

Error Geometry and Causal Robustness: A Unified Geometric Framework from Parameter Confidence Ellipsoids to Multi-Experiment Trust Regions

Haobo Ma¹

Wenlin Zhang²

¹Independent Researcher

²National University of Singapore

November 24, 2025

Abstract

This paper proposes a unified framework for transforming statistical errors into “geometric boundaries” and systematically embedding it into causal inference and experimental design. The core idea is: given an estimator and its error in parameter space, we no longer treat “confidence intervals/standard errors” as auxiliary information, but elevate them to “trust regions” in parameter space—geometric objects with metric structure; all causal conclusions, robustness judgments, and experimental planning are characterized through inclusion, intersection, union, and linear images among these geometric regions. Specifically, this paper first constructs confidence ellipsoids using typical asymptotic normality and information matrices under general parametric models, endowing parameter space with local Riemannian metric structure; second, in causal inference, we unify “identifiable sets” and “trust regions” as two types of sets in the same parameter space, characterizing provable causal conclusions and admissible extrapolation directions through their intersections; further, in multi-experiment/multi-model scenarios, we construct consensus regions and conflict regions via intersections, unions, and mappings of trust regions, forming a kind of “geometrized meta-analysis”; finally, in experimental design and observation planning, we formalize the objective of “shrinking trust region volume/semi-axes” as an optimization problem over design variables, providing several solvable instances under linear models and instrumental variable models. The appendix provides rigorous proofs of main theorems, including coverage properties of confidence ellipsoids, robustness criteria for causal effects, and equivalence relations between design criteria and Fisher information.

Keywords: Error Geometry; Confidence Ellipsoid; Causal Inference; Identifiable Set; Experimental Design; Fisher Information

MSC 2020: 62F12, 62K05, 62P25, 62R01

Contents

1 Introduction

Statistical inference is traditionally presented in the form of “point estimate + confidence interval”, while causal inference follows the basic structure of “identification assumptions + point estimate + sensitivity analysis”. In practice of decision-making and engineering applications, researchers often face three types of problems:

1. Given finite samples, **which causal conclusions are truly supported by data, rather than mere illusions of point estimates?**
2. When aggregating different experiments, different models, or even different data sources, **how should we systematically see consensus and conflicts among various results?**
3. Under limited resources, **how can we maximize “geometric resolution” for specific causal effects or parameter directions through experimental design?**

These problems differ in form but share a common feature: they all relate to “error”, and the essence of “error” is not just a variance or a confidence interval, but a geometric object with shape, direction, and boundary. The goal of this paper is to thoroughly formalize this intuitive “geometric nature”.

We adopt the following viewpoint:

- For any estimate $\hat{\theta}$ of parameter $\theta \in \Theta \subset \mathbb{R}^d$, its error naturally induces a region $\mathcal{R} \subset \Theta$ with metric structure, which can be called a “trust region”;
- Causal conclusions are not statements about single point $\hat{\theta}$, but about the range of some function $\psi(\theta)$ over $\theta \in \mathcal{R} \cap \mathcal{I}$, where \mathcal{I} is the identifiable set;
- Results from multiple experiments, models, and different assumptions can be unified as multiple trust regions \mathcal{R}_k in the same parameter space or its projection, whose intersections, unions, and symmetric differences naturally characterize parameter ranges that are “robustly consistent”, “contestable”, or “significantly conflicting”;
- Experimental design and observation planning can be viewed as an optimization problem of “actively shaping the geometric shape of future trust regions”, aiming to shrink semi-axes of trust regions in specific directions (such as some causal effect) or reduce their volume.

In this framework, “error geometry” is no longer just an appendage of results, but becomes the core structure of the entire causality-decision process. This paper will start from the most basic parametric models, construct this geometric framework, and provide provable properties and computable implementations under several specific models.

2 Parametric Models and Geometric Structure of Trust Regions

2.1 Parametric Models and Estimators

Let observed data be X_1, \dots, X_n , defined on sample space \mathcal{X} , assuming their distribution belongs to a family of probability measures $\{P_\theta : \theta \in \Theta \subset \mathbb{R}^d\}$. Let $\theta_0 \in \Theta$ be the “true

parameter”, $\hat{\theta}_n = \hat{\theta}_n(X_1, \dots, X_n)$ some estimator.

