

To derive the optimal tax formula, we need to extend the usual single crossing condition to our multidimensional setting:

Assumption 2. *$n = p$ and for each bundle $(c,\mathbf{x})$, the mapping $\mathbf{w} \mapsto \left(\mathcal{S}^1(c,\mathbf{x};\mathbf{w}),...,\mathcal{S}^n(c,\mathbf{x};\mathbf{w})\right)$ defined on $\mathcal{W}$ is injective and admits an invertible Jacobian.*

Assumption 2 is for instance verified under additively separable utility functions of the form:[6]

$$\mathcal{U}(c, \mathbf{x}; \mathbf{w}) = u(c) - \sum_{i=1}^{n} v^i(x_i, w_i) \qquad \text{where :} \qquad u', v^i_{x_i}, v^i_{x_i,x_i} > 0 > v^i_{x_i,w_i} \tag{18}$$

Combining Assumptions 1 and 2, we get:

Lemma 1. *Let $\mathcal{X} \stackrel{def}{\equiv} \{\mathbf{x} | \exists \mathbf{w} \in \mathcal{W} \ : \mathbf{x} = \mathbf{X}(\mathbf{w})\}$ be the range of incomes obtained. Under Assumptions 1 and 2, the mapping $\mathbf{w} \mapsto \mathbf{X}(\mathbf{w})$ is a bijection from $\mathcal{W}$ into $\mathcal{X}$.*

Proof: It is sufficient to show that the mapping $\mathbf{w} \mapsto \mathbf{X}(\mathbf{w})$ is injective. Assume there exists $\mathbf{x} \in \mathcal{X}$ and $\mathbf{w}, \widehat{\mathbf{w}} \in \mathcal{W}$ such that $\mathbf{X}(\mathbf{w}) = \mathbf{X}(\widehat{\mathbf{w}}) = \mathbf{x}$. From Assumption 1, the first-order conditions (3) have to be verified both at $(c, \mathbf{x}; \mathbf{w})$ and at $(c, \mathbf{x}; \widehat{\mathbf{w}})$, so we get $\mathcal{S}^i(c, \mathbf{x}, \mathbf{w}) = \mathcal{S}^i(c, \mathbf{x}, \widehat{\mathbf{w}})$ for all $i \in \{1, ..., n\}$. According to Assumption 2, these n equalities imply that $\mathbf{w} = \widehat{\mathbf{w}}$. □

To derive optimal tax formula, we have to define the (endogenous) density $h(\cdot)$ of incomes $\mathbf{X}(\mathbf{w})$. Via a change of variables we get:

$$\forall \mathbf{w} \in \mathcal{W} \qquad h(\mathbf{X}(\mathbf{w})) \stackrel{\text{def}}{\equiv} \frac{f(\mathbf{w})}{\left| \det \left[\frac{\partial X_i(\mathbf{w}))}{\partial w_j} \right]_{i,j} \right|} \tag{19}$$

We can then derive the optimal tax formula by rewriting (17) in terms of income densities and rearranging terms (see Appendix A.5).

Proposition 2. *Under Assumptions 1 and 2,[7] the optimal tax schedule verifies the Euler-Lagrange equation:*

$$\forall \mathbf{x} \in \mathcal{X} \qquad \left[1 - \overline{g(\mathbf{x})} - \sum_{i=1}^{n} T_{x_i}(\mathbf{x}) \overline{\frac{\partial X_i(\mathbf{x})}{\partial \rho}} \right] h(\mathbf{x}) = - \sum_{j=1}^{n} \frac{\partial \left[\sum_{i=1}^{n} T_{x_i}(\mathbf{x}) \overline{\frac{\partial X_i(\mathbf{x})}{\partial \tau_j}} h(\mathbf{x}) \right]}{\partial x_j} \tag{20a}$$

and the boundary conditions:

$$\forall \mathbf{x} \in \partial \mathcal{X} \qquad \sum_{1 \le i,j \le n} T_{x_i}(\mathbf{x}) \overline{\frac{\partial X_i(\mathbf{x})}{\partial \tau_j}} h(\mathbf{x}) \, e_j(\mathbf{x}) = 0 \tag{20b}$$

where $\overline{g(\mathbf{X}(\mathbf{w}))} = g(\mathbf{w})$, $\overline{\frac{\partial X_i(\mathbf{X}(\mathbf{w}))}{\partial \rho}} = \frac{\partial X_i(\mathbf{w})}{\partial \rho}$ *and* $\overline{\frac{\partial X_i(\mathbf{X}(\mathbf{w}))}{\partial \tau_j}} = \frac{\partial X_i(\mathbf{w})}{\partial \tau_j}$ *for all* $i, j \in \{1, ...n\}$ *and all* $\mathbf{w} \in \mathcal{W}$.

[6] When the utility function takes the form (18), we get that $\mathcal{S}^i(c, \mathbf{x}; \mathbf{w}) = v^i_{x_i}(x_i, w_i)/u'(c)$, which depends on $\mathbf{w}$ only through w_i. In such a case, Assumption 2 amounts to having the n one-dimensional mappings $w_i \mapsto v^i_{x_i}(x_i, w_i)/u'(c)$ being injective, which is guaranteed by $v^i_{x_i,w_i} < 0$.

[7] Assumptions 1 and 2 can be relaxed. Instead of Assumption 1, Golosov et al. (2014) assume that functions $t \mapsto \widetilde{\mathbf{X}}(\mathbf{w},)$ are Lipschitz continuous. Instead of Assumption 2, one can consider the case where $p > n$ provided that $\overline{\frac{\partial X_i(\mathbf{x})}{\partial \tau_j}}$, $\overline{\frac{\partial X_i(\mathbf{x})}{\partial \rho}}$ and $\overline{g(\mathbf{x})}$ are the means of $\frac{\partial X_i(\mathbf{w})}{\partial \tau_j}$, $\frac{\partial X_i(\mathbf{w})}{\partial \rho}$ and $g(\mathbf{w})$ among taxpayers whose type $\mathbf{w}$ is such that $\mathbf{X}(\mathbf{w}) = \mathbf{x}$, and provided that $h(\mathbf{x})$, $\overline{\frac{\partial X_i(\mathbf{x})}{\partial \tau_j}}$, $\overline{\frac{\partial X_i(\mathbf{x})}{\partial \rho}}$ and $\overline{g(\mathbf{x})}$ remain continuously differentiable in $\mathbf{x}$.

The Euler Lagrange Equation (20a) is a nonlinear second-order Partial Differential Equation. A more intuitive formulation can be obtained by integrating it on a subset $\Omega \subseteq \mathcal{X}$, with smooth boundary $\partial\Omega$ and outward unit normal vectors $\mathbf{e}(\mathbf{x}) = (e_1(\mathbf{x}), ..., e_n(\mathbf{x}))$. From the divergence theorem we get:

$$\oint_{\mathbf{x}\in\partial\Omega} \sum_{1\le i,j\le n} T_{x_i}(\mathbf{x}) \overline{\frac{\partial X_i(\mathbf{x})}{\partial \tau_j}} e_j(\mathbf{x})\, h(\mathbf{x}) d\Sigma(\mathbf{x}) = \iint_{\mathbf{x}\in\Omega} \left[1 - \overline{g(\mathbf{x})} - \sum_{i=1}^{n} T_{x_i}(\mathbf{x}) \overline{\frac{\partial X_i(\mathbf{x})}{\partial \rho}}\right] h(\mathbf{x}) d\mathbf{x} \quad (20c)$$

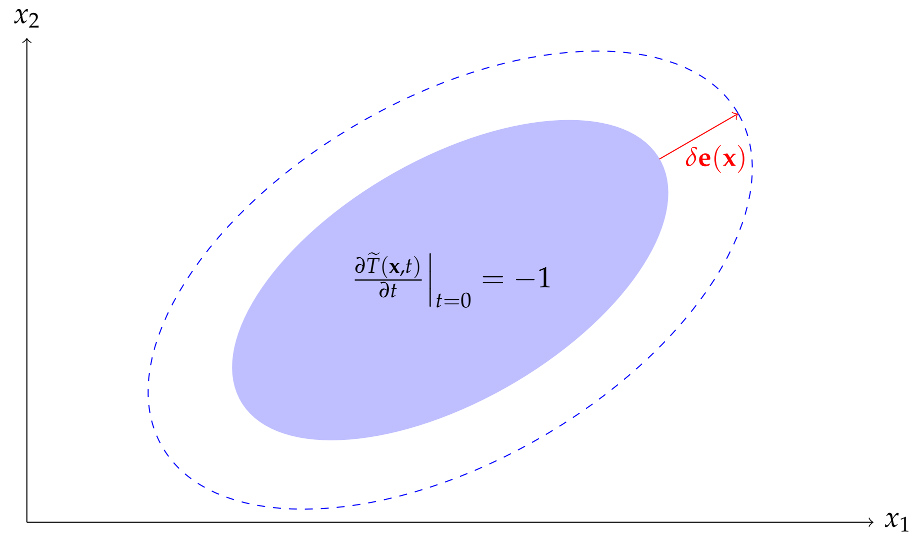

Figure 1: Intuition for Proposition 2

We now provide an heuristic derivation of Equation (20c) for the case with two incomes ($n = 2$). To do so, we extend the heuristic derivation of the optimal tax formula provided by Saez (2001) for the one dimensional case to the multidimensional case[8]. We consider a tax reform illustrated in Figure 1 that consists of two parts:

i. **Inside the subset of incomes** Ω (shaded area in Figure 1[9]): A perturbation that uniformly change (8a) for all individuals with incomes $\mathbf{x} \in \Omega$ before the reform. Using $\left.\frac{\partial \widetilde{T}(\mathbf{X}(\mathbf{w}),t)}{\partial t}\right|_{t=0} = -1$ and $\left.\frac{\partial \widetilde{T}_{x_i}(\mathbf{X}(\mathbf{w}),t)}{\partial t}\right|_{t=0} = 0$ inside Ω, only mechanical and income effects matter for types with incomes $\mathbf{X}(\mathbf{w}) \in \Omega$. From (14) and (15), the contributions of the mass $h(\mathbf{x})$ of taxpayers with initial income $\mathbf{x}$ inside Ω to the government's objective $\widetilde{\mathcal{L}}(t)$ is therefore given by:

$$\left[1 - \overline{g}(x) - \sum_{i=1}^{n} T_{x_i} \overline{\frac{\partial X_i(\mathbf{x})}{\partial \rho}}\right] h(\mathbf{x})$$

Integrating these effects over all incomes $\mathbf{x}$ inside the shade area Ω leads to the right-hand side of (20c).

[8]The derivation follows the graphical proof of Kleven et al. (2006) in the two dimensional case. We extend their proof as we do not assume quasilinear and additively separable preference, which they do. Golosov et al. (2014) provide graphical derivations which are very similar to ours.

[9]Note the area Ω does not have to be convex, unlike what might be suggested by Figure 1

ii. **Inside a ring of width δ around Ω** (The area between the shaded area and the dashed curve in Figure 1): The tax gradient $(T_{x_1}, ..., T_{x_n})$ must change to ensure tax liabilities are uniformly decreased by t inside Ω and are unchanged outside a ring of width δ around Ω and depicted in Figure 1 by the dashed line. For this purpose, along any radius normal to the Boundary $\partial\Omega$, the tax gradient $(T_{x_1}, ..., T_{x_n})$ has to be perturbed such that $\left.\frac{\partial \widetilde{T}_{x_i}(\mathbf{X}(\mathbf{w}),t)}{\partial t}\right|_{t=0} = -e_i(\mathbf{x})/\delta$ for all $i \in \{1, ..., n\}$, where $(e_1(\mathbf{x}), ..., e_n(\mathbf{x}))$ is the outward unit normal vector to $\partial\Omega$ at income $\mathbf{x}$. As the effects of changes in tax liabilities within the ring of width δ around Ω are of second-order compared to those inside Ω, we can approximate the tax perturbation in the ring by the n *compensated* tax perturbations (8c) times the size of the change $-e_i(\mathbf{x})/\delta$. This allows us to use $\left.\frac{\partial \widetilde{T}(\mathbf{X}(\mathbf{w}),t)}{\partial t}\right|_{t=0} \simeq 0$ to approximate the change in contribution to the governments objective $\widetilde{\mathcal{L}}(t)$ by taxpayers with initial income $\mathbf{x}$ inside the ring as:

$$\frac{1}{\delta} \sum_{1 \leq i,j \leq n} T_{x_i} \overline{\frac{\partial X_i(\mathbf{x})}{\partial \tau_j}} e_j(\mathbf{x}) \, h(\mathbf{x})$$

Integrating this expression, first along a radius of width δ normal to $\partial\Omega$, and second along the boundary $\partial\Omega$ of Ω leads to minus the left-hand side of (20c).

