Appendix A. First order conditions (3.3)-(3.4)

Consider the population version of the DRO problem (3.1)

$$\max_{x,c} G(x,c)$$

where

$$G(\delta,x,c) = -\frac{1}{\delta}\mathbb{E}_{\mathbb{P}}\Big[\phi^*\Big(-\delta\big[f(x,Y)+c\big]\Big)\Big] - c. \tag{A.1}$$

Differentiating with respect to x, the first order conditions are

$$\begin{aligned} \nabla_x G(\delta,x,c) &= \mathbb{E}_{\mathbb{P}}\Big[[\phi^*]'\Big(-\delta\big[f(x,Y)+c\big]\Big)\nabla_x f(x,Y)\Big] = 0, \\ \nabla_c G(\delta,x,c) &= \mathbb{E}_{\mathbb{P}}\Big[[\phi^*]'\Big(-\delta\big[f(x,Y)+c\big]\Big)\Big] - 1 = 0 \end{aligned} \tag{A.2}$$

with solutions $(x^\star(\delta), c^\star(\delta))$. We eventually wish to show that $(x^\star(\delta), c^\star(\delta))$ is continuously differentiable in a neighborhood of $\delta = 0$, which depends on the properties of the convex conjugate $\phi^*(\zeta)$. Under Assumption (3.1) (see Theorem 3.2 in [12]) $\phi^*(\zeta)$ is twice continuously differentiable in the neighborhood of $\zeta = 0$ and satisfies

$$\phi^*(\zeta) = \zeta + \frac{1}{2!}\Big(\frac{1}{\phi''(1)}\Big)\zeta^2 + o(\zeta^2). \tag{A.3}$$

It follows that $[\phi^*]'(\zeta)$ is continuously differentiable in a neighborhood of $\zeta = 0$ and

$$[\phi^*]'(\zeta) = 1 + \frac{\zeta}{\phi''(1)} + o(\zeta), \tag{A.4}$$

so by (A.2)

$$\nabla_c G(\delta,x,c) = \mathbb{E}_{\mathbb{P}}\Big[-\frac{\delta}{\phi''(1)}\big[f(x,Y)+c\big]\Big] + o(\delta)$$

and we can write the first order conditions as

$$\begin{aligned} \nabla_x G(\delta,x,c) &= \mathbb{E}_{\mathbb{P}}\Big[[\phi^*]'\Big(-\delta\big[f(x,Y)+c\big]\Big)\nabla_x f(x,Y)\Big] = 0, \\ -\frac{\phi''(1)}{\delta}\nabla_c G(\delta,x,c) &= \mathbb{E}_{\mathbb{P}}\Big[-\frac{\phi''(1)}{\delta}\Big\{[\phi^*]'\Big(-\delta\big[f(x,Y)+c\big]\Big) - 1\Big\}\Big] = 0. \end{aligned}$$

This does not change the solution of (A.2), but allows us to use the Implicit Function Theorem to establish level of smoothness for $(x^\star(\delta), c^\star(\delta))$ in Proposition 3.2 (in particular, that (B.3) is invertible in the proof in Appendix B). It follows that $\psi(x,c,Y)$ is given by (3.5). First order conditions for the sample DRO problem, and the DOO problems can be derived similarly.

Appendix B. Proof of Proposition 3.2

We use the Implicit Function Theorem to show existence of the solution $(x^\star(\delta),\, c^\star(\delta))$ of (3.4) and to characterize its smoothness. (Existence and smoothness of $(x_n(\delta),\, c_n(\delta))$ can be shown in a similar way). With a mild abuse of notation, let

$$g(\delta,\, x,\, c) \equiv \begin{bmatrix} g_1(\delta,\, x,\, c) \\ g_2(\delta,\, x,\, c) \end{bmatrix} := \mathbb{E}_{\mathbb{P}}[\psi(x,\, c,\, Y)].$$

The first-order conditions for the population problems (3.4) are

$$g(\delta,\, x,\, c) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \tag{B.1}$$

Since $[\phi^*]'(\zeta)$ is continuously differentiable in a neighborhood of $\zeta = 0$ (Theorem 3.2 in [12]), $g(\delta,\, x,\, c)$ is continuously differentiable in δ in some open neighborhood of $\delta = 0$ for every fixed $(x,\, c)$. By (A.4)

$$g(\delta,\, x,\, c) = \begin{bmatrix} \mathbb{E}_{\mathbb{P}}\Big[\nabla_x f(x, Y)\Big] - \frac{\delta}{\phi''(1)}\mathbb{E}_{\mathbb{P}}\Big[\Big(f(x, Y) + c\Big)\nabla_x f(x, Y)\Big] + o(\delta) \\ \mathbb{E}_{\mathbb{P}}\Big[f(x, Y) + c\Big] + O(\delta) \end{bmatrix},$$

and

$$g\big(0,\, x^\star(0),\, -\mathbb{E}_{\mathbb{P}}[f(x^\star(0),\, Y)]\big) = 0, \tag{B.2}$$

where $x^\star(0)$ is the solution of the empirical problem. Since $f(x,\, Y)$ is twice continuously differentiable in x, $g(\delta,\, x,\, c)$ is continuously differentiable in a neighborhood of $\big(0,\, x^\star(0),\, -\mathbb{E}_{\mathbb{P}}[f(x^\star(0),\, Y)]\big)$, and