We assume there exists the usual asymptotic linearity and normality structure:

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}(0, I(\theta_0)^{-1}),$$

where $I(\theta_0)$ is the Fisher information matrix, positive definite and continuous. Further assume there exists consistent estimate \hat{I}_n such that $\hat{I}_n \xrightarrow{P} I(\theta_0)$.

2.2 Local Metric Induced by Fisher Information

At each point θ of Θ , define bilinear form

$$g_\theta(u, v) := u^\top I(\theta)v, \quad u, v \in \mathbb{R}^d,$$

then g is a local Riemannian metric on Θ (under differentiability conditions). Intuitively, eigendirections of $I(\theta)$ describe “easy/difficult to distinguish” parameter directions: the smaller the variance along some direction, the larger the “unit length” in that direction under information metric, and vice versa.

Under finite sample n , using \hat{I}_n we obtain empirical metric

$$\hat{g}_n(u, v) := u^\top \hat{I}_n v,$$

which converges to g_{θ_0} in probability sense.

2.3 Confidence Ellipsoid as Trust Region

For given significance level $\alpha \in (0, 1)$, define the quantile $\chi_{d,1-\alpha}^2$ of d -dimensional chi-square distribution, construct trust region

$$\mathcal{R}_n(\alpha) := \{\theta \in \Theta : n(\theta - \hat{\theta}_n)^\top \hat{I}_n(\theta - \hat{\theta}_n) \leq \chi_{d,1-\alpha}^2\}.$$

It is an ellipsoid centered at $\hat{\theta}_n$, with shape determined by \hat{I}_n^{-1} , characterizing parameter uncertainty. Classical theory guarantees:

Theorem 2.1 (Asymptotic Coverage). *Under the above regularity conditions, for any fixed $\alpha \in (0, 1)$,*

$$P_{\theta_0}(\theta_0 \in \mathcal{R}_n(\alpha)) \rightarrow 1 - \alpha, \quad n \rightarrow \infty.$$

This theorem is proved in Appendix A.1.

Therefore, $\mathcal{R}_n(\alpha)$ can be viewed as a “trust region containing true value θ_0 with probability $1 - \alpha$ ”. Under information metric g_{θ_0} , semi-axis lengths of $\mathcal{R}_n(\alpha)$ are inversely proportional to eigenvalues of Fisher information and proportional to $1/\sqrt{n}$.

3 Error as Geometric Boundary: Operations and Projections of Trust Regions

This section systematically transforms “error” into “geometric boundary” and discusses several basic operations: projection, linear image, and nonlinear image.

3.1 Linear Function Image and Ellipsoid Projection

Let the target of interest be linear function $\psi(\theta) = c^\top \theta$, where $c \in \mathbb{R}^d$. On ellipsoid $\mathcal{R}_n(\alpha)$, the range of ψ is

$$\Psi_n(\alpha) := \{\psi(\theta) : \theta \in \mathcal{R}_n(\alpha)\} = [\psi_{\min,n}, \psi_{\max,n}].$$

This interval can be analytically calculated. Note that

$$\psi(\theta) = c^\top \theta = c^\top \hat{\theta}_n + c^\top (\theta - \hat{\theta}_n),$$

with constraint $n(\theta - \hat{\theta}_n)^\top \hat{I}_n(\theta - \hat{\theta}_n) \leq \chi_{d,1-\alpha}^2$. Let $h = \theta - \hat{\theta}_n$, the problem becomes maximizing/minimizing linear function $c^\top h$ under ellipsoid constraint. Classical optimization conclusion gives

$$\begin{aligned}\psi_{\max,n} &= c^\top \hat{\theta}_n + \sqrt{\frac{\chi_{d,1-\alpha}^2}{n} c^\top \hat{I}_n^{-1} c}, \\ \psi_{\min,n} &= c^\top \hat{\theta}_n - \sqrt{\frac{\chi_{d,1-\alpha}^2}{n} c^\top \hat{I}_n^{-1} c}.\end{aligned}$$

Therefore, confidence interval for linear target is naturally given by geometric relationship between ellipsoid and direction vector c .