If the initial tax schedule is optimal, the substitution effects inside the ring of width δ around Ω must be exactly offset by the mechanical and income effects inside Ω, which leads to (20c).

III.3 Optimal taxation given isotax curves

The tax perturbation approach enables one to decompose the design of the optimal tax schedule $\mathbf{x} \mapsto T(\mathbf{x})$ into two steps. First, the design of isotax curves[10] which are the locus of incomes $\mathbf{x}$ that are associated with the same tax liability. Second, the assignment of a specific tax liability to each isotax curve. In this subsection, we show that the solution to the second step is characterized by a tax formula reminiscent of the ABC optimal formula of Saez (2001). This implies that the the most difficult and novel issue in multi-dimensional tax problems is the design of isotax curves.

For this purpose, we decompose the tax schedule $\mathbf{x} \mapsto T(\mathbf{x})$ in two consecutive mappings. The first one denoted Γ defines a taxable income denoted $y = \Gamma(\mathbf{x}) \in \mathbb{R}$ through the isolines associated to function Γ. The second mapping denoted $\mathcal{T}$ assigns a tax liability to each taxable income y so that we have $T(\mathbf{x}) = \mathcal{T}(\Gamma(\mathbf{x}))$. We then consider perturbations of the form $\widetilde{T}(\mathbf{x}, t) = \mathcal{T}(\Gamma(\mathbf{x})) + t \; p(\Gamma(\mathbf{x}))$, where $p(\cdot)$ stands for the direction of the tax perturbation. Therefore only function $\mathcal{T}$ is perturbed while isotax curves $y = \Gamma(\mathbf{x})$ are preserved. We then denote $Y(\mathbf{w}) = \Gamma(\mathbf{X}(\mathbf{w}))$ the realized taxable income for taxpayers of type $\mathbf{w}$ and $\widetilde{Y}(\mathbf{w}, t) = \Gamma(\widetilde{\mathbf{X}}(\mathbf{w}, t))$ the realized taxable income of taxpayers of type $\mathbf{w}$ under the perturbed tax schedule $(\mathbf{x}, t) \mapsto \widetilde{T}(\mathbf{x}, t)$.

[10]Formally these locus are "curves" only if $n = 2$. If $n = 3$, they are isotax *surfaces*. If $n \geq 4$ they are isotax *hypersurfaces*. For simplicity, we will stick to the terminology isotax curves.

Using (11), the tax liability perturbation (8a) where the direction is $p(y) = -1$ defines the income response of taxable income as:

$$\frac{\partial Y(\mathbf{w})}{\partial \rho} = \sum_{i=1}^{n} \Gamma_{x_i}(\mathbf{X}(\mathbf{w})) \frac{\partial X_i(\mathbf{w})}{\partial \rho} \tag{22a}$$

The compensated tax perturbation at taxable income $Y(\mathbf{w})$ in the direction $p(y) = Y(\mathbf{w}) - y$ defines the compensated response of taxable income for taxpayers of type $\mathbf{w}$:

$$\frac{\partial Y(\mathbf{w})}{\partial \tau} \stackrel{\text{def}}{\equiv} \sum_{1 \le i,j \le n} \Gamma_{x_i}(\mathbf{X}(\mathbf{w})) \frac{\partial X_i(\mathbf{w})}{\partial \tau_j} \Gamma_{x_j}(\mathbf{X}(\mathbf{w})) \tag{22b}$$

Let $m(\cdot)$ denote the density of taxable income Y and let $M(\cdot)$ the corresponding CDF. For the taxpayers whose type $\mathbf{w}$ is such that their taxable income is $Y(\mathbf{w}) = y$, let $\overline{\frac{\partial Y(y)}{\partial \tau}}$, $\overline{\frac{\partial Y(y)}{\partial \rho}}$ and $\overline{g}(y)$ denote the mean value of the compensated response $\frac{\partial Y(\mathbf{w})}{\partial \tau}$, the income response $\frac{\partial Y(\mathbf{w})}{\partial \rho}$ and the welfare weight $g(\mathbf{w})$ respectively. We show in Appendix A.7 that:

$$\forall y: \qquad \mathcal{T}'(y) \overline{\frac{\partial Y(y)}{\partial \tau}} m(y) = \int_{z=y}^{\infty} \left[1 - \overline{g}(z) - \mathcal{T}'(z) \overline{\frac{\partial Y(z)}{\partial \rho}} \right] m(z) \mathrm{d}z \tag{23a}$$

which can also be obtained by applying (20c) to a set Ω delimited by the isotax curve $\Gamma(\mathbf{x}) = y$, and:

$$0 = \int_{z=0}^{\infty} \left[1 - \overline{g}(z) - \mathcal{T}'(z) \overline{\frac{\partial Y(z)}{\partial \rho}} \right] m(z) \mathrm{d}z \tag{23b}$$

If $\mathcal{T}'(y) < 1$, then one can define the compensated elasticity at income y to be

$$\varepsilon(y) \stackrel{\text{def}}{\equiv} \frac{1 - \mathcal{T}'(y)}{y} \overline{\frac{\partial Y(y)}{\partial \tau}}$$

in which case Equation (23a) simplifies to

$$\frac{\mathcal{T}'(y)}{1 - \mathcal{T}'(y)} = \frac{1}{\varepsilon(y)} \frac{1 - M(y)}{y\, m(y)} \int_{z=y}^{\infty} \left[1 - \overline{g}(z) - \mathcal{T}'(z) \overline{\frac{\partial Y(z)}{\partial \rho}} \right] \frac{m(z)}{1 - M(y)} \mathrm{d}z \tag{23c}$$

which corresponds to Equation (19) in Saez (2001). This formula looks at distortions arising from a change in marginal tax rate in the neighborhood of iso-tax curve y. These distortions are proportional to the compensated elasticity $\varepsilon(y)$ and to $y\, m(y)$, in optimum these distortions should be offset by the sum of mechanical effects $1 - \overline{g}(z)$ and the income response effects $\mathcal{T}'(z) \overline{\frac{\partial Y(z)}{\partial \rho}}$ for all iso-tax curves z above y. Rewriting the integral in terms of the taxable income density conditional on incomes above y enables one to rewrite the optimal tax formula in terms of the local Pareto coefficient $y\, m(y)/(1 - M(y))$ of taxable income as in Saez (2001).

Since we can replicate known results from the uni-dimensional problem so readily, the difficulty of solving the multidimensional tax problem does not appear to lie in assigning a tax liability to a given isotax curve. This problem is reminiscent of solving an optimal tax problem with one observable income and n dimensions of heterogeneity, whose solution has already been described in Saez (2001), Hendren (2017) and Jacquet and Lehmann (2017) through Equations (equivalent to) (23a) and (23b). However, the complementing step of of designing the optimal shape of isotax curves is novel and causes new difficulties and requires a new approach that we will develop in the next sections.

Two observations are worth emphasizing at this point. First, Equations (23a) and (23b) are also valid if the design of isotax curves are suboptimal. This is clear from Appendix A.7, where both equations are derived without assuming isotax curves are set optimally. Second, the decomposition of the tax schedule into a definition of taxable income, $\Gamma(\cdot)$, and an assignment of taxable income to a tax liability, $\mathcal{T}(\cdot)$, is not unique. If α is a differentiable and increasing function, the same tax schedule can also be decomposed by defining taxable income as $\widehat{\Gamma}(\mathbf{x}) \stackrel{\text{def}}{\equiv} \alpha(\Gamma(\mathbf{x}))$ and assigning tax liability by $\widehat{\mathcal{T}}(\hat{y}) \stackrel{\text{def}}{\equiv} \mathcal{T}\left(\alpha^{-1}(\widehat{y})\right)$. It is straightforward to verify that Equations (23a) and (23b) remain valid provided that taxable income density, compensated responses and income response are adequately redefined.

IV The Mechanism Design approach

There is an alternative approach to characterize the optimal schedule that was pioneered by Mirrlees (1971) in the one-dimensional case and by Mirrlees (1976) in the multidimensional case. This method relies on the Taxation Principle (Hammond, 1979, Guesnerie, 1995) according to which it is equivalent for the government to select a tax function $\mathbf{x} \mapsto T(\mathbf{x})$ taking into account the taxpayers' decisions through (1), or to directly select an allocation $\mathbf{w} \mapsto (C(\mathbf{w}), \mathbf{X}(\mathbf{w}))$ that verifies the incentives constraints:

$$\forall \mathbf{w}, \hat{\mathbf{w}} \in \mathcal{W} \qquad U(\mathbf{w}) \stackrel{\text{def}}{\equiv} \mathcal{U}\left(C(\mathbf{w}), \mathbf{X}(\mathbf{w}); \mathbf{w}\right) \geq \mathcal{U}\left(C(\hat{\mathbf{w}}), \mathbf{X}(\hat{\mathbf{w}}); \mathbf{w}\right) \tag{24}$$

To satisfy (24), the government must select an allocation that assigns a bundle $(c, \mathbf{x})$ to each type of taxpayers such that each taxpayer is better of with the bundle assigned to her type than with any bundle assigned to any other type. Hammond (1979) shows that for each allocation $\mathbf{w} \mapsto (C(\mathbf{w}), \mathbf{X}(\mathbf{w}))$ verifying (24), there exists a tax schedule $\mathbf{x} \mapsto T(\mathbf{x})$ such that for each type $\mathbf{w}$, the solution of the taxpayer's program (1) corresponds to the bundle $(c(\mathbf{w}), \mathbf{X}(\mathbf{w}))$ designed for him.

Instead of dealing with the double continuum of inequalities in (24), Mirrlees (1971, 1976) adopted a First Order Mechanism Design approach (henceforth the FOMD). This approach consists of first considering only "smooth" allocations $\mathbf{w} \mapsto (C(\mathbf{w}), \mathbf{X}(\mathbf{w}))$ and second to consider only the following first-order incentive constraints:

$$\forall \mathbf{w} \in \mathcal{W}, \forall i \in \{1, ..., p\} \qquad \frac{\partial U(w)}{\partial w_i} = \mathcal{U}_{w_i}\left(C(\mathbf{w}), \mathbf{X}(\mathbf{w}); \mathbf{w}\right) \tag{25}$$

From the revelation principle we know that any mechanism is outcome equivalent to a direct revelation game where individuals are assumed to declare a type $\hat{\mathbf{w}}$ and consume the corresponding bundle from the smooth allocation $(C(\mathbf{w}), \mathbf{X}(\mathbf{w}))$. In such a direct revelation game the incentive constraints (25) are obtained by applying the envelope theorem to the left hand-side of (24). As such, it is equivalent to the first-order condition of maximizing the right hand side of (24) with respect to $\hat{\mathbf{w}}$.

In this section, we thus consider only "smooth" allocations that verify:

Assumption 3. *The allocation* $\mathbf{w} \mapsto (C(\mathbf{w}), \mathbf{X}(\mathbf{w}))$ *is continuously differentiable and verifies* (24).

Under Assumption 3, the FOMD consists in finding a continuously differentiable allocation that maximizes the government's Lagrangian with the same resource constraint: $\iint_{\mathbf{w}\in\mathcal{W}} \left\{\sum_{i=1}^{n} X_i(\mathbf{w}) - C(\mathbf{w})\right.$ E among continuously differentiable allocations that verify the first-order incentive constraints (25).