$$\begin{aligned} J_{g,\,(x,c)}(\delta,\, x,\, c)\Big|_{(0,\, x^\star(0),\, -\mathbb{E}_{\mathbb{P}}[f(x^\star(0),\, Y)])} &\equiv \left.\begin{bmatrix} \nabla_x g_1(\delta,\, x,\, c) & \nabla_c g_1(\delta,\, x,\, c) \\ \nabla_x g_2(\delta,\, x,\, c) & \nabla_c g_2(\delta,\, x,\, c) \end{bmatrix}\right|_{(0,\, x^\star(0),\, -\mathbb{E}_{\mathbb{P}}[f(x^\star(0),\, Y)])} \\ &= \begin{bmatrix} \mathbb{E}_{\mathbb{P}}[\nabla_x^2 f(x^\star(0),\, Y)] & 0 \\ 0 & 1 \end{bmatrix} \end{aligned} \tag{B.3}$$

is invertible. It follows from the Implicit Function Theorem that $(x^\star(\delta),\, c^\star(\delta))$ exists and is continuously differentiable in an open neighborhood of $\big(0,\, x^\star(0),\, -\mathbb{E}_{\mathbb{P}}[f(x^\star(0),\, Y)]\big)$ so we can write

$$\begin{aligned} x^\star(\delta) &= x^\star(0) + \delta\pi + o(\delta) \\ c^\star(\delta) &= -\mathbb{E}_{\mathbb{P}}[f(x^\star(0), Y)] + \delta\kappa + o(\delta) \end{aligned}$$

where $\pi \in \mathbb{R}^d$ and κ is a constant. Substituting into the first equation in (B.2), a Taylor series expansion around $\delta = 0$ gives

$$\begin{aligned}&\mathbb{E}_{\mathbb{P}}\big[\nabla_x f(x^\star(0), Y)\big] + \delta\pi'\mathbb{E}_{\mathbb{P}}\big[\nabla_x^2 f(x^\star(0), Y)\big]\\ &\quad -\frac{\delta}{\phi''(1)}\mathbb{E}_{\mathbb{P}}\Big[\Big(f(x^\star(0), Y) - \mathbb{E}_{\mathbb{P}}[f(x^\star(0), Y)\Big)\nabla_x f(x^\star(0), Y)\Big] + o(\delta) = 0.\end{aligned}$$

Optimality of $x^\star(0)$ for the population problem implies $\mathbb{E}_{\mathbb{P}}\big[\nabla_x f(x^\star(0), Y)\big] = 0$ while the coefficient of the δ term is zero if

$$\pi = \frac{1}{\phi''(1)}\Big(\mathbb{E}_{\mathbb{P}}\big[\nabla_x^2 f(x^\star(0), Y)\big]\Big)^{-1}\mathrm{Cov}_{\mathbb{P}}\Big(f(x^\star(0), Y), \nabla_x f(x^\star(0), Y)\Big).$$

Smoothness of $(x_n(\delta),\, c_n(\delta))$ and the expression (3.6)-(3.7) can be established in a similar manner.

Appendix C. Proof of Proposition 3.6

We begin with some preliminaries. Let $\psi(x, c, Y)$ be defined by (3.5), and

$$\nabla\psi(x^\star(\delta),\, c^\star(\delta), Y) = \left[\begin{array}{ccc|c} \nabla_{x_1}\psi_1 & \ldots & \nabla_{x_m}\psi_1 & \nabla_c\psi_1 \\ \hline \nabla_{x_1}\psi_2 & \ldots & \nabla_{x_m}\psi_2 & \nabla_c\psi_2 \end{array}\right](x^\star(\delta),\, c^\star(\delta), Y)$$

be the Jacobian of $\psi(x, c, Y)$ and the matrices

$$\begin{aligned}\widetilde{A}(\delta) &= \mathbb{E}_{\mathbb{P}}[\nabla\psi(x^\star(\delta),\, c^\star(\delta), Y)] \in \mathbb{R}^{(d+1)\times(d+1)},\\ \widetilde{B}(\delta) &= \mathbb{E}_{\mathbb{P}}[\psi(x^\star(\delta),\, c^\star(\delta), Y)\,\psi(x^\star(\delta),\, c^\star(\delta), Y)'] \in \mathbb{R}^{(d+1)\times(d+1)}.\end{aligned}$$

We also define the $\mathbb{R}^{d+1}$ valued random vectors

$$\begin{aligned}\widetilde{W}_n(\delta) &= -\widetilde{A}(\delta)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \psi(x^\star(\delta),\, c^\star(\delta),\, Y_i),\\ \widetilde{V}_n(\delta) &= \widetilde{H}_n(\delta)\widetilde{W}_n(\delta) - \widetilde{I}_n(\delta)\end{aligned} \tag{C.1}$$

where

$$\begin{aligned}\widetilde{H}_n(\delta) &= -\frac{1}{\sqrt{n}}\widetilde{A}(\delta)^{-1}\sum_{i=1}^n\Big\{\nabla\psi(x^\star(\delta),\, c^\star(\delta), Y_i) - \mathbb{E}_{\mathbb{P}}\big[\nabla\psi(x^\star(\delta),\, c^\star(\delta), Y)\big]\Big\}\\ \tilde{I}_n(\delta) &= \widetilde{A}(\delta)^{-1}\begin{bmatrix}\widetilde{W}_n(\delta)'\mathbb{E}_{\mathbb{P}}\big[\nabla^2\psi^{(1)}(x^\star(\delta),\, c^\star(\delta),\, Y)\big]\widetilde{W}_n(\delta)\\ \vdots \\ \widetilde{W}_n(\delta)'\mathbb{E}_{\mathbb{P}}\big[\nabla^2\psi^{(d+1)}(x^\star(\delta),\, c^\star(\delta),\, Y)\big]\widetilde{W}_n(\delta)\end{bmatrix}\end{aligned}$$