Proposition 3.1 (Optimal Bounds for Linear Target). *For any $c \in \mathbb{R}^d$, the minimum and maximum values of $\psi(\theta) = c^\top \theta$ on $\mathcal{R}_n(\alpha)$ are as shown above, and this interval has coverage probability $1 - \alpha$ in asymptotic sense.*

Proof is in Appendix A.2.

3.2 Local Linear Approximation of Nonlinear Functions

If $\psi : \Theta \rightarrow \mathbb{R}^k$ is differentiable, then near $\hat{\theta}_n$ we can make first-order approximation

$$\psi(\theta) \approx \psi(\hat{\theta}_n) + D\psi(\hat{\theta}_n)(\theta - \hat{\theta}_n),$$

where Jacobian matrix $D\psi(\hat{\theta}_n) \in \mathbb{R}^{k \times d}$ has i -th row as $\nabla \psi_i(\hat{\theta}_n)^\top$. Then $\psi(\mathcal{R}_n(\alpha))$ under first-order approximation is an ellipsoid in \mathbb{R}^k :

$$\mathcal{S}_n(\alpha) := \left\{ y \in \mathbb{R}^k : n(y - \psi(\hat{\theta}_n))^\top (D\psi(\hat{\theta}_n) \hat{I}_n^{-1} D\psi(\hat{\theta}_n)^\top)^{-1} (y - \psi(\hat{\theta}_n)) \leq \chi_{k,1-\alpha}^2 \right\},$$

where existence of inverse matrix requires row vectors of $D\psi(\hat{\theta}_n)$ to be linearly independent under information metric. This result is essentially a restatement of Delta method in geometric language.

3.3 Support Function Form for Multi-Parameter and Multi-Target

In more general cases, we can use support functions to characterize arbitrary convex target sets. On convex ellipsoid $\mathcal{R}_n(\alpha)$, its support function is

$$h_{\mathcal{R}_n(\alpha)}(u) := \sup_{\theta \in \mathcal{R}_n(\alpha)} u^\top \theta = u^\top \hat{\theta}_n + \sqrt{\frac{\chi^2_{d,1-\alpha}}{n} u^\top \hat{I}_n^{-1} u}.$$

Therefore, any target set defined by collection of linear functionals $\{u : u \in \mathcal{U}\}$ can have its boundary directly calculated through support function. An important application in causal inference scenarios is: when we care about a family of linear causal effects (such as heterogeneous effects across multiple groups), we can uniformly provide their worst/best cases on trust regions through support functions.

4 Identifiable Sets and Trust Regions in Causal Inference

4.1 Causal Models and Identifiable Sets

In causal inference, parameter θ usually has structural interpretation, such as average treatment effect in potential outcome models, path coefficients in structural equation models, local average treatment effect in instrumental variable models, etc. In cases of non-identification or partial identification, what data and assumptions can determine is only an **identifiable set**

$$\mathcal{I} := \{\theta \in \Theta : \theta \text{ is compatible with observed distribution and causal assumptions}\}.$$

For example, when violating certain exclusion restrictions or with selection bias, identifiable sets are often convex sets, semi-algebraic sets, or general closed sets, rather than single points.

4.2 Data-Driven Estimation of Identifiable Sets

Under finite samples, we typically approximate \mathcal{I} through estimated inequalities. For example, if causal constraints can be expressed as parameter constraints

$$g_j(\theta) \leq 0, \quad j = 1, \dots, m,$$

while we can only estimate empirical version $\hat{g}_{j,n}(\theta)$ of $g_j(\theta)$, common practice is using “relaxed inequalities”

$$\hat{g}_{j,n}(\theta) \leq b_{j,n},$$

where $b_{j,n}$ is upper bound (such as threshold after multiple testing correction), thus obtaining data-driven identifiable set estimate

$$\hat{\mathcal{I}}_n := \{\theta \in \Theta : \hat{g}_{j,n}(\theta) \leq b_{j,n}, j = 1, \dots, m\}.$$

Under suitable conditions it can be proved that $\hat{\mathcal{I}}_n$ converges to \mathcal{I} in appropriate sense.