To solve the government's relaxed program using calculus of variations, we need to express the bundle $(C(\mathbf{x}), \mathbf{X}(\mathbf{w}))$ assigned to taxpayers of type $\mathbf{w}$ as a function of the utility $U(\mathbf{w})$ assigned to them and of the gradient of the mapping $\mathbf{w} \mapsto U(\mathbf{w})$. For this purpose, we first note that consumption can be re-expressed in terms of incomes and utility through $C(\mathbf{w}) = \mathcal{C}(U(\mathbf{w}), \mathbf{X}(\mathbf{w}); \mathbf{w})$. Second, we note that for each type $\mathbf{w}$ and each utility level u, the first-order incentive constraints (25) express the gradient of $\mathbf{w} \mapsto U(\mathbf{w})$ as a function of incomes trough:

$$\forall i \in \{1, ..., n\} \qquad z_i = \mathcal{U}_{w_i}\left(\mathcal{C}(u, \mathbf{x}; \mathbf{w}), \mathbf{x}; \mathbf{w}\right) \tag{26}$$

where z_i denotes the i^{th} component of the gradient of $\mathbf{w} \mapsto U(\mathbf{w})$, so that $U_{w_i}(\mathbf{w}) = z_i(\mathbf{w}) = \mathcal{U}_{w_i}\left(\mathcal{C}(U(\mathbf{w}), \mathbf{X}(\mathbf{w}); \mathbf{w}), \mathbf{X}(\mathbf{w}); \mathbf{w}\right)$ and $\mathbf{z} \stackrel{\text{def}}{\equiv} (z_1, ..., z_n)^T$. We then need to assume

Assumption 4. *For each utility level u and each type* $\mathbf{w} \in \mathcal{W}$*, the mapping*

$$\mathbf{x} \mapsto \left(\mathcal{U}_{w_1}\left(\mathcal{C}(u, \mathbf{x}; \mathbf{w}), \mathbf{x}; \mathbf{w}\right), ..., \mathcal{U}_{w_p}\left(\mathcal{C}(u, \mathbf{x}; \mathbf{w}), \mathbf{x}; \mathbf{w}\right)\right)^T$$

is a continuously differentiable bijection which admits an invertible Jacobian and whose reciprocal is denoted:

$$(u, \mathbf{z}; \mathbf{w}) \mapsto \left(\mathbb{X}_1(u, \mathbf{z}; \mathbf{w}), ..., \mathbb{X}_n(u, \mathbf{z}; \mathbf{w})\right)^T \tag{27}$$

The FOMD approach then consists of selecting the twice differentiable scalar field $\mathbf{w} \mapsto U(\mathbf{w})$ that maximizes

$$\iint_{\mathbf{w}\in\mathcal{W}} L\left(U(\mathbf{w}), U_{w_1}(\mathbf{w}), ..., U_{w_1}(\mathbf{w}); \mathbf{w}\right) \mathrm{d}\mathbf{w}$$

subject to incentive constraints (25), where

$$L\left(u, \mathbf{z}; \mathbf{w}, \lambda\right) \stackrel{\text{def}}{\equiv} \left[\sum_{i=1}^{n} \mathbb{X}_i(u, \mathbf{z}; \mathbf{w}) - \mathcal{C}\left(U, \mathbb{X}_1(u, \mathbf{z}; \mathbf{w}), ..., \mathbb{X}_n(u, \mathbf{z}; \mathbf{w}); \mathbf{w}\right) + \frac{\Phi(u)}{\lambda}\right] f(\mathbf{w}) - E \tag{28}$$

In Appendix B.1), we compute the partial (Gâteaux) derivative of L when the function $\mathbf{w} \mapsto U(\mathbf{w})$ is perturbed to $\mathbf{w} \mapsto U(\mathbf{w}) + t\, p(\mathbf{w})$, for any twice differentiable direction $p(\cdot)$. This lead us to the Euler-Lagrange Equation:

$$\forall \mathbf{w} \in \mathcal{W}: \qquad L_U \langle \mathbf{w} \rangle = \sum_{j=1}^{n} \frac{\partial L_{z_j} \langle \mathbf{w} \rangle}{\partial w_j} \tag{29a}$$

and to the Boundary conditions

$$\forall \mathbf{w} \in \partial \mathcal{W}: \qquad 0 = \sum_{j=1}^{n} L_{z_j} \langle \mathbf{w} \rangle \; e_j \tag{29b}$$

where the notation $\langle \mathbf{w} \rangle$ is a shortcut to denote that functions $C(\cdot), \mathbf{X}(\cdot), U(\cdot)$ are evaluated at $\mathbf{w}$. We then obtain:

Proposition 3. *The optimal allocation $\mathbf{w} \mapsto U(\mathbf{w})$ has to verify:*

$$\forall \mathbf{w} \in \mathcal{W} \quad : \quad \left(1 - \mathcal{S}^i \langle \mathbf{w} \rangle\right) f(\mathbf{w}) = \mathcal{U}_c \langle \mathbf{w} \rangle \sum_{j=1}^{n} \theta_j(\mathbf{w}) \, \mathcal{S}^i_{w_j} \langle \mathbf{w} \rangle \qquad \forall i \in \{1, ..., n\} \tag{30a}$$

$$\forall \mathbf{w} \in \mathcal{W} \quad : \quad \sum_{j=1}^{n} \frac{\partial \theta_j}{\partial w_j}(\mathbf{w}) = \left(\frac{1}{\mathcal{U}_c \langle \mathbf{w} \rangle} - \frac{\Phi_u(U(\mathbf{w}); \mathbf{w})}{\lambda} \right) f(\mathbf{w}) - \sum_{j=1}^{n} \theta_j(\mathbf{w}) \frac{\mathcal{U}_{c,w_j} \langle \mathbf{w} \rangle}{\mathcal{U}_c \langle \mathbf{w} \rangle} \tag{30b}$$

$$\forall \mathbf{w} \in \partial \mathcal{W} \quad : \quad 0 = \sum_{j=1}^{n} \theta_j(\mathbf{w}) e_j(\mathbf{w}) \tag{30c}$$

where we denote:

$$\forall \mathbf{w} \in \mathcal{W}: \qquad \theta_j(\mathbf{w}) \overset{def}{\equiv} -L_{z_j}\left(U(\mathbf{w}), U_{w_1}(\mathbf{w}), ..., U_{w_n}(\mathbf{w}); \mathbf{w}, \lambda\right) \qquad \forall j \in \{1, ..., n\} \tag{30d}$$

Equation (30b) is the Euler-Lagrange equation characterizing the optimal allocation for interior types. (30c) corresponds to the Boundary conditions that have to hold on the boundary of the type space $\partial \mathcal{W}$. These two necessary conditions are expressed in terms of the cost $\theta(w_j)$ of distorting the j^{th} component of the gradient of $\mathbf{w} \mapsto U(\mathbf{w})$ (see Equation (30d)).[11] These costs have to verify the system given by Equation (30a) in optimum. The conditions in Proposition 3 are equivalent to those obtained by Mirrlees (1976, 1986).

In Appendix B.2, we rewrite these conditions in terms of behavioral elasticities, type density and welfare weights. The Euler-Lagrange Equation becomes: $\forall \; \mathbf{w} \in \mathcal{W}$:

$$\left[1 - g(\mathbf{w}) - \sum_{i=1}^{n} T_{x_i}(\mathbf{X}(\mathbf{w})) \frac{\partial X_i(\mathbf{w})}{\partial \rho}\right] f(\mathbf{w}) = \sum_{j=1}^{n} \frac{\partial \sum_{i=1}^{n} \left(T_{x_i}(\mathbf{X}(\mathbf{w})) \, \mathcal{A}_{j,i}(\mathbf{w}) \, f(\mathbf{w})\right)}{\partial w_j} \tag{31a}$$

while the Boundary conditions becomes: $\forall \; \mathbf{w} \in \partial \mathcal{W}$:

$$\sum_{1 \le i,j \le n} T_{x_i}(\mathbf{X}(\mathbf{w})) \; \mathcal{A}_{j,i}(\mathbf{w}) \; e_j(\mathbf{w}) = 0 \tag{31b}$$

where matrix $\mathcal{A}$ is defined by:

$$\left[\mathcal{A}_{i,j}\right]_{i,j} \overset{\text{def}}{\equiv} \left[\mathcal{S}^i_{w_j}\right]^{-1}_{i,j} = -\left[\frac{\partial X_i(\mathbf{w})}{\partial w_j}\right]^{-1} \cdot \left[\frac{\partial X_i(\mathbf{w})}{\partial \tau_j}\right] \tag{31c}$$

where the last equality in (31c) is derived in Equation (34c) of Appendix A.2. In Appendix A.6, we also show how to retrieve this optimal tax formulas by applying the tax perturbation

[11] An alternative method to derive this equation is to attach incentive constraints to the Lagrangian in the form of equality constraints (26) with corresponding multipliers θ_i. The multipliers would then be the shadow price of the incentive constraint and their value would be defined by Equation (30d).

approach and expressing these conditions in terms of the type density instead of the income density. Hence, we show that the optimal conditions derived by Mirrlees (1976, 1986) through a mechanism design approach are consistent with the optimal conditions in terms of sufficient statistics derived by Golosov et al. (2014) using a tax perturbation approach. We thus confirm these two approaches are consistent, as Saez (2001) shows formally for the one-dimensional case. The tax perturbation approach and FOMD are the two faces of the same coin: while the FOMD computes the effects of a perturbing an allocation directly, the tax perturbation considers the effects of the tax perturbation that decentralizes the allocation. The tax perturbation thus indirectly deals with perturbed allocations and yields similar conditions for optimality as the direct perturbation.

The tax perturbation et MD approach actually requires different assumptions.

Even though the approaches yield consistent conditions, the mechanism approach has an advantage in that allows us to show the following result as is shown in Appendix B.3:

Proposition 4. *If for each type $\mathbf{w} \in \mathcal{W}$ and each $\lambda \in \mathbb{R}_+$ the mapping $(U, \mathbf{z}) \mapsto L(U, \mathbf{z}; \mathbf{w}, \lambda)$ is concave. If $\mathbf{w} \mapsto U^\star(\mathbf{w})$ verifies Equations (30), then $\mathbf{w} \mapsto U^\star(\mathbf{w})$ is the unique solution to the relaxed problem and there is no other $\mathbf{w} \mapsto U(\mathbf{w})$ that verifies (30).*

This result is especially important to derive for the numerical part because it ensures that, whenever $(U, \mathbf{z}) \mapsto L(U, \mathbf{z}; \mathbf{w}, \lambda)$ is concave, and an allocation verifies the necessary conditions, this allocation is the unique solution to the government's problem.

V Numerical simulations

In this section, we apply our framework to investigate optimal income taxation for couples. We consider an economy where couples differ in the productivity of males (w_1) and of females (w_2), so unobservable heterogeneity is bi-dimensional ($p = 2$). Each couple chooses the labor supply of both spouses, so there are two actions (i.e. $n = p = 2$). Following Kleven et al. (2007), preferences over the couple's consumption c, male income x_1 and female income x_2 are assumed to be quasilinear in consumption, additively separable and isoelastic in each income:

$$\mathcal{U}(c, x_1, x_2; w_1, w_2) = c - \varepsilon_1\, x_1^{\frac{1+\varepsilon_1}{\varepsilon_1}}\, w_1^{-\frac{1}{\varepsilon_1}} - \varepsilon_2\, x_2^{\frac{1+\varepsilon_2}{\varepsilon_2}}\, w_2^{-\frac{1}{\varepsilon_2}} \tag{32}$$

Therefore, income effects are assumed away (i.e. $\frac{\partial X_1(\mathbf{w})}{\partial \rho} = \frac{\partial X_2(\mathbf{w})}{\partial \rho} = 0$). Moreover, the cross responses equal zero (i.e. $\frac{\partial X_1^\star(\mathbf{w})}{\partial \tau_2} = \frac{\partial X_2(\mathbf{w})}{\partial \tau_1} = 0$), so that matrix of direct (compensated) responses is diagonal. Finally, ε_1 and ε_2 denote the direct elasticity of male and female income with respect to own net-of-marginal tax rates respectively.

Instead of constraining the income tax schedule to be *individual* of the form $T(x_1, x_2) = \mathcal{T}(x_1) + \mathcal{T}(x_2)$ or *joint* of the form $T(x_1, x_2) = \mathcal{T}(x_1 + x_2)$, we put no a priori restriction on the form of the tax schedule. We only presume the optimal schedule $(x_1, x_2) \mapsto T(x_1, x_2)$ verifies Assumption 1.

We calibrate a CARA social welfare function $\Phi(u, w_1, w_2) = -\exp(-\beta\, u)$, where $\beta > 0$ stands for the degree of inequality aversion.

With these functional specifications, Equation (25) implies that sensitivities are given by $z_i = (x_i/w_i)^{\frac{1+\varepsilon_i}{\varepsilon_i}}$, for $i = 1, 2$. Hence, we have that $\mathbb{X}_i(u, \mathbf{z}, \mathbf{w}) = w_i\, z_i^{\frac{\varepsilon_i}{1+\varepsilon_i}}$ and from (28) we get:

$$L(u, z_1, z_2; w_1, w_2, \lambda) = \left[w_1\, z_1^{\frac{\varepsilon_1}{1+\varepsilon_1}} - \varepsilon_1\, z_1\, w_1 + w_2\, z_2^{\frac{\varepsilon_2}{1+\varepsilon_2}} - \varepsilon_2\, z_2\, w_2 - u - \frac{\exp(-\beta\, u)}{\lambda} \right] f(w_1, w_2)$$

which is concave in (u, z_1, z_2). With a concave Lagrangian, Proposition 4 applies, meaning that optimal tax formulas are both necessary and sufficient.

We calibrate the type density, by applying a bivariate kernel to the joint distribution of male and female income among couples in the US Current Population Survey data (wave of March 2016, further CPS). With the empirical correlation, we deviate from the independent type distribution assumed in Kleven et al. (2007).