Here, $\psi^{(i)}(x,c,Y)$ is the i^{th} component of the $d+1$ dimensional function

$$\psi(x,c,Y) = \begin{bmatrix} \psi^{(1)}(x,c,Y) \\ \vdots \\ \psi^{(d+1)}(x,c,Y) \end{bmatrix}$$

that was defined in (3.5), and $\nabla^2\psi^{(i)}(x^\star(\delta),\, c^\star(\delta),\, Y)$ is the Hessian $\nabla^2\psi^{(i)}(x,\, c,\, Y)$ evaluated at $(x^\star(\delta), c^\star(\delta))$. Note that condition (2) of Assumption 3.5 implies that $\phi^*(\zeta)$ is three times continuously differentiable[3] in a neighborhood of $\zeta = 0$. Together with condition (1), it follows that the Hessian $\psi^{(i)}(x,\, c,\, Y)$ evaluated at $(x^\star(\delta), c^\star(\delta))$ above is well defined.

The following result characterizes the distribution of $(x_n(\delta),\, c_n(\delta))$ and is obtained by applying Theorem 5.21 in [27] and Lemma 1 from [17] to (3.5).

Proposition C.1. *Suppose that data $\{Y_1, \cdots, Y_n\}$ is drawn iid from $\mathbb{P}$, $f(x,\, Y)$ satisfies Assumption 2.1, $\phi(z)$ satisfies Assumption 3.1, and δ is such that Assumption 3.5 holds. Then there is a unique solution $(x^\star(\delta),\, c^\star(\delta))$ of the first-order conditions (3.4), the matrix $\widetilde{A}(\delta)$ is invertible, $(x_n(\delta),\, c_n(\delta)) \xrightarrow{P} (x^\star(\delta),\, c^\star(\delta))$, and for n sufficiently large*

$$\begin{bmatrix} x_n(\delta) \\ c_n(\delta) \end{bmatrix} = \begin{bmatrix} x^\star(\delta) \\ c^\star(\delta) \end{bmatrix} + \frac{1}{\sqrt{n}}\widetilde{W}_n(\delta) + \frac{1}{n}\widetilde{V}_n(\delta) + o_P(n^{-3/2}).$$

$\widetilde{W}_n(\delta)$ has mean 0 and covariance matrix

$$\widetilde{\xi}(\delta) \equiv \mathbb{V}_{\mathbb{P}}[\widetilde{W}_n(\delta)] = \widetilde{A}(\delta)^{-1}\widetilde{B}(\delta)\widetilde{A}(\delta)^{-1'},$$

and $\widetilde{V}_n(\delta)$ is $O_P(1)$ with mean

$$\begin{aligned} \mathbb{E}_{\mathbb{P}}[\widetilde{V}_n(\delta)] \;&=\; \mathbb{E}_{\mathbb{P}}\Big[\widetilde{A}(\delta)^{-1}\nabla\psi(x^\star(\delta),\, c^\star(\delta), Y)\widetilde{A}(\delta)^{-1}\psi(x^\star(\delta),\, c^\star(\delta), Y)\Big] \\ &\quad -\widetilde{A}(\delta)^{-1}\begin{bmatrix} \mathrm{tr}\big\{\widetilde{\xi}(\delta)\,\mathbb{E}_{\mathbb{P}}\big[\nabla^2\psi^{(1)}(x^\star(\delta),\, c^\star(\delta),\, Y)\big]\big\} \\ \vdots \\ \mathrm{tr}\big\{\widetilde{\xi}(\delta)\,\mathbb{E}_{\mathbb{P}}\big[\nabla^2\psi^{(d+1)}(x^\star(\delta),\, c^\star(\delta),\, Y)\big]\big\} \end{bmatrix}. \end{aligned}$$

Proposition 3.6 is largely a restatement of Proposition C.1 for the $x_n(\delta)$ component of $(x_n(\delta), c_n(\delta))$. In particular, expression (3.12) for $x_n(\delta)$ holds because we can extract the $\mathbb{R}^d$ valued random vectors

---

[3]It can be shown, along the lines of the Proof of Theorem 3.2 in [12], that

$$\phi^*(\zeta) = \zeta + \frac{\zeta^2}{2!}\Big(\frac{1}{\phi^{(2)}(1)}\Big) - \frac{\zeta^3}{3!}\Big(\frac{1}{\phi^{(3)}(1)}\frac{1}{\phi^{(2)}(1)}\Big) + o(\zeta^3)$$

where $\phi^{(i)}(z)$ is the i^{th} derivative of $\phi(z)$ with respect to z.

$W_n(\delta)$ and $V_n(\delta)$ of (C.1) corresponding to $x_n(\delta)$ in (C.2)

$$\widetilde{W}_n(\delta) = \begin{bmatrix} W_n(\delta) \\ U_n(\delta) \end{bmatrix} \in \mathbb{R}^{d+1}, \quad \widetilde{V}_n(\delta) = \begin{bmatrix} V_n(\delta) \\ T_n(\delta) \end{bmatrix} \in \mathbb{R}^{d+1},$$

and $\xi(\delta) \in \mathbb{R}^{d\times d}$ and $\overline{V}(\delta) \in \mathbb{R}^d$ from

$$\widetilde{\xi}(\delta) = \begin{bmatrix} \xi(\delta) & \kappa(\delta) \\ \kappa(\delta)' & \eta(\delta) \end{bmatrix}, \quad \mathbb{E}_{\mathbb{P}}[\widetilde{V}_n(\delta)] = \begin{bmatrix} \overline{V}(\delta) \\ \overline{S}(\delta) \end{bmatrix}.$$

$\xi(\delta)$ and $\overline{V}(\delta)$ are continuously differentiable in a neighborhood of $\delta = 0$ because the pair $(x^\star(\delta), c^\star(\delta))$ is continuously differentiable in a neighborhood of $\delta = 0$ (Proposition 3.2) and the conditions imposed on $\psi(x, c, Y)$ in Assumption 3.5.