4.3 Intersection of Identifiable Set and Trust Region

This paper proposes: **Causal conclusions should be based on $\mathcal{R}_n(\alpha) \cap \hat{\mathcal{I}}_n$ rather than merely on $\hat{\theta}_n$.** For given causal function $\psi : \Theta \rightarrow \mathbb{R}^k$, we care about

$$\mathcal{C}_n(\alpha) := \{\psi(\theta) : \theta \in \mathcal{R}_n(\alpha) \cap \hat{\mathcal{I}}_n\}.$$

If $\mathcal{C}_n(\alpha)$ has consistent sign in some direction or component, or is restricted to some desired interval, we can say this causal conclusion is “geometrically robust” at significance level α .

Definition 4.1 (Geometric Robustness). Let $\psi : \Theta \rightarrow \mathbb{R}$ be a scalar causal target, $\mathcal{R}_n(\alpha)$ a $1 - \alpha$ level trust region, $\hat{\mathcal{I}}_n$ sample approximation of identifiable set. If there exists interval $[L, U] \subset \mathbb{R}$ such that

$$\mathcal{C}_n(\alpha) = \{\psi(\theta) : \theta \in \mathcal{R}_n(\alpha) \cap \hat{\mathcal{I}}_n\} \subset [L, U],$$

then say “at level α , causal conclusion $\psi(\theta) \in [L, U]$ is geometrically robust”.

In particular, when $L > 0$ (or $U < 0$), robust judgment can be made about effect direction.

4.4 A Typical Criterion: Linear Causal Effect

Let $\psi(\theta) = c^\top \theta$ be linear causal effect (such as some linear combination in multi-parameter model corresponding to average treatment effect), and identifiable set can be represented as linear inequalities

$$A\theta \leq b,$$

then $\mathcal{R}_n(\alpha) \cap \hat{\mathcal{I}}_n$ is intersection of ellipsoid and polyhedron, a convex set. Extreme values of causal effect can be given by following convex optimization problem:

$$\psi_{\max,n}^* := \sup\{c^\top \theta : \theta \in \mathcal{R}_n(\alpha), A\theta \leq b\},$$

$$\psi_{\min,n}^* := \inf\{c^\top \theta : \theta \in \mathcal{R}_n(\alpha), A\theta \leq b\}.$$

In many applications, this problem can be efficiently solved through quadratic programming or semidefinite programming. Clearly,

$$\mathcal{C}_n(\alpha) = [\psi_{\min,n}^*, \psi_{\max,n}^*].$$

Theorem 4.2 (Geometric Robustness Criterion for Linear Causal Effect). *Under above setting, if for some $\delta > 0$,*

$$\psi_{\min,n}^* \geq \delta > 0,$$

then at significance level α , can conclude causal conclusion “effect is positive and at least δ ”, and this conclusion holds for all $\theta \in \mathcal{R}_n(\alpha) \cap \hat{\mathcal{I}}_n$.

Proof is in Appendix A.3.

This criterion elevates “point estimate significance” to “significance over all trusted candidate parameters”, naturally excluding spurious significance that may arise from relying only on point estimates while ignoring parameter correlations.

5 Multiple Experiments and Models: Intersection, Union, and Conflict Structure of Trust Regions

In reality, we often need to synthesize results from different experiments, different data sources, or even different models. This paper advocates: **The natural objects for multi-experiment aggregation are not “several point estimates”, but “several trust regions”.**

5.1 Intersection and Consensus of Multiple Trust Regions

Suppose there are K experiments/data sources/models giving trust regions $\mathcal{R}_{n_k}^{(k)}(\alpha_k)$ on same parameter space Θ , $k = 1, \dots, K$. Define overall consensus region

$$\mathcal{R}_{\text{cons}} := \bigcap_{k=1}^K \mathcal{R}_{n_k}^{(k)}(\alpha_k).$$

If the causal function of interest is $\psi : \Theta \rightarrow \mathbb{R}^d$, then its image on consensus region is

$$\mathcal{C}_{\text{cons}} := \{\psi(\theta) : \theta \in \mathcal{R}_{\text{cons}}\}.$$

If $\mathcal{R}_{\text{cons}}$ is non-empty and “small”, it indicates high consistency among different experiments; conversely, if $\mathcal{R}_{\text{cons}}$ is empty, can clearly say “there exists fundamental conflict among these experiments/models”, rather than vaguely relying on “some estimation differences”.