We then numerically solve for the optimal tax formulas (20a)-(20b) in an iterative process. To solve the equations we start with a guess for total elasticities $\left[\frac{\partial X_1(\mathbf{w})}{\partial \tau_1}, \frac{\partial X_1(\mathbf{w})}{\partial \tau_2}\right]_l = \frac{\partial X_2(\mathbf{w})}{\partial \tau_1}, \frac{\partial X_2(\mathbf{w})}{\partial \tau_2}$ and Lagrange multiplier λ_0. We then use the MATLAB PDE solver to solve (20a)-(20b). We then compute the Hessian of the tax function and use this to update total elasticities . With those values we update λ_{l+1}, derive new guesses for the total elasticities and repeat until convergence.

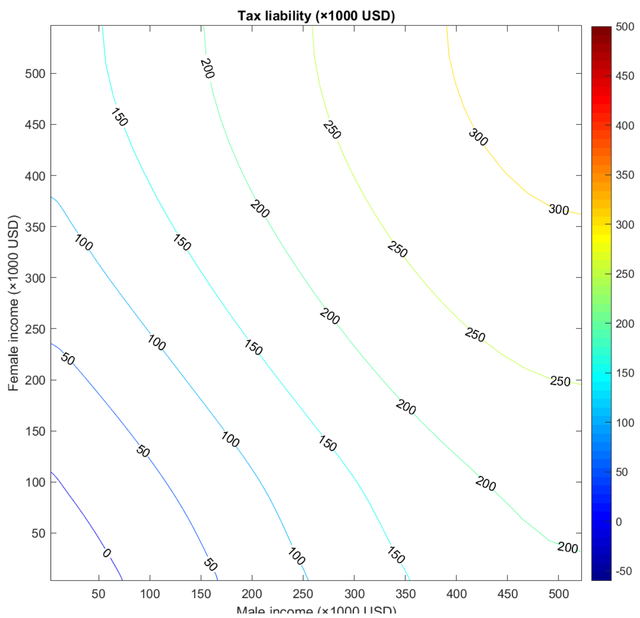

Figure 2: Isotax curves in the baseline case

Figure 2 displays the solution in our baseline economy with a direct elasticity $\varepsilon_1 = 0.25$ for males, direct elasticity $\varepsilon_1 = 0.5$ for females, and inequality aversion $\beta = 0.03$. Strikingly, isotax curves are almost linear and parallel, except where they approach the boundaries. There, the curvature of isotax curves are driven by the necessity to satisfy boundary constraints (20b).

A Appendices on the Tax Perturbation approach

A.1 Convexity of the indifference surfaces

We get from (2) that:

$$\mathcal{C}_{x_i}(u,\mathbf{x};\mathbf{w}) = -\frac{\mathcal{U}_{x_i}\left(\mathcal{C}(u,\mathbf{x};\mathbf{w}),\mathbf{x};\mathbf{w}\right)}{\mathcal{U}_c\left(\mathcal{C}(u,\mathbf{x};\mathbf{w}),\mathbf{x};\mathbf{w}\right)} = \mathcal{S}^i\left(\mathcal{C}(u,\mathbf{x};\mathbf{w}),\mathbf{x};\mathbf{w}\right)$$

The Hessian of the indifference surfaces is therefore a matrix whose i^{th} row and j^{th} column is:

$$\begin{aligned}\mathcal{C}_{x_i,x_j} &= -\frac{\mathcal{U}_{x_i,x_j}\mathcal{U}_c - \mathcal{U}_{c,x_i}\frac{\mathcal{U}_{x_j}}{\mathcal{U}_c}\mathcal{U}_c - \mathcal{U}_{c,x_j}\mathcal{U}_{x_i} + \mathcal{U}_{c,c}\frac{\mathcal{U}_{x_j}}{\mathcal{U}_c}\mathcal{U}_{x_i}}{\mathcal{U}_c^2} \\ &= -\frac{\mathcal{U}_{x_i,x_j}\mathcal{U}_c^2 - \mathcal{U}_{c,x_i}\mathcal{U}_{x_j}\mathcal{U}_c - \mathcal{U}_{c,x_j}\mathcal{U}_{x_i}\mathcal{U}_c + \mathcal{U}_{c,c}\mathcal{U}_{x_j}\mathcal{U}_{x_i}}{\mathcal{U}_c^3}\end{aligned}$$

On the other hand, from (2), we get:

$$\begin{aligned}\mathcal{S}^i_{x_j} + \mathcal{S}^j\mathcal{S}^i_c &= -\frac{\mathcal{U}_{x_i,x_j}\mathcal{U}_c - \mathcal{U}_{c,x_j}\mathcal{U}_{x_i}}{\mathcal{U}_c^2} + \frac{\mathcal{U}_{x_j}}{\mathcal{U}_c}\frac{\mathcal{U}_{c,x_i}\mathcal{U}_c - \mathcal{U}_{c,c}\mathcal{U}_{x_i}}{\mathcal{U}_c^2} \\ &= -\frac{\mathcal{U}_{x_i,x_j}\mathcal{U}_c^2 - \mathcal{U}_{c,x_i}\mathcal{U}_{x_j}\mathcal{U}_c - \mathcal{U}_{c,x_j}\mathcal{U}_{x_i}\mathcal{U}_c + \mathcal{U}_{c,c}\mathcal{U}_{x_j}\mathcal{U}_{x_i}}{\mathcal{U}_c^3} = \mathcal{C}_{x_i,x_j}\end{aligned}$$

Therefore matrix $\left[\mathcal{S}^i_{x_j} + \mathcal{S}^j\mathcal{S}^i_c\right]_{i,j}$ is symmetric and positive definite given the assumption that indifference surfaces are convex. Moreover, matrix $\left[\mathcal{S}^i_{x_j} + \mathcal{S}^j\mathcal{S}^i_c + T_{x_i,x_j}\right]_{i,j}$ is symmetric.

A.2 Behavioral Responses

Differentiating (10) with respect to t, $\mathbf{x}$ and $\mathbf{w}$ and using (3) leads to:

$$\begin{aligned}&\left[\mathcal{S}^i_{x_j} + \mathcal{S}^j\mathcal{S}^i_c + T_{x_ix_j}\right]_{i,j}\cdot[\mathrm{d}x_i]_i = \\ &\left[-\left.\frac{\partial\widetilde{T}_{x_i}(\mathbf{X}(\mathbf{w}),t)}{\partial t}\right|_{t=0}\right]_i \mathrm{d}t + \left[\mathcal{S}^i_c\right]_i \left.\frac{\partial\widetilde{T}(\mathbf{X}(\mathbf{w}),t)}{\partial t}\right|_{t=0}\mathrm{d}t - \left[\mathcal{S}^i_{w_j}\right]\cdot\left[\mathrm{d}w_j\right]_j\end{aligned} \tag{33}$$

where the various derivatives are evaluated at $\mathbf{x} = \mathbf{X}(\mathbf{w})$ and $c = C(\mathbf{w})$. From (8c), one has $\left.\frac{\partial\widetilde{T}(\mathbf{X}(\mathbf{w}),t)}{\partial t}\right|_{t=0} = 0$, $\left.\frac{\partial\widetilde{T}_{x_j}(\mathbf{X}(\mathbf{w}),t)}{\partial t}\right|_{t=0} = -1$ and $\left.\frac{\partial\widetilde{T}_{x_k}(\mathbf{X}(\mathbf{w}),t)}{\partial t}\right|_{t=0} = 0$ for $k \neq j$. The matrix of compensated responses is therefore given by:

$$\left[\frac{\partial X_i(\mathbf{w})}{\partial\tau_j}\right]_{i,j} = \left[\mathcal{S}^i_{x_j} + \mathcal{S}^j\mathcal{S}^i_c + T_{x_ix_j}\right]^{-1}_{i,j} \tag{34a}$$

The Matrix of compensated responses being the inverse of a symmetric and positive definite matrix, it is also symmetric and positive definite.

From (8a), one has $\left.\frac{\partial\widetilde{T}(\mathbf{X}(\mathbf{w}),t)}{\partial t}\right|_{t=0} = 0$, $\left.\frac{\partial\widetilde{T}_{x_j}(\mathbf{X}(\mathbf{w}),t)}{\partial t}\right|_{t=0} = -1$. The vector of income responses is therefore given by:

$$\left[\frac{\partial X_i(\mathbf{w})}{\partial\rho}\right]_i = -\left[\mathcal{S}^i_{x_j} + \mathcal{S}^j\mathcal{S}^i_c + T_{x_ix_j}\right]^{-1}_{i,j}\cdot\left[\mathcal{S}^i_c\right]_i = -\left[\frac{\partial X_i(\mathbf{w})}{\partial\tau_j}\right]_{i,j}\cdot\left[\mathcal{S}^i_c\right]_i \tag{34b}$$

Multiplying both sides of (33) by Matrix $\left[\mathcal{S}^i_{x_j} + \mathcal{S}^j\mathcal{S}^i_c + T_{x_ix_j}\right]^{-1}_{i,j}$ and using Equations (34) leads to (11). Using (8b) in (11) leads to (12).

Finally, the implicit function theorem ensure that the mapping $\mathbf{w} \mapsto \widetilde{\mathbf{X}}(\mathbf{w}, t)$ is differentiable for all $\mathbf{w} \in \mathcal{W}$ and for $t = 0$ with a Jacobian given by:

$$\left[\frac{\partial X_i(\mathbf{w})}{\partial w_j}\right]_{i,j} = -\left[\mathcal{S}^i_{x_j} + \mathcal{S}^j\mathcal{S}^i_c + T_{x_ix_j}\right]^{-1}_{i,j} \cdot \left[\mathcal{S}^i_{w_j}\right]_{i,j} = -\left[\frac{\partial X_i(\mathbf{w})}{\partial \tau_j}\right]_{i,j} \cdot \left[\mathcal{S}^i_{w_j}\right]_{i,j} \tag{34c}$$

A.3 Total versus Direct Responses

Let $\frac{\partial X_i^\star(\mathbf{w})}{\partial \rho}$ and $\frac{\partial X_i^\star(\mathbf{w})}{\partial \tau_j}$ denoted direct income and compensated direct responses and let $\frac{\partial X_i^\star(\mathbf{w})}{\partial w_j}$ denote the direct responses to a change in types if the tax schedule was linear. To understand the difference between direct and total responses, let $\Delta_1\mathbf{x}$ denote the change in income in response to a tax perturbation or a perturbation in types if the tax schedule was linear. This vector is obtained by taking $[T_{x_ix_j}]_{i,j} = 0$ in (33). We thus get:

$$\Delta_1\mathbf{x} = \left[\mathcal{S}^i_{x_j} + \mathcal{S}^j\mathcal{S}^i_c\right]^{-1}_{i,j} \cdot \mathrm{d}B$$

where $\mathrm{d}B$ is the vector column in the right-hand side of (33).

When the tax function is nonlinear, this "first" change $\Delta_1\mathbf{x}$ in income induces a change $[T_{x_i,w_j}]_{i,j} \cdot \Delta_1\mathbf{x}$ in the vector of marginal tax rates that generates a "second" change in income through compensated responses that are given by:

$$\Delta_2\mathbf{x} = -\left[\frac{\partial X_i^\star}{\partial \tau_j}\right]_{i,j} \cdot [T_{x_i,w_j}]_{i,j} \cdot \Delta_1\mathbf{x}$$

which in turn generates a further change in marginal tax rates. Hence, the k^{th} change in income $\Delta_k\mathbf{x}$ is related to $k-1^{\text{th}}$ change in income $\Delta_{k-1}\mathbf{x}$ by:

$$\Delta_k\mathbf{x} = -\left[\frac{\partial X_i^\star}{\partial \tau_j}\right]_{i,j} \cdot [T_{x_i,w_j}]_{i,j} \cdot \Delta_{k-1}\mathbf{x}$$

and so:

$$\Delta_k\mathbf{x} = \left(-\left[\frac{\partial X_i^\star}{\partial \tau_j}\right]_{i,j} \cdot [T_{x_i,w_j}]_{i,j}\right)^{k-1} \cdot \Delta_1\mathbf{x}$$