Appendix D. Assumptions 4.1 and 4.2

We know from (3.10) that the solution of the SAA problem has an expansion of the form

$$x_n(0) = x^\star(0) + \frac{1}{\sqrt{n}} W_n(0) + \frac{1}{n} V_n(0) + H_n(0) \tag{D.1}$$

where the error term $H_n(0)$ is $O_p(n^{-\frac{3}{2}})$. Assumption 4.1 holds if $\sup_n \mathbb{E}_{\mathbb{P}}\|n^{\frac{3}{2}} H_n(0)\|^p < \infty$ for $p = 1, 2$. We now derive an expression for $H_n(0)$ to determine conditions under which this holds. To ease notation, we assume x is scalar. Our derivation shows that $n^{\frac{3}{2}} H_n(\delta)$ is a product of derivatives of $f(x, Y)$ with respect to x of up to order 4, and Assumption 4.1 holds if the first two moments of $n^{\frac{3}{2}} H_n(\delta)$ are finite.

Our approach generalizes to the case where the decision variable is multi-dimensional, which is required for DRO and DOO (Assumption 4.2) where the pair (x, c) is the decision and SAA with vector-valued x, at the cost of even uglier equations and tensor notation. Not surprisingly, the same insight, that products of (partial) derivatives of order up to 4 have finite first- and second moment, holds. The equations are long but not particularly insightful (beyond this observation) and the derivation is arduous. We outline how this can be done and leave details to the reader.

To derive the expansion (D.1), we embed the first order conditions for our sample problem in a family of fixed-point equations $G(y,\varepsilon)=0$ parameterized by $\varepsilon\in[0,1]$ where

$$\begin{aligned} G(y,\varepsilon) &= \mathbb{E}_{\mathbb{P}}[\nabla_x f(y,Y)] + \varepsilon\frac{1}{n}\sum_{i=1}^{n}\Big\{\nabla_x f(y,Y_i) - \mathbb{E}_{\mathbb{P}}[\nabla_x f(y,Y)]\Big\} \\ &= \mathbb{E}_{\mathbb{P}}[\nabla_x f(y,Y)] + \varepsilon\frac{1}{n}\sum_{i=1}^{n}\nabla_x \widetilde{f}(y,Y_i). \end{aligned}$$

Here

$$\widetilde{f}(y,Y) = f(y,Y) - \mathbb{E}_{\mathbb{P}}[f(y,Y)]$$

is the deviation of the random reward $f(y,Y)$ from its expectation $\mathbb{E}_{\mathbb{P}}[f(y,Y)]$, while $\nabla_x\widetilde{f}(y,Y)$ is the derivative of the centered random variable. We denote the solution of $G(y,\varepsilon)=0$ as $y(\varepsilon)$. (We change the argument in G from x to y to minimize the chance of confusion between $x_n(0)$, the solution of the SAA problem, and the solution $y(\varepsilon)$ of the fixed point problem). The case $\varepsilon=0$

$$G(y(0),0) = \mathbb{E}_{\mathbb{P}}[\nabla_x f(y(0),Y)] = 0$$

corresponds to the population problem with solution $y(0)=x^\star(0)$; $\varepsilon=1$ is the SAA problem with solution $y(1)=x_n(0)$

$$G(x_n(0),1) = \frac{1}{n}\sum_{i=1}^{n}\nabla_x f(x_n(0),Y_i) = 0.$$

To begin, we find the derivatives of $y(\varepsilon)$ with respect to ε, or $A(\varepsilon)$, $B(\varepsilon)$ and $C(\varepsilon)$, such that

$$y(\varepsilon+\Delta) = y(\varepsilon) + \Delta A(\varepsilon) + \frac{\Delta^2}{2}B(\varepsilon) + \frac{\Delta^3}{3!}C(\varepsilon) + o(\Delta^3).$$

Since $G(y(\varepsilon+\Delta),\varepsilon+\Delta)=0$ we do this by expanding $G(y(\varepsilon+\Delta),\varepsilon+\Delta)$ around $G(y(\varepsilon),\varepsilon)$ and choosing $A(\varepsilon)$, $B(\varepsilon)$ and $C(\varepsilon)$ to ensure the first-order condition holds.

Specifically, a Taylor series expansion around ε gives (after some work)