5.2 Union and Admissible Sets

On the other hand, define admissible region as

$$\mathcal{R}_{\text{perm}} := \bigcup_{k=1}^K \mathcal{R}_{n_k}^{(k)}(\alpha_k),$$

which characterizes the parameter set “supported by at least one experiment”. In some decision problems (such as tolerating partial experiment failure or model misspecification), we may only require conclusions to hold on $\mathcal{R}_{\text{perm}}$.

5.3 Conflict Region and Uncertainty Decomposition

Define conflict region as symmetric difference

$$\mathcal{R}_{\text{conflict}} := \left(\bigcup_{k=1}^K \mathcal{R}_{n_k}^{(k)} \right) \setminus \left(\bigcap_{k=1}^K \mathcal{R}_{n_k}^{(k)} \right),$$

where if some point θ is only supported by partial experiments (while excluded by others), then belongs to this region. By visualizing parameter space as partition of “consensus–conflict–unconstrained”, can intuitively identify which parameter directions’ conclusions are most sensitive to experiment selection.

6 Experimental Design and Observation Planning: Trust Region as Objective Function

In above framework, the essence of experimental design is: **Shaping the shape and size of future trust regions through choosing experimental schemes or observation strategies**. This section provides formalized characterization under classical linear models and general Fisher information background.

6.1 Fisher Information and Region Volume

In regular models, with sample size n , information matrix can usually be expressed as

$$I_n(\theta) = nI_1(\theta),$$

where $I_1(\theta)$ is information from single observation. For given design parameter ξ (such as distribution of samples over different treatment/covariate configurations), single observation information can be written as $I_1(\theta; \xi)$. Therefore,

$$I_n(\theta; \xi) = nI_1(\theta; \xi).$$

Volume of ellipsoidal trust region is proportional to $\det(I_n(\theta_0; \xi))^{-1/2}$. More precisely, when $\mathcal{R}_n(\alpha; \xi)$ is ellipsoid based on $I_n(\theta_0; \xi)$, its Lebesgue volume is

$$\text{Vol}(\mathcal{R}_n(\alpha; \xi)) = C_{d,\alpha} (\det I_n(\theta_0; \xi))^{-1/2},$$

where constant $C_{d,\alpha}$ only depends on dimension d and α . Therefore, minimizing region volume is equivalent to maximizing $\det I_n(\theta_0; \xi)$.

Definition 6.1 (Error Geometric Characterization of D-Optimal Design). If design ξ^* satisfies

$$\det I_n(\theta_0; \xi^*) = \sup_{\xi} \det I_n(\theta_0; \xi),$$

then call ξ^* D-optimal design. Geometrically, it makes trust region volume minimal under given n , thus most compact overall.

Classical D-optimality conclusions are restated in error geometric language in Appendix A.4.

6.2 Directional Resolution: A-Optimal and c-Optimal

If focus is on specific linear causal effect $\psi(\theta) = c^\top \theta$, then its asymptotic variance is

$$\text{Var}(\hat{\psi}) \approx \frac{1}{n} c^\top I_1(\theta_0; \xi)^{-1} c.$$

From error geometry perspective, this is exactly the squared length of principal semi-axis of trust ellipsoid in direction c (ignoring constants). Therefore, minimizing this variance is equivalent to maximizing resolution in direction c .

Definition 6.2 (c-Optimal Design). If design ξ^* satisfies

$$c^\top I_1(\theta_0; \xi^*)^{-1} c = \inf_{\xi} c^\top I_1(\theta_0; \xi)^{-1} c,$$

then call ξ^* c-optimal design.

From geometric perspective: c-optimal design does not pursue minimal overall ellipsoid volume, but specifically compresses semi-axis in direction c , i.e., focuses on enhancing geometric resolution of this causal effect.