Adding all the effects leads to a total effect:

$$\begin{aligned}
\Delta\mathbf{x} &= \sum_{k=1}^{\infty} \Delta_k\mathbf{x} = \sum_{k=1}^{\infty} \left(-\left[\frac{\partial X_i^\star}{\partial \tau_j}\right]_{i,j} \cdot [T_{x_i,w_j}]_{i,j}\right)^{k-1} \cdot \Delta_1\mathbf{x} = \left(I_n + \left[\frac{\partial X_i^\star}{\partial \tau_j}\right]_{i,j} \cdot [T_{x_i,w_j}]_{i,j}\right)^{-1} \cdot \Delta_1\mathbf{x} \\
&= \left(I_n + \left[\mathcal{S}^i_{x_j} + \mathcal{S}^j\mathcal{S}^i_c\right]^{-1}_{i,j} \cdot [T_{x_i,w_j}]_{i,j}\right)^{-1} \cdot \left[\mathcal{S}^i_{x_j} + \mathcal{S}^j\mathcal{S}^i_c\right]^{-1}_{i,j} \cdot \mathrm{d}B \\
&= \left[\mathcal{S}^i_{x_j} + \mathcal{S}^j\mathcal{S}^i_c + T_{x_ix_j}\right]^{-1}_{i,j} \cdot \mathrm{d}B
\end{aligned}$$

where I_n denotes the identity matrix of rank n. We thus retrieve (11) and get:

$$\left[\frac{\partial X_i(\mathbf{w})}{\partial \tau_j}\right]_{i,j} = \left(I_n + \left[\frac{\partial X_i^\star}{\partial \tau_j}\right]_{i,j} \cdot [T_{x_i,w_j}]_{i,j}\right)^{-1} \cdot \left[\frac{\partial X_i^\star(\mathbf{w})}{\partial \tau_j}\right]_{i,j} \tag{35a}$$

$$\left[\frac{\partial X_i(\mathbf{w})}{\partial \rho}\right]_{i} = \left(I_n + \left[\frac{\partial X_i^\star}{\partial \tau_j}\right]_{i,j} \cdot [T_{x_i,w_j}]_{i,j}\right)^{-1} \cdot \left[\frac{\partial X_i^\star(\mathbf{w})}{\partial \rho}\right]_{i} \tag{35b}$$

$$\left[\frac{\partial X_i(\mathbf{w})}{\partial w_j}\right]_{i,j} = \left(I_n + \left[\frac{\partial X_i^\star}{\partial \tau_j}\right]_{i,j} \cdot [T_{x_i,w_j}]_{i,j}\right)^{-1} \cdot \left[\frac{\partial X_i^\star(\mathbf{w})}{\partial w_j}\right]_{i,j} \tag{35c}$$

It is worth noting that Part ii) of Assumption 1 is equivalent to assuming matrix $I_n + \left[\frac{\partial X_i^\star}{\partial \tau_j}\right]_{i,j} \cdot [T_{x_i,w_j}]_{i,j}$ is positive definite despite the nonlinearity of the tax schedule.

A.4 Proof of Lemma 1

For all t, let $b(t)$ denote the lump-sum rebate that makes budget balanced the tax perturbation $\widehat{T}(\cdot,\cdot)$ defined by:

$$\forall \mathbf{x}, \forall t \qquad \widehat{T}(\mathbf{x},t) = \widetilde{T}(\mathbf{x},t) - b(t)$$

Equations (11), (14) and (15) applies to the tax perturbation $\widehat{T}(\cdot,\cdot)$ if one substitutes $\left.\frac{\partial \widehat{T}(\mathbf{X}(\mathbf{w}),t)}{\partial t}\right|_{t=0}$ for $\left.\frac{\partial \widetilde{T}(\mathbf{X}(\mathbf{w}),t)}{\partial t}\right|_{t=0}$ and $\left.\frac{\partial \widehat{T}_{x_j}(\mathbf{X}(\mathbf{w}),t)}{\partial t}\right|_{t=0}$ for $\left.\frac{\partial \widetilde{T}_{x_j}(\mathbf{X}(\mathbf{w}),t)}{\partial t}\right|_{t=0}$.

Using:

$$\left.\frac{\partial \widehat{T}(\mathbf{x},t)}{\partial t}\right|_{t=0} = \left.\frac{\partial \widetilde{T}(\mathbf{x},t)}{\partial t}\right|_{t=0} + b'(0) \qquad \text{and} \qquad \left.\frac{\partial \widehat{T}_{x_j}(\mathbf{x},t)}{\partial t}\right|_{t=0} = \left.\frac{\partial \widetilde{T}_{x_j}(\mathbf{x},t)}{\partial t}\right|_{t=0}$$

and Equation (14), the tax perturbation $\widehat{T}(\cdot,\cdot)$ is budget-balanced if and only if:

$$\begin{aligned} 0 = & \iint_{\mathbf{w}\in\mathcal{W}} \Bigg\{ \left[1 - \sum_{i=1}^{n} T_{x_i}(\mathbf{X}(\mathbf{w})) \frac{\partial X_i(\mathbf{w})}{\partial \rho}\right] \left[\left.\frac{\partial \widetilde{T}\left(\mathbf{X}(\mathbf{w}),t\right)}{\partial t}\right|_{t=0} - b'(0)\right] \\ & - \sum_{1\le i,j\le n} T_{x_i}(\mathbf{X}(\mathbf{w})) \frac{\partial x_i(\mathbf{w})}{\partial \tau_j} \left.\frac{\partial \widetilde{T}_{x_j}\left(\mathbf{X}(\mathbf{w}),t\right)}{\partial t}\right|_{t=0} \Bigg\} f(\mathbf{w}) d\mathbf{w} \end{aligned}$$

The latter Equation can be rewritten as:

$$\begin{aligned} & b'(0) \iint_{\mathbf{w}\in\mathcal{W}} \left[1 - \sum_{i=1}^{n} T_{x_i}(\mathbf{X}(\mathbf{w})) \frac{\partial X_i(\mathbf{w})}{\partial \rho}\right] f(\mathbf{w}) d\mathbf{w} \\ = & \iint_{\mathbf{w}\in\mathcal{W}} \Bigg\{ \left[1 - \sum_{i=1}^{n} T_{x_i}(\mathbf{X}(\mathbf{w})) \frac{\partial X_i(\mathbf{w})}{\partial \rho}\right] \left.\frac{\partial \widetilde{T}\left(\mathbf{X}(\mathbf{w}),t\right)}{\partial t}\right|_{t=0} \\ & - \sum_{1\le i,j\le n} T_{x_i}(\mathbf{X}(\mathbf{w})) \frac{\partial x_i(\mathbf{w})}{\partial \tau_j} \left.\frac{\partial \widetilde{T}_{x_j}\left(\mathbf{X}(\mathbf{w}),t\right)}{\partial t}\right|_{t=0} \Bigg\} f(\mathbf{w}) d\mathbf{w} \end{aligned}$$

Using (16), we thus get:

$$
\begin{aligned}
& b'(0) \iint_{\mathbf{w}\in\mathcal{W}} g(\mathbf{w})\, f(\mathbf{w}) \mathrm{d}\mathbf{w} \\
= & \iint_{\mathbf{w}\in\mathcal{W}} \left\{ \left[1 - \sum_{i=1}^{n} T_{x_i}(\mathbf{X}(\mathbf{w})) \frac{\partial X_i(\mathbf{w})}{\partial \rho} \right] \left. \frac{\partial \widetilde{T}(\mathbf{X}(\mathbf{w}), t)}{\partial t} \right|_{t=0} \right. \\
- & \left. \sum_{1\le i,j\le n} T_{x_i}(\mathbf{X}(\mathbf{w})) \frac{\partial X_i(\mathbf{w})}{\partial \tau_j} \left. \frac{\partial \widetilde{T}_{x_j}(\mathbf{X}(\mathbf{w}), t)}{\partial t} \right|_{t=0} \right\} f(\mathbf{w}) \mathrm{d}\mathbf{w}
\end{aligned} \tag{36}
$$

Using Equations (5) and (15), the effects of the tax perturbation $\widehat{T}(\cdot,\cdot)$ on social welfare $t \mapsto \widehat{\mathcal{O}}(t)$ is given by:

$$
\frac{1}{\lambda} \left. \frac{\partial \widehat{\mathcal{O}}(t)}{\partial t} \right|_{t=0} = \iint_{\mathbf{w}\in\mathcal{W}} g(\mathbf{w}) \left[b'(0) - \left. \frac{\partial \widetilde{T}(\mathbf{X}(\mathbf{w}), t)}{\partial t} \right|_{t=0} \right] f(\mathbf{w}) \mathrm{d}\mathbf{w}
$$

Plugging (36) in the latter equation and using (17) leads finally to:

$$
\frac{1}{\lambda} \left. \frac{\partial \widehat{\mathcal{O}}(t)}{\partial t} \right|_{t=0} = \left. \frac{\partial \widetilde{\mathcal{L}}}{\partial t} \right|_{t=0}
$$

It is worth noting that (16) ensures that implementing the lump-sum perturbation (8a) and rebating the net surplus in a lump sum fashion (i.e. do not perturb the tax schedule, which has by definition no effect on the social objective $\widehat{\mathcal{O}}(\cdot)$) has no first-order effect on the Lagrangian $\widetilde{\mathcal{L}}(\cdot)$.

A.5 Optimal tax formula in income densities

Rewriting Equation (17) in terms of the income density $h(\cdot)$, rather than the type density $f(\cdot)$ leads to:

$$
\begin{aligned}
\left. \frac{\partial \widetilde{\mathcal{L}}(t)}{\partial t} \right|_{t=0} = & \iint_{\mathbf{x}\in\mathcal{X}} \left\{ \left[1 - \overline{g(\mathbf{x})} - \sum_{i=1}^{n} T_{x_i}(\mathbf{x}) \overline{\frac{\partial X_i(\mathbf{x})}{\partial \rho}} \right] \left. \frac{\partial \widetilde{T}(\mathbf{x}, t)}{\partial t} \right|_{t=0} \right. \\
& - \left. \sum_{1\le i,j\le n} T_{x_i}(\mathbf{x}) \overline{\frac{\partial X_i(\mathbf{x})}{\partial \tau_j}} \left. \frac{\partial \widetilde{T}_{x_j}(\mathbf{x}, t)}{\partial t} \right|_{t=0} \right\} h(\mathbf{x}) \mathrm{d}\mathbf{x}
\end{aligned}
$$

From the integration by parts corollary of the divergence theorem (See e.g. Theorem 13 page 13 in Basov (2005)), we get:

$$
\begin{aligned}
\left. \frac{\partial \widetilde{\mathcal{L}}(t)}{\partial t} \right|_{t=0} = & \oint_{\mathbf{x}\in\partial\mathcal{X}} \sum_{1\le i,j\le n} T_{x_i}(\mathbf{x}) \overline{\frac{\partial X_i(\mathbf{x})}{\partial \tau_j}} h(\mathbf{x})\, e_j(\mathbf{x}) \left. \frac{\partial \widetilde{T}(\mathbf{x}, t)}{\partial t} \right|_{t=0} \mathrm{d}\Sigma(\mathbf{x}) \\
+ & \iint_{\mathbf{x}\in\mathcal{X}} \left\{ \left[1 - \overline{g(\mathbf{x})} - \sum_{i=1}^{n} T_{x_i}(\mathbf{x}) \overline{\frac{\partial X_i(\mathbf{x})}{\partial \rho}} \right] h(\mathbf{x}) + \sum_{j=1}^{n} \frac{\partial \left[\sum_{i=1}^{n} T_{x_i}(\mathbf{x}) \overline{\frac{\partial X_i(\mathbf{x})}{\partial \tau_j}} h(\mathbf{x}) \right]}{\partial x_j} \right\} \left. \frac{\partial \widetilde{T}(\mathbf{x}, t)}{\partial t} \right|_{t=0} \mathrm{d}\mathbf{x}
\end{aligned}
$$

If the tax schedule $T(\cdot)$ is optimal, the above expression has to be nil for any tax perturbation, i.e. for any values of $\left. \frac{\partial \widetilde{T}(\mathbf{x},t)}{\partial t} \right|_{t=0}$. This is only possible if the Euler Lagrange Partial Differential Equation (20a) and the boundary conditions (20b) are satisfied.