$$
\begin{aligned}
&G(y(\varepsilon+\Delta),\varepsilon+\Delta) = G(y(\varepsilon),\varepsilon) \\
&\quad +\Delta\left\{\Big[\mathbb{E}_{\mathbb{P}}[\nabla_x^2 f(y(\varepsilon),Y)] + \varepsilon\frac{1}{n}\sum_{i=1}^n \nabla_x^2 \widetilde{f}(y(\varepsilon),Y_i)\Big]A(\varepsilon) + \frac{1}{n}\sum_{i=1}^n \nabla_x f(y(\varepsilon),Y_i)\right\} \\
&\quad +\frac{\Delta^2}{2!}\left\{B(\varepsilon)\Big(\mathbb{E}_{\mathbb{P}}[\nabla_x^2 f(y(\varepsilon),Y)] + \varepsilon\frac{1}{n}\sum_{i=1}^n \nabla_x^2 \widetilde{f}(y(\varepsilon),Y_i)\Big)\right. \\
&\quad \left. +A(\varepsilon)^2\Big(\mathbb{E}_{\mathbb{P}}[\nabla_x^3 f(y(\varepsilon),Y)] + \varepsilon\frac{1}{n}\sum_{i=1}^n \nabla_x^3 \widetilde{f}(y(\varepsilon),Y_i)\Big) + A(\varepsilon)\frac{1}{n}\sum_{i=1}^n 2\nabla_x^2 \widetilde{f}(y(\varepsilon),Y)\right\} \\
&\quad +\frac{\Delta^3}{3!}\left\{C(\varepsilon)\Big(\mathbb{E}_{\mathbb{P}}[\nabla_x^2 f(y(\varepsilon),Y)] + \varepsilon\frac{1}{n}\sum_{i=1}^n \nabla_x^2 \widetilde{f}(y(\varepsilon),Y_i)\Big)\right. \\
&\quad +A(\varepsilon)^3\Big(\mathbb{E}_{\mathbb{P}}[\nabla_x^4 f(y(\varepsilon),Y)] + \varepsilon\frac{1}{n}\sum_{i=1}^n \nabla_x^4 \widetilde{f}(y(\varepsilon),Y_i)\Big) \\
&\quad +3A(\varepsilon)B(\varepsilon)\Big(\mathbb{E}_{\mathbb{P}}[\nabla_x^3 f(y(\varepsilon),Y)] + \varepsilon\frac{1}{n}\sum_{i=1}^n \nabla_x^3 \widetilde{f}(y(\varepsilon),Y_i)\Big) \\
&\quad \left. +3\Big(\frac{1}{n}\sum_{i=1}^n \nabla_x^2 \widetilde{f}(y(\varepsilon),Y_i)\Big)B(\varepsilon) + 3\Big(\frac{1}{n}\sum_{i=1}^n \nabla_x^3 \widetilde{f}(y(\varepsilon),Y_i)\Big)A(\varepsilon)^2\right\} + o(\Delta^3).
\end{aligned}
$$

Since $G(y(\varepsilon+\Delta),\varepsilon+\Delta)=0$, the coefficients of the powers of Δ must be zero and hence

$$
A(\varepsilon) = \frac{1}{\sqrt{n}}\Lambda_n(\varepsilon), \qquad B(\varepsilon) = \frac{1}{n}\Gamma_n(\varepsilon), \qquad C(\varepsilon) = n^{-\frac{3}{2}}\Theta_n(\varepsilon)
$$

where

$$
\begin{aligned}
\Lambda_n(\varepsilon) &= \frac{-\frac{1}{\sqrt{n}}\sum_{i=1}^n \nabla_x f(y(\varepsilon),Y_i)}{\mathbb{E}_{\mathbb{P}}[\nabla_x^2 f(y(\varepsilon),Y)] + \varepsilon\frac{1}{n}\sum_{i=1}^n \nabla_x^2 \widetilde{f}(y(\varepsilon),Y_i)} \\
\Gamma_n(\varepsilon) &= -[\Lambda_n(\varepsilon)]^2 \left\{ \frac{\mathbb{E}_{\mathbb{P}}[\nabla_x^3 f(y(\varepsilon),Y)] + \varepsilon\frac{1}{n}\sum_{i=1}^n \nabla_x^3 \widetilde{f}(y(\varepsilon),Y_i)}{\mathbb{E}_{\mathbb{P}}[\nabla_x^2 f(y(\varepsilon),Y)] + \varepsilon\frac{1}{n}\sum_{i=1}^n \nabla_x^2 \widetilde{f}(y(\varepsilon),Y_i)} \right\} \\
&\quad -\Lambda_n(\varepsilon) \left\{ \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n 2\nabla_x^2 \widetilde{f}(y(\varepsilon),Y)}{\mathbb{E}_{\mathbb{P}}[\nabla_x^2 f(y(\varepsilon),Y)] + \varepsilon\frac{1}{n}\sum_{i=1}^n \nabla_x^2 \widetilde{f}(y(\varepsilon),Y_i)} \right\} \\
\Theta_n(\varepsilon) &= -3\Gamma_n(\varepsilon) \left\{ \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n \nabla_x^2 \widetilde{f}(y(\varepsilon),Y_i)}{\mathbb{E}_{\mathbb{P}}[\nabla_x^2 f(y(\varepsilon),Y)] + \varepsilon\frac{1}{n}\sum_{i=1}^n \nabla_x^2 \widetilde{f}(y(\varepsilon),Y_i)} \right\} \\
&\quad -3[\Lambda_n(\varepsilon)]^2 \left\{ \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n \nabla_x^3 \widetilde{f}(y(\varepsilon),Y_i)}{\mathbb{E}_{\mathbb{P}}[\nabla_x^2 f(y(\varepsilon),Y)] + \varepsilon\frac{1}{n}\sum_{i=1}^n \nabla_x^2 \widetilde{f}(y(\varepsilon),Y_i)} \right\} \\
&\quad -[\Lambda_n(\varepsilon)]^3 \left\{ \frac{\mathbb{E}_{\mathbb{P}}[\nabla_x^4 f(y(\varepsilon),Y)] + \varepsilon\frac{1}{n}\sum_{i=1}^n \nabla_x^4 \widetilde{f}(y(\varepsilon),Y_i)}{\mathbb{E}_{\mathbb{P}}[\nabla_x^2 f(y(\varepsilon),Y)] + \varepsilon\frac{1}{n}\sum_{i=1}^n \nabla_x^2 \widetilde{f}(y(\varepsilon),Y_i)} \right\} \\
&\quad -3\Lambda_n(\varepsilon)\Gamma_n(\varepsilon) \left\{ \frac{\mathbb{E}_{\mathbb{P}}[\nabla_x^3 f(y(\varepsilon),Y)] + \varepsilon\frac{1}{n}\sum_{i=1}^n \nabla_x^3 \widetilde{f}(y(\varepsilon),Y_i)}{\mathbb{E}_{\mathbb{P}}[\nabla_x^2 f(y(\varepsilon),Y)] + \varepsilon\frac{1}{n}\sum_{i=1}^n \nabla_x^2 \widetilde{f}(y(\varepsilon),Y_i)} \right\}
\end{aligned}
$$