7 Examples of Error Geometry in Typical Models

This section briefly demonstrates specific forms of error geometry framework under several common models for readers to gain intuitive impression.

7.1 Confidence Ellipsoid and Effect Interval in Linear Regression Models

Consider linear regression model

$$Y_i = x_i^\top \beta + \varepsilon_i, \quad \varepsilon_i \sim \mathcal{N}(0, \sigma^2),$$

where $x_i \in \mathbb{R}^p$ are known covariates, $\beta \in \mathbb{R}^p$ are regression coefficients. Let X be design matrix, OLS estimate is

$$\hat{\beta} = (X^\top X)^{-1} X^\top Y.$$

Classical result gives

$$\hat{\beta} \sim \mathcal{N}(\beta, \sigma^2(X^\top X)^{-1}).$$

Therefore, information matrix is $I(\beta) = \sigma^{-2} X^\top X$, trust ellipsoid is

$$\mathcal{R}(\alpha) = \{\beta : (\beta - \hat{\beta})^\top X^\top X(\beta - \hat{\beta}) \leq \sigma^2 \chi_{p,1-\alpha}^2\}.$$

For any linear prediction $\psi(\beta) = x_{\text{new}}^\top \beta$, its interval on $\mathcal{R}(\alpha)$ is

$$x_{\text{new}}^\top \hat{\beta} \pm \sqrt{\chi_{p,1-\alpha}^2 \sigma^2 x_{\text{new}}^\top (X^\top X)^{-1} x_{\text{new}}},$$

which is completely consistent with classical linear regression confidence interval, but in this framework is interpreted as “geometric projection of trust ellipsoid in direction x_{new} ”.

7.2 Identifiable Set and Ellipsoid Constraint in Instrumental Variable Models

In simple linear IV models, when weak instruments exist or exclusion hypothesis is violated, structural parameters may only be partially identifiable. At this time identifiable set \mathcal{I} is usually some polyhedron or more complex convex set. Intersecting it with confidence ellipsoid $\mathcal{R}_n(\alpha)$, can obtain geometric constraints on structural parameters, thus providing robust intervals for range of local average treatment effect.

In many cases, this problem can be reduced to “extremum of linear function on intersection of ellipsoid and polyhedron”, solvable using quadratic programming, essentially the geometric robustness criterion described in Section 4.4.

8 Discussion and Outlook

This paper provides a unified perspective that elevates “error” to “geometric boundary”, and systematically restates classical topics such as confidence intervals, causal identifiable sets, multi-experiment aggregation, and experimental design on this basis. Core points can be summarized as:

1. Trust region $\mathcal{R}_n(\alpha)$ is an ellipsoid or more general convex set in parameter space, carrying local metric structure induced by Fisher information;
2. Causal conclusions should be tested on $\mathcal{R}_n(\alpha) \cap \hat{\mathcal{I}}_n$ rather than merely relying on point estimates;
3. Multiple experiments and models naturally induce intersections, unions, and symmetric differences of trust regions, whose geometric decomposition characterizes consensus, conflict, and uncertainty structure;
4. Experimental design can be viewed as an optimization problem of “shaping future trust region geometry”, formally equivalent to classical criteria such as D-optimal, A-optimal, and c-optimal.

Further directions include: combining this paper’s error geometry framework with information geometry (Fisher–Rao metric), functional parameters (such as density curves or response surfaces), high-dimensional sparse structures (such as polyhedral regions under ℓ_1 constraints), and dynamic causal structures (time series and event history models); in these extensions, “trust regions” will no longer be finite-dimensional ellipsoids, but complex sets on Banach spaces or manifolds, yet their essence as “error boundaries” remains unchanged.