A.6 Optimal tax formula in type density

We first need to redefine the perturbed tax schedule $\mathbf{x} \mapsto \widetilde{T}(\mathbf{x}, t)$ in the type space though the mapping $\mathbf{w} \mapsto \mathbf{X}(\mathbf{w})$. Importantly, the new coordinates corresponds to incomes in the not perturbed allocation. These coordinates thus don't depend on t. The perturbed tax schedule in the type space is thus defined: by $(\mathbf{w}, t) \mapsto \widetilde{\mathcal{T}}(\mathbf{w}, t) \stackrel{\text{def}}{\equiv} \widetilde{T}(\mathbf{X}(\mathbf{w}), t)$. We get: $\widetilde{\mathcal{T}}_{w_j}(\mathbf{w}, t) = \sum_{i=1}^{n} \frac{\partial X_i(\mathbf{w})}{\partial w_j} \widetilde{T}_{x_i}(\mathbf{X}(\mathbf{w}), t)$. Hence, in matrix term, we get:

$$\left[\widetilde{\mathcal{T}}_{w_j}(\mathbf{w}, t)\right]_j^T = \left[\widetilde{T}_{x_i}(\mathbf{X}(\mathbf{w}), t)\right]_i^T \cdot \left[\frac{\partial X_i(\mathbf{w})}{\partial w_j}\right]_{i,j} \qquad \Leftrightarrow \qquad \left[\widetilde{T}_{x_i}(\mathbf{X}(\mathbf{w}), t)\right]_i^T = \left[\widetilde{\mathcal{T}}_{w_j}(\mathbf{w}, t)\right]_j^T \cdot \left[\frac{\partial X_i(\mathbf{w})}{\partial w_j}\right]_{i,j}^{-1}$$

We thus get:

$$\left[\left.\frac{\partial \widetilde{T}_{x_j}(\mathbf{X}(\mathbf{w}), t)}{\partial t}\right|_{t=0}\right]_j^T = \left[\left.\frac{\partial \widetilde{\mathcal{T}}_{w_j}(\mathbf{w}, t)}{\partial t}\right|_{t=0}\right]_j^T \cdot \left[\frac{\partial X_i(\mathbf{w})}{\partial w_j}\right]_{i,j}^{-1}$$

Using the symmetry of matrix $\left[\frac{\partial X_i(\mathbf{w})}{\partial \tau_j}\right]_{i,j}$, the term $\sum_{1 \le i,j \le n} T_{x_i}(\mathbf{X}(\mathbf{w})) \frac{\partial X_i(\mathbf{w})}{\partial \tau_j} \left.\frac{\partial \widetilde{T}_{x_j}(\mathbf{X}(\mathbf{w}), t)}{\partial t}\right|_{t=0}$ in Equation (17) can be rewritten as:

$$\begin{aligned}
&\sum_{1 \le i,j \le n} T_{x_i}(\mathbf{X}(\mathbf{w})) \frac{\partial X_i(\mathbf{w})}{\partial \tau_j} \left.\frac{\partial \widetilde{T}_{x_j}(\mathbf{X}(\mathbf{w}), t)}{\partial t}\right|_{t=0} \\
= &\left[\left.\frac{\partial \widetilde{T}_{x_j}(\mathbf{X}(\mathbf{w}), t)}{\partial t}\right|_{t=0}\right]_j^T \cdot \left[\frac{\partial X_i(\mathbf{w})}{\partial \tau_j}\right]_{i,j} \cdot \left[T_{x_i}(\mathbf{X}(\mathbf{w}))\right]_i \\
= &\left[\left.\frac{\partial \widetilde{\mathcal{T}}_{w_j}(\mathbf{w}, t)}{\partial t}\right|_{t=0}\right]_j^T \cdot \left[\frac{\partial X_i(\mathbf{w})}{\partial w_j}\right]_{i,j}^{-1} \cdot \left[\frac{\partial X_i(\mathbf{w})}{\partial \tau_j}\right]_{i,j} \cdot \left[T_{x_i}(\mathbf{X}(\mathbf{w}))\right]_i \\
= &-\left[\left.\frac{\partial \widetilde{\mathcal{T}}_{w_j}(\mathbf{w}, t)}{\partial t}\right|_{t=0}\right]_j^T \cdot \left[\mathcal{S}_{w_j}^i\right]_{i,j}^{-1} \cdot \left[T_{x_i}(\mathbf{X}(\mathbf{w}))\right]_i
\end{aligned}$$

where the last Equality combines (34a) with (34c). Using the definition of Matrix $\mathcal{A}_{i,j}(\mathbf{w})$ in (31c), Equation (17) can be rewritten as:

$$\begin{aligned}
\left.\frac{\partial \widetilde{\mathcal{L}}(t)}{\partial t}\right|_{t=0} = &\iint_{\mathbf{w} \in \mathcal{W}} \left\{ \left[1 - g(\mathbf{w}) - \sum_{i=1}^{n} T_{x_i}(\mathbf{X}(\mathbf{w})) \frac{\partial X_i(\mathbf{w})}{\partial \rho}\right] \left.\frac{\partial \widetilde{\mathcal{T}}(\mathbf{w}, t)}{\partial t}\right|_{t=0} \right. \\
&\left. + \sum_{1 \le i,j \le n} T_{x_i}(\mathbf{X}(\mathbf{w}))\, \mathcal{A}_{j,i}(\mathbf{w}) \left.\frac{\partial \widetilde{\mathcal{T}}_{w_j}(\mathbf{w}, t)}{\partial t}\right|_{t=0} \right\} f(\mathbf{w}) d\mathbf{w}
\end{aligned}$$

Using the Divergence theorem, we get:

$$\begin{aligned}
\left.\frac{\partial \widetilde{\mathcal{L}}(t)}{\partial t}\right|_{t=0} = &\oint_{\mathbf{w} \in \partial \mathcal{W}} \sum_{1 \le i,j \le n} T_{x_i}(\mathbf{X}(\mathbf{w}))\, \mathcal{A}_{j,i}(\mathbf{w})\, e_j(\mathbf{w})\, f(\mathbf{w}) \left.\frac{\partial \widetilde{\mathcal{T}}(\mathbf{w}, t)}{\partial t}\right|_{t=0} d\Sigma(\mathbf{w}) \\
&+ \iint_{\mathbf{w} \in \mathcal{W}} \left\{ \left[1 - g(\mathbf{w}) - \sum_{i=1}^{n} T_{x_i}(\mathbf{X}(\mathbf{w})) \frac{\partial X_i(\mathbf{w})}{\partial \rho}\right] f(\mathbf{w}) \right. \\
&\left. - \sum_{j=1}^{n} \frac{\partial \sum_{i=1}^{n} T_{x_i}(\mathbf{X}(\mathbf{w}))\, \mathcal{A}_{j,i}(\mathbf{w})\, f(\mathbf{w})}{\partial w_j} \right\} \left.\frac{\partial \widetilde{\mathcal{T}}(\mathbf{w}, t)}{\partial t}\right|_{t=0} d\mathbf{w}
\end{aligned}$$

This Partial (Gâteaux) derivative is equal to zero for any tax perturbation $\widetilde{\mathcal{T}}(\cdot, \cdot)$ if and only if the Euler Lagrange Equation (31a) and Boundary conditions (31b) are verified.

A.7 Optimal Tax for given isotax curves

From $\widetilde{Y}(\mathbf{w},t) = \Gamma\left(\widetilde{\mathbf{X}}(\mathbf{w},t)\right)$ and Equations (11), (22a) and (22b), the response of taxable income to a generic tax perturbation $p(\cdot)$ is given by:

$$\left.\frac{\partial \widetilde{Y}(\mathbf{w},t)}{\partial t}\right|_{t=0} = -\frac{\partial Y(\mathbf{w})}{\partial \rho}\, p(Y(\mathbf{w})) - \frac{\partial Y(\mathbf{w})}{\partial \tau}\, p'(Y(\mathbf{w}))$$

The response of tax liability to a generic tax perturbation in the direction $p(\cdot)$ is thus given by:

$$\left.\frac{\partial \widetilde{T}\left(\widetilde{Y}(\mathbf{w},t),t\right)}{\partial t}\right|_{t=0} = \left[1-\mathcal{T}'(Y(\mathbf{w}))\frac{\partial Y(\mathbf{w})}{\partial \rho}\right] p(Y(\mathbf{w})) - \mathcal{T}'(Y(\mathbf{w}))\,\frac{\partial Y(\mathbf{w})}{\partial \tau}\, p'(Y(\mathbf{w}))$$

The response of the perturbed Lagrangian to a generic tax perturbation in the direction p is thus given by:

$$\begin{aligned}
\left.\frac{\partial \mathcal{L}(t)}{\partial t}\right|_{t=0} &= \iint_{\mathbf{w}\in\mathcal{W}} \left\{\left[1-g(\mathbf{w})-\mathcal{T}'(Y(\mathbf{w}))\frac{\partial Y(\mathbf{w})}{\partial \rho}\right] p(Y(\mathbf{w}))\right. \\
&- \left.\mathcal{T}'(Y(\mathbf{w}))\,\frac{\partial Y(\mathbf{w})}{\partial \tau}\, p'(Y(\mathbf{w}))\right\} f(\mathbf{w})\mathrm{d}\mathbf{w} \qquad (37) \\
&= \int_y \left\{\int_{Y(\mathbf{w})=y}\left\{\left[1-g(\mathbf{w})-\mathcal{T}'(y)\frac{\partial Y(\mathbf{w})}{\partial \rho}\right] p(y)\right.\right. \\
&- \left.\left.\mathcal{T}'(y)\,\frac{\partial Y(\mathbf{w})}{\partial \tau}\, p'(y)\right\} f\left(\mathbf{w}|Y(\mathbf{w})=y\right)\mathrm{d}\mathbf{w}\right\} m(y)\mathrm{d}y \qquad (38)
\end{aligned}$$

Note that Equation (37) can be obtain directly from (17) by noting that under the perturbation $\widetilde{T}(\mathbf{x},t) = \mathcal{T}\left(\Gamma(\mathbf{x})\right) + t\, p\left(\Gamma(\mathbf{x})\right)$ we have:

$$\left.\frac{\partial \widetilde{T}(\mathbf{x},t) = p\left(\Gamma(\mathbf{x})\right)}{\partial t}\right|_{t=0} \qquad \left.\frac{\partial \widetilde{T}_{x_j}(\mathbf{x},t)}{\partial t}\right|_{t=0} = p'\left(\Gamma(\mathbf{x})\right)\,\Gamma_{x_j}(\mathbf{x})$$

Using Equations (22a) and (22b) leads to (37).

Let

$$\overline{\frac{\partial Y(y)}{\partial \tau}} \stackrel{\text{def}}{\equiv} \int_{Y(\mathbf{w})=y} \frac{\partial Y(\mathbf{w})}{\partial \tau}\, f\left(\mathbf{w}|Y(\mathbf{w})=y\right)\mathrm{d}\mathbf{w} \qquad (39a)$$

denote the mean of the compensated elasticity among taxpayers of type $\mathbf{w}$ earning $Y(\mathbf{w}) = y$. Let similarly

$$\overline{\frac{\partial Y(y)}{\partial \rho}}(y) \stackrel{\text{def}}{\equiv} \int_{Y(\mathbf{w})=y} \frac{\partial Y(\mathbf{w})}{\partial \rho}\, f\left(\mathbf{w}|Y(\mathbf{w})=y\right)\mathrm{d}\mathbf{w} \qquad (39b)$$

the mean of the income responses among taxpayers of type $\mathbf{w}$ earning $Y(\mathbf{w}) = y$. Finally, let

$$\overline{g}(y) \stackrel{\text{def}}{\equiv} \int_{Y(\mathbf{w})=y} g(\mathbf{w})\, f\left(\mathbf{w}|Y(\mathbf{w})=y\right)\mathrm{d}\mathbf{w} \qquad (39c)$$

denote the mean of welfare weights among taxpayers of type $\mathbf{w}$ earning $Y(\mathbf{w}) = y$. Equation (38) simplifies to:

$$\left.\frac{\partial \mathcal{L}(t)}{\partial t}\right|_{t=0} = \int_y \left\{\left[1-\overline{g}(y)-\mathcal{T}'(y)\overline{\frac{\partial Y(y)}{\partial \rho}}\right] p(y) - \mathcal{T}'(y)\,\overline{\frac{\partial Y(y)}{\partial \tau}}\, p'(y)\right\} m(y)\mathrm{d}y \qquad (40)$$

Integrating by parts leads to:

$$\frac{\partial \mathcal{L}(t)}{\partial t}\Big|_{t=0} = \int_y \left\{ \int_{z=y}^{\infty} \left[1 - \overline{g}(z) - \mathcal{T}'(z)\overline{\frac{\partial Y(z)}{\partial \rho}}\right] m(z)dz - \mathcal{T}'(y)\, \overline{\frac{\partial Y(y)}{\partial \tau}}\, m(y) \right\} p'(y)dy$$
$$- \quad p(0) \int_{z=0}^{\infty} \left[1 - \overline{g}(z) - \mathcal{T}'(z)\overline{\frac{\partial Y(z)}{\partial \rho}}\right] m(z)dz \tag{41}$$

The effect of perturbation on the Lagrangian is nil whatever the direction p if and only if Equations (23a) and (23b) are valid.