which gives us the derivatives of $y(\varepsilon)$ we are looking for. This gives the Taylor series expansion

$$y(\varepsilon) = y(0) + \varepsilon A(0) + \frac{\varepsilon^2}{2}B(0) + \frac{\varepsilon^3}{3!}C(\eta)$$

where η is a random variable taking values in $[0,1]$ that depends on the data. In particular, with $\varepsilon = 1$ and noting that $y(\varepsilon) = x_n(0)$ when $\varepsilon = 1$

$$x_n(0) = x^\star(0) + \frac{1}{\sqrt{n}}\Lambda_n(0) + \frac{1}{n}\Gamma_n(0) + n^{-\frac{3}{2}}\Theta_n(\eta)$$

where

$$
\begin{aligned}
\Lambda_n(0) &= \frac{-\frac{1}{\sqrt{n}}\sum_{i=1}^n \nabla_x f(x^\star(0),Y_i)}{\mathbb{E}_{\mathbb{P}}[\nabla_x^2 f(x^\star(0),Y)]} \\
\Gamma_n(0) &= -[\Lambda_n(0)]^2 \left\{ \frac{\mathbb{E}_{\mathbb{P}}[\nabla_x^3 f(x^\star(0),Y)]}{\mathbb{E}_{\mathbb{P}}[\nabla_x^2 f(x^\star(0),Y)]} \right\} - \Lambda_n(0) \left\{ \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n 2\nabla_x^2 \widetilde{f}(x^\star(0),Y)}{\mathbb{E}_{\mathbb{P}}[\nabla_x^2 f(x^\star(0),Y)]} \right\},
\end{aligned}
$$

and η is a random variable taking values between 0 and 1. Observe that $\Lambda_n(0) = W_n(0)$ and $\Gamma_n(0) = V_n(0)$, as defined in (3.10) and that the error term in (D.1) is $H_n(0) = n^{-\frac{3}{2}}\Theta_n(\eta)$. Assumption 4.1 holds if $\mathbb{E}_{\mathbb{P}}\|\Theta_n(\eta)\|^p < \infty$ for $p = 1, 2$. This will be the case if there is a constant $\gamma > 0$ such that

$\nabla_x^2 f(x,Y) > \gamma$ for every (x,Y) and the derivatives $\|\nabla_x^k f(x,Y)\|$ of order $k = 1,2,3,4$ are uniformly bounded, though this is not the mildest assumption.

In the case of DRO/DOO, we have first-order conditions

$$\mathbb{E}_{\mathbb{P}}\left[\begin{array}{c}[\phi^*]'\Big(-\delta\big[f(x,Y)+c\big]\Big)\nabla_x f(x,Y) = 0\\ -\frac{\phi''(1)}{\delta}\Big\{[\phi^*]'\Big(-\delta\big[f(x,Y)+c\big]\Big)-1\Big\}\end{array}\right] = 0 \tag{D.2}$$

for the solution $(x^\star(\delta), c^\star(\delta))$ of the population problem, and

$$\frac{1}{n}\sum_{i=1}^{n}\left[\begin{array}{c}[\phi^*]'\Big(-\delta\big[f(x,Y)+c\big]\Big)\nabla_x f(x,Y)\\ -\frac{\phi''(1)}{\delta}\Big\{[\phi^*]'\Big(-\delta\big[f(x,Y)+c\big]\Big)-1\Big\}\end{array}\right] = 0 \tag{D.3}$$

for the solution $(x_n^\star(\delta), c_n^\star(\delta))$ of the empirical version. Analogous to the case of SAA, we can define $(y(\varepsilon),\omega(\varepsilon))$ as the solution of $G((y(\varepsilon),\omega(\varepsilon)),\varepsilon) = [0,0]'$ ($\varepsilon \in [0,1]$) where

$$\begin{aligned} G((y(\varepsilon),\omega(\varepsilon)),\varepsilon) &= (1-\varepsilon)\mathbb{E}_{\mathbb{P}}\left[\begin{array}{c}[\phi^*]'\Big(-\delta\big[f(y(\varepsilon),Y)+\omega(\varepsilon)\big]\Big)\nabla_x f(y(\varepsilon),Y)\\ -\frac{\phi''(1)}{\delta}\Big\{[\phi^*]'\Big(-\delta\big[f(y(\varepsilon),Y)+c(\varepsilon)\big]\Big)-1\Big\}\end{array}\right]\\ &\quad +\varepsilon\frac{1}{n}\sum_{i=1}^{n}\left[\begin{array}{c}[\phi^*]'\Big(-\delta\big[f(y(\varepsilon),Y)+\omega\big]\Big)\nabla_x f(y(\varepsilon),Y)\\ -\frac{\phi''(1)}{\delta}\Big\{[\phi^*]'\Big(-\delta\big[f(y(\varepsilon),Y)+\omega(\varepsilon)\big]\Big)\Big\}\end{array}\right]. \end{aligned}$$