A Proofs of Main Theorems and Propositions

A.1 Proof of Theorem ?? (Asymptotic Coverage of Confidence Ellipsoid)

Theorem Restatement. Under regularity conditions, let $\hat{\theta}_n$ satisfy

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}(0, I(\theta_0)^{-1}),$$

and $\hat{I}_n \xrightarrow{P} I(\theta_0)$, then for any fixed $\alpha \in (0, 1)$,

$$P_{\theta_0}(\theta_0 \in \mathcal{R}_n(\alpha)) \rightarrow 1 - \alpha.$$

Proof. Let

$$Z_n := \sqrt{n}(\hat{\theta}_n - \theta_0), \quad I_0 := I(\theta_0),$$

then $Z_n \xrightarrow{d} Z \sim \mathcal{N}(0, I_0^{-1})$. Note that event $\theta_0 \in \mathcal{R}_n(\alpha)$ is equivalent to

$$n(\theta_0 - \hat{\theta}_n)^\top \hat{I}_n(\theta_0 - \hat{\theta}_n) \leq \chi_{d,1-\alpha}^2,$$

i.e.,

$$Z_n^\top \hat{I}_n Z_n \leq \chi_{d,1-\alpha}^2.$$

By Slutsky's theorem, $\hat{I}_n \xrightarrow{P} I_0$, so $Z_n^\top \hat{I}_n Z_n \xrightarrow{d} Z^\top I_0 Z$. For $Z \sim \mathcal{N}(0, I_0^{-1})$,

$$Z^\top I_0 Z \sim \chi_d^2.$$

Therefore,

$$P_{\theta_0}(\theta_0 \in \mathcal{R}_n(\alpha)) = P(Z_n^\top \hat{I}_n Z_n \leq \chi_{d,1-\alpha}^2) \rightarrow P(Z^\top I_0 Z \leq \chi_{d,1-\alpha}^2) = 1 - \alpha.$$

QED.

A.2 Proof of Proposition ?? (Optimal Bounds for Linear Target)

Proposition Restatement. On $\mathcal{R}_n(\alpha)$, minimum and maximum values of linear function $\psi(\theta) = c^\top \theta$ are

$$\psi_{\max,n} = c^\top \hat{\theta}_n + \sqrt{\frac{\chi_{d,1-\alpha}^2}{n} c^\top \hat{I}_n^{-1} c},$$

$$\psi_{\min,n} = c^\top \hat{\theta}_n - \sqrt{\frac{\chi_{d,1-\alpha}^2}{n} c^\top \hat{I}_n^{-1} c}.$$

Proof. Let $h = \theta - \hat{\theta}_n$, constraint becomes

$$nh^\top \hat{I}_n h \leq \chi_{d,1-\alpha}^2.$$

Objective function is

$$\psi(\theta) = c^\top \hat{\theta}_n + c^\top h.$$

Therefore maximizing $\psi(\theta)$ is equivalent to maximizing $c^\top h$ on ellipsoid

$$\mathcal{E} := \{h : h^\top \hat{I}_n h \leq \chi_{d,1-\alpha}^2 / n\}.$$

Standard quadratically constrained linear programming conclusion or Cauchy–Schwarz inequality gives

$$\sup_{h \in \mathcal{E}} c^\top h = \sqrt{\chi_{d,1-\alpha}^2/n} \sqrt{c^\top \hat{I}_n^{-1} c},$$

minimum value is opposite number. Therefore

$$\psi_{\max,n} = c^\top \hat{\theta}_n + \sqrt{\frac{\chi_{d,1-\alpha}^2}{n} c^\top \hat{I}_n^{-1} c},$$

$$\psi_{\min,n} = c^\top \hat{\theta}_n - \sqrt{\frac{\chi_{d,1-\alpha}^2}{n} c^\top \hat{I}_n^{-1} c}.$$

QED.

A.3 Proof of Theorem ?? (Geometric Robustness Criterion for Linear Causal Effect)

Theorem Restatement. Let $\mathcal{R}_n(\alpha)$ be trust ellipsoid, $\hat{\mathcal{I}}_n = \{\theta : A\theta \leq b\}$ be linear identifiable set, $\psi(\theta) = c^\top \theta$. If

$$\psi_{\min,n}^* := \inf\{c^\top \theta : \theta \in \mathcal{R}_n(\alpha), A\theta \leq b\} \geq \delta > 0,$$

then at significance level α , can assert that for all $\theta \in \mathcal{R}_n(\alpha) \cap \hat{\mathcal{I}}_n$, $\psi(\theta) \geq \delta > 0$, i.e., “causal effect is positive and at least δ ” is geometrically robust.