Let $\alpha(\cdot)$ be an increasing mapping, let $\widehat{\Gamma}(\mathbf{x}) \stackrel{\text{def}}{\equiv} \alpha(\Gamma(\mathbf{x}))$ be an alternative definition of taxable income that we denote $\widehat{y} = \alpha(y)$ and let $\widehat{\mathcal{T}}(\hat{y}) \stackrel{\text{def}}{\equiv} \mathcal{T}\left(\alpha^{-1}(\widehat{y})\right)$ be the associated assignment of tax liability to taxable income. Finally let $\widehat{m}(\cdot)$ and $\widehat{M}(\cdot)$ be the PDF and CDF of $\widehat{y}$. We get

$$\widehat{\mathcal{T}}'(\widehat{y}) = \frac{\mathcal{T}'\left(\alpha^{-1}(\widehat{y})\right)}{\alpha'\left(\alpha^{-1}(\widehat{y})\right)} = \frac{\mathcal{T}'(y)}{\alpha'(y)}$$

Differentiating both sides of $\widehat{M}(\alpha(y)) = M(y)$ leads to:

$$\widehat{m}(\widehat{y}) = \frac{m(y)}{\alpha'(y)}$$

Applying respectively (22) and (22b) to $\widehat{Y}(\mathbf{w}) = \alpha(Y(\mathbf{w}))$ leads to:

$$\frac{\partial \widehat{Y}(\mathbf{w})}{\partial \rho} = \alpha'(Y(\mathbf{w}))\frac{\partial Y(\mathbf{w})}{\partial \rho} \qquad \text{and} \qquad \frac{\partial \widehat{Y}(\mathbf{w})}{\partial \tau} = \left(\alpha'(Y(\mathbf{w}))\right)^2 \frac{\partial Y(\mathbf{w})}{\partial \tau}$$

Hence

$$\widehat{\mathcal{T}}'(\widehat{y})\, \overline{\frac{\partial \widehat{Y}(\widehat{y})}{\partial \rho}} = \mathcal{T}'(y)\, \overline{\frac{\partial Y(y)}{\partial \rho}} \qquad \text{and} \qquad \widehat{\mathcal{T}}'(\widehat{y})\, \overline{\frac{\partial \widehat{Y}(\widehat{y})}{\partial \tau}}\, \widehat{m}(\widehat{y}) = \mathcal{T}'(y)\, \overline{\frac{\partial Y(y)}{\partial \tau}}\, m(y)$$

Therefore (23a) and (23b) are equivalent in terms of y or in terms of $\widehat{y}$.

B Appendices on the First-Order Mechanism Design approach (FOMD)

B.1 Proof of Proposition 3

Let p be a function defined over $\mathcal{W}$ into $\mathbb{R}$. We consider the effects of perturbing the allocation $\mathbf{w} \mapsto U(\mathbf{w})$ in the direction p. We hence consider the function:

$$\widetilde{L}^p(t) \stackrel{\text{def}}{\equiv} \iint_{\mathbf{w}\in\mathcal{W}} L\left(U(\mathbf{w}) + t\, p(\mathbf{w}), U_{w_1}(\mathbf{w}) + t\, p_{w_1}(\mathbf{w}), ..., U_{w_n}(\mathbf{w}) + t\, p_{w_n}(\mathbf{w}); \mathbf{w}, \lambda\right) d\mathbf{w} \tag{42}$$

Applying the chain rule and denoting $\langle \mathbf{w} \rangle$ as a shortcut to denote that a function is evaluated at $(U(\mathbf{w}), U_{w_1}(\mathbf{w}), ..., U_{w_n}(\mathbf{w}); \mathbf{w}, \lambda)$, we get using (30d):

$$\frac{\partial \widetilde{L}^p(t)}{\partial t}\Big|_{t=0} = \iint_{\mathbf{w}\in\mathcal{W}} \left\{ L_u \langle \mathbf{w} \rangle\, p(\mathbf{w}) - \sum_{j=1}^{n} \theta_j(\mathbf{w})\, p_{w_j}(\mathbf{w}) \right\} d\mathbf{w}$$

Applying the divergence theorem (See e.g. Theorem 13 page 13 in Basov (2005)) leads to:

$$\frac{\partial \widetilde{L}^p(t)}{\partial t}\Big|_{t=0} = \iint_{\mathbf{w}\in\mathcal{W}} \left\{ L_u \langle \mathbf{w} \rangle + \sum_{j=1}^{n} \frac{\partial \theta_j(\mathbf{w})}{\partial w_j} \right\} p(\mathbf{w})d\mathbf{w} - \oint_{\mathbf{w}\in\partial\mathcal{W}} \sum_{j=1} \theta_j(\mathbf{w}) e_j(\mathbf{w}) p(\mathbf{w}) d\Sigma(\mathbf{w})$$

At the optimal allocation, this expression is nil for any perturbation p. This implies that the Boundary conditions (30c) must hold together with the Euler Lagrange Equation:

$$\forall \mathbf{w} \in \mathcal{W}: \qquad \sum_{j=1}^{n} \frac{\partial \theta_j(\mathbf{w})}{\partial w_j} = -L_u \langle \mathbf{w} \rangle \tag{43}$$

Combining (26) and (28) leads to:

$$\left[\sum_{i=1} X_i(\mathbf{w}) - \mathcal{C}(U(\mathbf{w}), \mathbf{X}(\mathbf{w}); \mathbf{w}) + \frac{\Phi(U(\mathbf{w}); \mathbf{w})}{\lambda}\right] f(\mathbf{w}) = \tag{44}$$
$$L\left(U(\mathbf{w}), \mathcal{U}_{w_1}\left(\mathcal{C}(U(\mathbf{w}), \mathbf{X}(\mathbf{w}); \mathbf{w}), \mathbf{X}(\mathbf{w}); \mathbf{w}\right), ..., \mathcal{U}_{w_n}\left(\mathcal{C}(U(\mathbf{w}), \mathbf{X}(\mathbf{w}); \mathbf{w}), \mathbf{X}(\mathbf{w}); \mathbf{w}\right); \mathbf{w}, \lambda\right)$$

which are equivalent to Equations (29)

Differentiating both sides of (44) with respect to $X_i(\mathbf{w})$ and using (2), (3), $\mathcal{C}_{x_i} = \mathcal{S}^i$ and (30d) leads to:

$$\left(1 - \mathcal{S}^i \langle \mathbf{w} \rangle\right) f(\mathbf{w}) = -\sum_{j=1} \theta_j(\mathbf{w}) \left[\mathcal{U}_{x_i w_j} \langle \mathbf{w} \rangle + \mathcal{S}^i \langle \mathbf{w} \rangle \mathcal{U}_{c w_j} \langle \mathbf{w} \rangle\right]$$

which leads to (30a). Differentiating (44) with respect to $U(\mathbf{w})$ and using $\mathcal{C}_u = 1/\mathcal{U}_c$ and (30d) leads to:

$$\left(-\frac{1}{\mathcal{U}_c \langle \mathbf{w} \rangle} + \frac{\Phi_u(U(\mathbf{w}); \mathbf{w})}{\lambda}\right) f(\mathbf{w}) = L_u \langle \mathbf{w} \rangle - \sum_{j=1}^{n} \theta_j(\mathbf{w}) \frac{\mathcal{U}_{c,w_j} \langle \mathbf{w} \rangle}{\mathcal{U}_c \langle \mathbf{w} \rangle}$$

Plugging the latter equation into (43) leads to (30b).

B.2 Derivation of an optimal tax formula

Let use denote $\mu_j(\mathbf{w}) \stackrel{\text{def}}{\equiv} \theta_j(\mathbf{w})\, \mathcal{U}_c\left(C(\mathbf{w}), \mathbf{X}(\mathbf{w}); \mathbf{w}\right)$. Using (3), Equation (30a) leads to:

$$T_{x_i}(\mathbf{X}(\mathbf{w}))\, f(\mathbf{w}) = \sum_{j=1}^{n} \mu_j(\mathbf{w})\, \mathcal{S}^i_{w_j} \langle \mathbf{w} \rangle \tag{45}$$

This can be rewritten $[T_{x_i}(\mathbf{w})]_i\, f(\mathbf{w}) = \left[\mathcal{S}^i_{w_j}\right]_{i,j} \cdot \left[\mu_j(\mathbf{w})\right]_j$ in matrix term, which leads to: $\left[\mu_j(\mathbf{w})\right]_j = \left[\mathcal{S}^i_{w_j}\right]^{-1}_{i,j} \cdot [T_{x_i}(\mathbf{w})]_i\, f(\mathbf{w})$. We therefore get using (31c):

$$\forall \mathbf{w} \in \mathcal{W}, \forall i \in \{1, ..., n\} \qquad \mu_i(\mathbf{w}) = \sum_{i=1}^{n} \mathcal{A}_{i,j}(\mathbf{w})\, T_{x_j}(\mathbf{X}(\mathbf{w}))\, f(\mathbf{w}) \tag{46}$$

Combining Equation (30c) with (46) thus leads to (31b).

Using Equation (7), Equation (30b) implies that:

$$\begin{aligned}
\sum_{j=1}^{n} \frac{\partial \mu_j(\mathbf{w})}{\partial w_j} &= (1 - g(\mathbf{w})) f(\mathbf{w}) - \sum_{j=1}^{n} \theta_j(\mathbf{w})\, \mathcal{U}_{c,w_j} \langle \mathbf{w} \rangle \\
&+ \sum_{j=1}^{n} \theta_j(\mathbf{w}) \left[\mathcal{U}_{cc} \langle \mathbf{w} \rangle \frac{\partial C(\mathbf{w})}{\partial w_j} + \sum_{i=1}^{n} \mathcal{U}_{cx_i} \langle \mathbf{w} \rangle \frac{\partial X_i(\mathbf{w})}{\partial w_j} + \mathcal{U}_{c,w_j} \langle \mathbf{w} \rangle\right] \\
&= 1 - g(\mathbf{w})) f(\mathbf{w}) + \sum_{j=1}^{n} \theta_j(\mathbf{w}) \left[\mathcal{U}_{cc} \langle \mathbf{w} \rangle \frac{\partial C(\mathbf{w})}{\partial w_j} + \sum_{i=1}^{n} \mathcal{U}_{cx_i} \langle \mathbf{w} \rangle \frac{\partial X_i(\mathbf{w})}{\partial w_j}\right]
\end{aligned} \tag{47}$$

Differentiating $C(\mathbf{w}) = \mathcal{C}(U(\mathbf{w}), \mathbf{X}(\mathbf{w}); \mathbf{w})$ with respect to w_j and using $\mathcal{C}_u = 1/\mathcal{U}_c$, $\mathcal{C}_{x_i} = -\mathcal{U}_{x_i}/\mathcal{U}_c$, $\mathcal{C}_{w_j} = -\mathcal{U}_{w_j}/\mathcal{U}_c$ and (25) leads to:

$$\frac{\partial C(\mathbf{w})}{\partial w_j} = \frac{\mathcal{U}_{w_j} \langle \mathbf{w} \rangle}{\mathcal{U}_c \langle \mathbf{w} \rangle} - \sum_{i=1}^{n} \frac{\mathcal{U}_{x_i} \langle \mathbf{w} \rangle}{\mathcal{U}_c \langle \mathbf{w} \rangle} \frac{\partial X_i(\mathbf{w})}{\partial w_j} - \frac{\mathcal{U}_{w_j} \langle \mathbf{w} \rangle}{\mathcal{U}_c \langle \mathbf{w} \rangle} = -\sum_{i=1}^{n} \frac{\mathcal{U}_{x_i} \langle \mathbf{w} \rangle}{\mathcal{U}_c \langle \mathbf{w} \rangle} \frac{\partial X_i(\mathbf{w})}{\partial w_j}$$

Plugging this equality into (47) leads to

$$\begin{aligned}\sum_{j=1}^{n} \frac{\partial \mu_j(\mathbf{w})}{\partial w_j} &= (1-g(\mathbf{w}))f(\mathbf{w}) + \sum_{1\leq i,j\leq n} \theta_j(\mathbf{w}) \left[\mathcal{U}_{cx_i}\langle \mathbf{w}\rangle - \frac{\mathcal{U}_{x_i}\langle \mathbf{w}\rangle}{\mathcal{U}_c\langle \mathbf{w}\rangle}\mathcal{U}_{cc}\langle \mathbf{w}\rangle\right] \frac{\partial X_i(\mathbf{w})}{\partial w_j} \\ &= (1-g(\mathbf{w}))f(\mathbf{w}) - \sum_{1\leq i,j\leq n} \mu_j(\mathbf{w})\, \mathcal{S}_c^i\langle \mathbf{w}\rangle\, \frac{\partial X_i(\mathbf{w})}{\partial w_j} \end{aligned} \tag{48}$$

Combining (34a) with (34c) leads to $\frac{\partial X_i}{\partial w_j} = -\sum_{k=1}^{n} \frac{\partial X_i(\mathbf{w})}{\partial \tau_k} \mathcal{S}_{w_j}^k(\mathbf{w})$ which transforms (48) into:

$$\sum_{j=1}^{n} \frac{\partial \mu_j(\mathbf{w})}{\partial w_j} = (1-g(\mathbf{w}))f(\mathbf{w}) + \sum_{1\leq i,j,k\leq n} \mu_j(\mathbf{w})\, \mathcal{S}_c^i\langle \mathbf{w}\rangle\, \frac{\partial X_i(\mathbf{w})}{\partial \tau_k}\, \mathcal{S}_{w_j}^k(\mathbf{w}) \tag{49}$$

Combining (34a) with (34b) and using the symmetry of matrix $\left[\frac{\partial X_i(\mathbf{w})}{\partial \tau_j}\right]$ leads to $\frac{\partial X_k}{\partial \rho} = -\sum_{i=1}^{n} \frac{\partial X_i(\mathbf{w})}{\partial \tau_k}\mathcal{S}_c^i(\mathbf{w})$, which transforms (49) into:

$$\sum_{j=1}^{n} \frac{\partial \mu_j(\mathbf{w})}{\partial w_j} = (1-g(\mathbf{w}))f(\mathbf{w}) - \sum_{1\leq j,k\leq n} \mu_j(\mathbf{w})\, \mathcal{S}_{w_j}^k(\mathbf{w})\, \frac{\partial X_k(\mathbf{w})}{\partial \rho} \tag{50}$$

Plugging (45) into (50) leads to:

$$\sum_{j=1}^{n} \frac{\partial \mu_j(\mathbf{w})}{\partial w_j} = \left(1-g(\mathbf{w}) - \sum_{k=1}^{n} T_{x_k}(\mathbf{X}(\mathbf{w}))\, \frac{\partial X_k(\mathbf{w})}{\partial \rho}\right) f(\mathbf{w}) \tag{51}$$

Plugging (46) into (51) leads to (31a).