This allows us to embed (D.2) and (D.3) in the problem of solving $G((y(\varepsilon),\omega(\varepsilon)),\varepsilon) = [0,0]'$. Expressions for $\Lambda_n(\varepsilon)$, $\Gamma_n(\varepsilon)$ and the remainder term $\Theta_n(\varepsilon)$ such that

$$\left[\begin{array}{c} y(\varepsilon)\\ \omega(\varepsilon)\end{array}\right] = \frac{1}{\sqrt{n}}\Lambda_n(\varepsilon) + \frac{1}{n}\Gamma_n(\varepsilon) + n^{-\frac{3}{2}}\Theta_n(\eta),$$

where η is a random variable between 0 and ε, can be derived, though equations are long and unwieldy. Setting $\varepsilon = 1$ we get

$$\left[\begin{array}{c} x_n(\delta)\\ c_n(\delta)\end{array}\right] = \frac{1}{\sqrt{n}}W_n(\delta) + \frac{1}{n}V_n(\delta) + n^{-\frac{3}{2}}\Theta_n(\eta),$$

where $W_n(\delta) = \Lambda_n(1)$, $V_n(\delta) = \Gamma_n(1)$ and the remainder term $\Theta_n(\eta)$ is a product of partial derivatives of

$$\Phi(y,\omega,Y) := -\frac{1}{\delta}\phi^*\Big(-\delta\Big[f(y,Y)+\omega\Big]\Big) - \omega$$

with respect to y and ω of up to order 4, evaluated at $y(\eta)$ for some random $\eta \in [0,1]$ that depends on the data. Assumption 4.2 holds if $\sup_n \mathbb{E}_{\mathbb{P}}\|\Theta_n(\eta)\|^p < \infty$ for $p = 1, 2$. Again, this will be the case if, for example, the partial derivatives up to order 4 are uniformly bounded.

As a final note, [22] provides conditions for the error term in an M-estimation problem to converge in L^p for every $p \geq 1$. It can also be used to derive conditions for $\sup_n \mathbb{E}_{\mathbb{P}}\|n^{\frac{3}{2}} H_n\|^p < \infty$ for every $p \geq 1$, and hence for Assumptions 4.1 and 4.2 to hold.

Appendix E. Proof of Proposition 4.3

Recall from (4.4) that

$$\mathbb{E}_{\mathbb{P}}[x_n(0)] = x^\star(0) + \frac{1}{n}\overline{V}(0) + o(n^{-1}) \tag{E.1}$$

and hence

$$\begin{aligned}
\nabla_x g(\mathbb{E}_{\mathbb{P}}[x_n(0)]) &= \nabla_x g\Big(x^\star(0) + \frac{1}{n}\overline{V}(0) + o(n^{-1})\Big) \\
&= \nabla_x g(x^\star(0)) + \nabla_x^2 g(x^\star(0))\Big(\frac{1}{n}\overline{V}(0)\Big) + o(n^{-1}) \\
&= \frac{1}{n}\nabla_x^2 g(x^\star(0))\overline{V}(0) + o(n^{-1}) \\
\nabla_x^2 g(\mathbb{E}_{\mathbb{P}}[x_n(0)]) &= \nabla_x^2 g(x^\star(0)) + O(n^{-1}).
\end{aligned} \tag{E.2}$$

From Proposition 3.6, the DRO/DOO solution

$$x_n(\delta) = x^\star(\delta) + \frac{1}{\sqrt{n}}W_n(\delta) + \frac{1}{n}V_n(\delta) + o_P(n^{-1})$$

where $x^\star(\delta) = x^\star(0) + \delta\pi + o(\delta)$. Taking expectations gives

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}}[x_n(\delta)] &= x^\star(\delta) + \frac{1}{n}\overline{V}(\delta) + o(n^{-1}) \\
\mathbb{V}_{\mathbb{P}}[x_n(\delta)] &= \frac{1}{n}\xi(\delta) + o(n^{-1}).
\end{aligned} \tag{E.3}$$

Using (E.1) we can write (E.3) as

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}}[x_n(\delta)] &= \mathbb{E}_{\mathbb{P}}[x_n(0)] + [x^\star(\delta) - x^\star(0)] + \frac{1}{n}[\overline{V}(\delta) - \overline{V}(0)] + o(n^{-1}) \\
&= \mathbb{E}_{\mathbb{P}}[x_n(0)] + \delta\pi + o(\delta) + \frac{1}{n}\{\delta\overline{V}_\delta(0) + o(\delta)\} + o(n^{-1}).
\end{aligned} \tag{E.4}$$

We now expand $g(\mathbb{E}_{\mathbb{P}}[x_n(\delta)])$ around $g(\mathbb{E}_{\mathbb{P}}[x_n(0)])$:

$$\begin{aligned}
&g(\mathbb{E}_{\mathbb{P}}[x_n(\delta)])\\
&= g(\mathbb{E}_{\mathbb{P}}[x_n(0)]) + \Big(\nabla_x g\big(\mathbb{E}_{\mathbb{P}}[x_n(0)]\big)\Big)'\Big\{\mathbb{E}_{\mathbb{P}}[x_n(\delta)] - \mathbb{E}_{\mathbb{P}}[x_n(0)]\Big\}\\
&\quad +\frac{1}{2}\Big\{\mathbb{E}_{\mathbb{P}}[x_n(\delta)] - \mathbb{E}_{\mathbb{P}}[x_n(0)]\Big\}'\nabla_x^2 g(\mathbb{E}_{\mathbb{P}}[x_n(0)])\Big\{\mathbb{E}_{\mathbb{P}}[x_n(\delta)] - \mathbb{E}_{\mathbb{P}}[x_n(0)]\Big\}\\
&\quad +o\big(\|x_n(\delta)] - \mathbb{E}_{\mathbb{P}}[x_n(0)]\|^2\big)\\
&= g(\mathbb{E}_{\mathbb{P}}[x_n(0)]) + \Big(\frac{1}{n}\,\nabla_x^2 g(x^\star(0))\overline{V}(0) + o(n^{-1})\Big)'\Big\{\delta\pi + O(\delta^2) + \frac{1}{n}O(\delta)\Big\}\\
&\quad +\frac{1}{2}\Big(\delta\pi + O(\delta^2) + \frac{1}{n}O(\delta)\Big)'\Big\{\nabla_x^2 g(x^\star(0)) + O(n^{-1})\Big\}\Big(\delta\pi + O(\delta^2) + \frac{1}{n}O(\delta)\Big)
\end{aligned}$$

where the second equality follows from (E.2) and (E.4). It now follows that the expected reward at the mean of the robust solution

$$\begin{aligned}
&g(\mathbb{E}_{\mathbb{P}}[x_n(\delta)])\\
&= g(\mathbb{E}_{\mathbb{P}}[x_n(0)]) + \underbrace{\frac{\delta}{n}\Big\{\pi'\nabla_x^2 g(x^\star(0))\overline{V}(0)\Big\} + \frac{\delta^2}{2}\pi'\big[\nabla_x^2 g(x^\star(0))\big]\pi}_{\text{Impact of robustness}}\\
&\quad +o(1/n^2) + o(\delta^2) + \frac{1}{n}o(\delta^2) + \frac{1}{n^2}O(\delta)
\end{aligned} \tag{E.5}$$

In the case of the loss from Jensen's inequality, recall that the variance of the solution

$$\begin{aligned}
&\mathbb{E}_{\mathbb{P}}\Big[\big(x_n(\delta)) - \mathbb{E}_{\mathbb{P}}[x_n(\delta)]\big)\big(x_n(\delta)) - \mathbb{E}_{\mathbb{P}}[x_n(\delta)]\big)'\Big]\\
&= \frac{1}{n}\xi(\delta) + O(n^{-3/2})\\
&= \frac{1}{n}\Big\{\xi(0) + \delta\xi_\delta(0)\Big\} + \frac{1}{n}o(\delta) + O(n^{-3/2}).
\end{aligned}$$

It now follows from a Taylor series expansion that

$$\begin{aligned}
&\mathbb{E}_{\mathbb{P}}\Big[g(x_n(\delta)) - g(\mathbb{E}_{\mathbb{P}}[x_n(\delta)])\Big] \\
&= \mathbb{E}_{\mathbb{P}}\Big[\nabla_x g(\mathbb{E}_{\mathbb{P}}[x_n(\delta)])\Big(x_n(\delta) - \mathbb{E}_{\mathbb{P}}[x_n(\delta)]\Big)\Big] \\
&\quad + \frac{1}{2}\mathbb{E}_{\mathbb{P}}\Big[\big(x_n(\delta)) - \mathbb{E}_{\mathbb{P}}[x_n(\delta)]\big)'\nabla_x^2 g(\mathbb{E}_{\mathbb{P}}[x_n(\delta)])\big(x_n(\delta)) - \mathbb{E}_{\mathbb{P}}[x_n(\delta)]\big)\Big] + O(n^{-3/2}) \\
&= \frac{1}{2}\mathrm{tr}\Big\{\nabla_x^2 g\Big(x^\star(\delta) + \frac{1}{n}\overline{V}(\delta) + o(n^{-1})\Big)\frac{\xi(\delta)}{n}\Big\} + O(n^{-3/2}) \\
&= \frac{1}{2n}\mathrm{tr}\Big\{\nabla_x^2 g(x^\star(\delta))\xi(\delta)\Big\} + O(n^{-3/2}) \\
&= \frac{1}{2n}\mathrm{tr}\Big\{\nabla_x^2 g(x^\star(0))\xi(0)\Big\} + \frac{\delta}{2n}\frac{\mathrm{d}}{\mathrm{d}\delta}\Big\{\mathrm{tr}\Big(\nabla_x^2 g(x^\star(\delta))\xi(\delta)\Big)\Big\}\Big|_{\delta=0} + \frac{1}{n}O(\delta^2) + O(n^{-3/2}) \\
&= \mathbb{E}_{\mathbb{P}}\Big[g(x_n(0)) - g(\mathbb{E}_{\mathbb{P}}[x_n(0)])\Big] + \underbrace{\Big(\frac{\delta}{2n}\Big)\frac{\mathrm{d}}{\mathrm{d}\delta}\Big\{\mathrm{tr}\Big(\nabla_x^2 g(x^\star(\delta))\xi(\delta)\Big)\Big\}\Big|_{\delta=0}}_{\text{Impact of robustness}} \qquad \text{(E.6)} \\
&\quad + \frac{1}{n}O(\delta^2) + O(n^{-3/2})
\end{aligned}$$

where the second equality follows from (E.3), the fourth is a Taylor series expansion of the function $\mathrm{tr}\Big\{\nabla_x^2 g(x^\star(\delta))\xi(\delta)\Big\}$ around $\delta = 0$, and the last equality follows from (4.7). Proposition 4.3 follows after substituting (E.5) and (E.6) into the decomposition (4.9).