Proof. By definition, any θ in $\mathcal{R}_n(\alpha) \cap \hat{\mathcal{I}}_n$ satisfies $A\theta \leq b$ and $\theta \in \mathcal{R}_n(\alpha)$. Therefore, for any $\theta \in \mathcal{R}_n(\alpha) \cap \hat{\mathcal{I}}_n$,

$$\psi(\theta) = c^\top \theta \geq \inf\{c^\top \theta : \theta \in \mathcal{R}_n(\alpha), A\theta \leq b\} = \psi_{\min,n}^* \geq \delta.$$

That is

$$\mathcal{C}_n(\alpha) = \{\psi(\theta) : \theta \in \mathcal{R}_n(\alpha) \cap \hat{\mathcal{I}}_n\} \subset [\delta, \infty).$$

Since construction of $\mathcal{R}_n(\alpha)$ guarantees its asymptotic coverage probability is $1 - \alpha$, under usual conditions can consider “true parameter θ_0 ” falls in $\mathcal{R}_n(\alpha)$ with probability approximately $1 - \alpha$; if $\theta_0 \in \mathcal{I}$ and $\hat{\mathcal{I}}_n$ is good approximation, then above conclusion can be interpreted in asymptotic sense as level α significant conclusion. Rigorous probabilistic statement requires introducing joint coverage and uniform convergence, not elaborated here. QED.

A.4 Equivalence of D-Optimal Design and Trust Region Volume

Proposition A.1 (Error Geometric Characterization of D-Optimal Design). *In regular model with sample size n , information matrix $I_n(\theta_0; \xi) = nI_1(\theta_0; \xi)$, volume of confidence ellipsoid $\mathcal{R}_n(\alpha; \xi)$ constructed based on $I_n(\theta_0; \xi)$ satisfies*

$$\text{Vol}(\mathcal{R}_n(\alpha; \xi)) = C_{d,\alpha} (\det I_n(\theta_0; \xi))^{-1/2},$$

where $C_{d,\alpha}$ is independent of ξ . Therefore, maximizing $\det I_n(\theta_0; \xi)$ is equivalent to minimizing $\text{Vol}(\mathcal{R}_n(\alpha; \xi))$.

Proof. Let $A(\xi) := I_n(\theta_0; \xi)$, then ellipsoid is defined as

$$\mathcal{R}_n(\alpha; \xi) = \{\theta : (\theta - \hat{\theta}_n)^\top A(\xi)(\theta - \hat{\theta}_n) \leq c\},$$

where $c = \chi_{d,1-\alpha}^2$ is constant. Let $y = A(\xi)^{1/2}(\theta - \hat{\theta}_n)$, then mapping $\theta \mapsto y$ is linear homeomorphism with Jacobian $\det A(\xi)^{1/2}$. In y -space, ellipsoid becomes Euclidean ball $B_d(\sqrt{c})$ with radius \sqrt{c} , whose volume is

$$\text{Vol}(B_d(\sqrt{c})) = \frac{\pi^{d/2} c^{d/2}}{\Gamma(d/2 + 1)}.$$

Therefore

$$\text{Vol}(\mathcal{R}_n(\alpha; \xi)) = \frac{\text{Vol}(B_d(\sqrt{c}))}{\det A(\xi)^{1/2}} = C_{d,\alpha} (\det A(\xi))^{-1/2},$$

where $C_{d,\alpha} := \text{Vol}(B_d(\sqrt{c}))$ is independent of ξ . Thus minimizing volume is equivalent to maximizing $\det A(\xi) = \det I_n(\theta_0; \xi)$. QED.

References (Indicative)

1. Bickel, P. J., & Doksum, K. A. (2015). *Mathematical Statistics: Basic Ideas and Selected Topics*.
2. van der Vaart, A. W. (1998). *Asymptotic Statistics*.
3. Lehmann, E. L., & Casella, G. (1998). *Theory of Point Estimation*.
4. Pukelsheim, F. (2006). *Optimal Design of Experiments*.
5. Manski, C. F. (2003). *Partial Identification of Probability Distributions*.
6. Imbens, G. W., & Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences*.
7. Pearl, J. (2009). *Causality: Models, Reasoning, and Inference*.