B.3 Proof of Proposition 4

If $(u, \mathbf{z}) \mapsto L(u, \mathbf{z}; \mathbf{w}, \lambda)$ is concave then for any perturbation p, function $t \mapsto \widetilde{L}^p(t)$ defined in (42) is concave. Let $\mathbf{w} \mapsto U(\mathbf{w})$ be another allocation and take the perturbation $p(\mathbf{w}) = U(\mathbf{w}) - U^\star(\mathbf{w})$. As the allocation $\mathbf{w} \mapsto U(\mathbf{w})$ verifies Equations (30), we get that function $t \mapsto \widetilde{L}^p(t)$ admits a zero derivative at $t = 0$ from and is concave. So $\widetilde{L}^p(0) > \widetilde{L}^p(1)$ and $U^\star(\cdot)$ provides a strictly higher welfare than $U(\cdot)$.

If by contradiction, two distinct allocations $\mathbf{w} \mapsto U^\star(\mathbf{w})$ and $\mathbf{w} \mapsto U(\mathbf{w})$ verify Equations (30) then, from the above reasoning $U(\cdot)$ strictly dominates $U^\star(\cdot)$ and $U^\star(\cdot)$ strictly dominates $U(\cdot)$, a contradiction. So at mowt one allocation can verify Equations (30).

C Appendix on the Numerical Simulations

To be written

References

Basov, Suren, *Multidimensional Screening* Studies in Economic Theory, Springer-Verlag Berlin Heidelberg, 2005.

Diamond, Peter, "Optimal Income Taxation: An Example with U-Shaped Pattern of Optimal Marginal Tax Rates," *American Economic Review*, 1998, *88* (1), 83–95.

Golosov, Mikhail, Aleh Tsyvinski, and Nicolas Werquin, "Dynamic Tax Reforms," NBER Working Papers 20 780 2014.

Guesnerie, Roger, *A Contribution to the Pure Theory of Taxation*, Cambridge: Cambridge University Press, 1995.

Hammond, Peter, "Straightforward Individual Incentive Compatibility in Large Economies," *The Review of Economic Studies*, 1979, *46* (2), 263–282.

Hendren, Nathaniel, "Efficient Welfare Weights," *NBER Working Papers*, 2017, (20351). Revise and Resubmit, *Journal of Political Economy*.

Jacquet, Laurence M. and Etienne Lehmann, "Optimal income taxation with composition effects," CEPR Discussion Papers 12276 2017.

Kleven, Henrik Jacobsen, Claus Thustrup Kreiner, and Emmanuel Saez, "The Optimal Income Taxation of Couples," NBER Working Papers 12685, National Bureau of Economic Research, Inc 2006.

_ , _ , and _ , "The Optimal Income Taxation of Couples as a Multi-Dimensional Screening Problem," CESifo Working Paper 2092, CESifo 2007.

_ , _ , and _ , "The Optimal Income Taxation of Couples," *Econometrica*, 2009, *77* (2), 537–560.

Mirrlees, J. A., "An Exploration in the Theory of Optimum Income Taxation," *Review of Economic Studies*, 1971, *38* (2), 175–208.

Mirrlees, James, "Optimal tax theory : A synthesis," *Journal of Public Economics*, 1976, *6* (4), 327–358.

_ , "The theory of optimal taxation," in K. J. Arrow and M.D. Intriligator, eds., *Handbook of Mathematical Economics*, Vol. 3, Elsevier, 1986, chapter 24, pp. 1197–1249.

Saez, Emmanuel, "Using Elasticities to Derive Optimal Income Tax Rates," *Review of Economic Studies*, 2001, *68* (1), 205–229.

A Mathematical Appendix not for publication

Multidimensional screening problems requires mathematical tools that are unusual to most economists. For the sake of self-completeness, this appendix states those results which are not new. Hence, this appendix is not intended to be published. For additional results and for proofs, an excellent reference is Basov (2005).

A mapping from a set $\mathcal{X}$ of $\mathbf{R^n}$ onto $\mathbf{R}$ is a *scalar field*. A mapping from a set $\mathcal{X}$ of $\mathbf{R}^n$ onto $\mathbf{R}^n$ is a *vector field*. When a scalar field $\varphi : \mathbf{x} \mapsto \varphi(\mathbf{x})$ is differentiable, the vector of its partial derivatives $\left(\frac{\partial \varphi}{\partial x_1}, ..., \frac{\partial \varphi}{\partial x_n}\right)$ is called the *gradient*. The gradient of a scalar field is therefore a vector field. When a vector field is the gradient of a scalar field, this vector field is said to be *conservative*. When $n = 1$, any continuous (vector) field is conservative, simply because any continuous function over an interval of $\mathbb{R}$ is integrable. This result does not easily generalize when $n > 1$. This is because if a scalar field is twice continuously differentiable, the Hessian matrix of its second order derivative has to be symmetric. This imposes a necessary condition on the partial derivatives of a vector field to be conservative. The following theorem ensures that this condition is also sufficient.

Theorem 1. *Let $\mathbf{f} : \mathbf{x} \mapsto \mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), ..., f_n(\mathbf{x})) \in \mathbb{R}^n$ be a continuously differentiable vector field defined over a convex subset $\mathcal{X}$ of $\mathbb{R}^n$. Then $\mathbf{f}$ is conservative if and only if:*

$$\forall \mathbf{x} \in \mathcal{X} \quad \textit{and} \quad \forall (i,j) \in \{1, ..., n\}^2 \qquad \frac{\partial f_i}{\partial x_j}(\mathbf{x}) = \frac{\partial f_j}{\partial x_i}(\mathbf{x})$$

The next result is the divergence theorem that generalizes the one-dimensional result according to which $f(b) - f(a) = \int_a^b f'(x)\mathrm{d}x$ to the case of multidimensional vector fields.

Theorem 2 (Divergence theorem). *Let $\mathbf{f} : \mathbf{x} \mapsto \mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), ..., f_n(\mathbf{x})) \in \mathbb{R}^n$ be a continuously differentiable vector field defined over a convex subset $\mathcal{X}$ of $\mathbb{R}^n$. Let Ω be a connected compact subset of $\mathcal{X}$ which admits a continuously differentiable boundary denoted $\partial\Omega$. For any $\mathbf{x}$ on the boundary $\partial\Omega$ of Ω, let $\mathbf{e}(\mathbf{x}) = (e_1(\mathbf{x}), ..., e_n(\mathbf{x}))$ denote the outward unit surface normal vector to the boundary of Ω. Then:*

$$\iint_{\mathbf{x}\in\Omega} \sum_{i=1}^{n} \frac{\partial f_i}{\partial x_i}(\mathbf{x})\ \mathrm{d}\Sigma(\mathbf{x}) = \oint_{\mathbf{x}\in\partial\Omega} \sum_{i=1}^{n} f_i(\mathbf{x}) e_i(\mathbf{x})\ \mathrm{d}\mathbf{x}$$

The Divergence theorem implies two important results.

Theorem 3 (Multidimensional integration by parts)**.** *Let $\mathbf{f} : \mathbf{x} \mapsto \mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), ..., f_n(\mathbf{x})) \in \mathbb{R}^n$ and $a : \mathbf{x} \mapsto a(\mathbf{x}) \in \mathbb{R}$ be respectively a vector and a scalar field that are continuously differentiable over a convex subset Ω of $\mathbb{R}^n$, with a smooth boundary $\partial\Omega$ and let $\mathbf{e}(\mathbf{x}) = (e_1(\mathbf{x}), ...,)$ be the outward unit surface normal vector defined on the boundary $\partial\Omega$. Then:*

$$\iint_{\mathbf{x}\in\Omega} \left(\sum_{i=1}^{n} f_i(\mathbf{x}) \frac{\partial a}{\partial x_i}(\mathbf{x}) \right) \mathrm{d}\mathbf{x} = \oint_{\mathbf{x}\in\partial\Omega} \sum_{i=1}^{n} a(\mathbf{x}) f_i(\mathbf{x}) e_i(\mathbf{x}) \mathrm{d}\Sigma(\mathbf{x}) - \iint_{\mathbf{x}\in\Omega} \left(\sum_{i=1}^{n} \frac{\partial f_i}{\partial x_i}(\mathbf{w}) a(\mathbf{x}) \right) \mathrm{d}\mathbf{x}$$

Proof: Let $\mathbf{Q}(\mathbf{x}) = (Q_1(\mathbf{x}), ..., Q_n(\mathbf{x})) \stackrel{\text{def}}{\equiv} a(\mathbf{x})\mathbf{f}(\mathbf{x}) = (a(\mathbf{x})f_1(\mathbf{x}), ..., a(\mathbf{x})f_n(\mathbf{x}))$. Since $\frac{\partial Q_i}{\partial x_i}(\mathbf{x}) = \frac{\partial a}{\partial x_i}(\mathbf{x}) f_i(\mathbf{x}) + a(\mathbf{x}) \frac{\partial f_i}{\partial x_i}(\mathbf{x})$, applying the divergence theorem to $\mathbf{Q}$ gives the result. □

Theorem 4 (Integral form of a divergence PDE). *Let $a : \mathbf{x} \mapsto a(\mathbf{x}) \in \mathbf{R}$ and $\mathbf{f} : \mathbf{x} \mapsto \mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), ..., f_n(\mathbf{x})) \in \mathbb{R}^n$ be respectively a continuously differentiable scalar and vector fields defined over a convex subset $\mathcal{X}$ of $\mathbb{R}^n$. Then f verifies the divergence partial differential equation:*

$$\sum_{i=1}^{n} \frac{\partial f_i}{\partial x_i}(\mathbf{x}) = a(\mathbf{x})$$

if and only if it verifies the integrated divergence equation

$$\oint_{\mathbf{x}\in\partial\Omega} \sum_{i=1}^{n} f_i(\mathbf{x}) e_i(\mathbf{x})\ \mathrm{d}\Sigma(\mathbf{x}) = \iint_{\mathbf{x}\in\Omega} a(\mathbf{x})\ \mathrm{d}\mathbf{x}$$

for any connected compact subset Ω of $\mathcal{X}$ which admits a continuously differentiable boundary denoted $\partial\Omega$ with an outward unit surface normal vector to the boundary of Ω denoted $\mathbf{e}(\mathbf{x}) = (e_1(\mathbf{x}), ..., e_n(\mathbf{x}))$.

Proof: This result is obtained by applying the Divergence theorem to left-hand side of the *divergence* partial differential equation. For the reciprocal, applying the Divergence theorem to left-hand side of the integrated *divergence* equation leads to:

$$\iint_{\mathbf{x}\in\Omega} \sum_{i=1}^{n} \frac{\partial f_i}{\partial x_i}(\mathbf{x})\ \mathrm{d}\mathbf{x} = \iint_{\mathbf{x}\in\Omega} a(\mathbf{x})\ \mathrm{d}\mathbf{x}$$

As this equality should be verified for any smooth compact subset of $\mathcal{X}$, and f is assumed continuously differentiable, then the two integrand must be equal everywhere inside Ω, which leads to the divergence PDE. $\square$