On and Off Semiparametric Models

Jon A. Wellner

University of Washington

- joint work with Takumi Saegusa,
  University of Washington, Biostatistics
- Talk at First IMS Asia-Pacific Rim Conference,
- Seoul, Korea
- June 29, 2009
- *Email: jaw@stat.washington.edu*
  http: //www.stat.washington.edu/jaw/jaw.research.html

Outline

- Introduction: estimation theory on a model $\mathcal{P}$

Outline

- Introduction: estimation theory on a model $\mathcal{P}$
- Estimation theory off a model $\mathcal{P}$: notation and concepts

Outline

- Introduction: estimation theory on a model $\mathcal{P}$
- Estimation theory off a model $\mathcal{P}$: notation and concepts
- Estimation theory off a model $\mathcal{P}$: three questions

Outline

- Introduction: estimation theory on a model $\mathcal{P}$
- Estimation theory off a model $\mathcal{P}$: notation and concepts
- Estimation theory off a model $\mathcal{P}$: three questions
- An answer to question 1

Outline

- Introduction: estimation theory on a model $\mathcal{P}$
- Estimation theory off a model $\mathcal{P}$: notation and concepts
- Estimation theory off a model $\mathcal{P}$: three questions
- An answer to question 1
- An answer to questions 2 and 3

Outline

- Introduction: estimation theory on a model $\mathcal{P}$
- Estimation theory off a model $\mathcal{P}$: notation and concepts
- Estimation theory off a model $\mathcal{P}$: three questions
- An answer to question 1
- An answer to questions 2 and 3
- Examples
  - Example 1. Symmetric location model
  - Example 2. Paired exponential mixture model

Outline

- Introduction: estimation theory on a model $\mathcal{P}$
- Estimation theory off a model $\mathcal{P}$: notation and concepts
- Estimation theory off a model $\mathcal{P}$: three questions
- An answer to question 1
- An answer to questions 2 and 3
- Examples
  - Example 1. Symmetric location model
  - Example 2. Paired exponential mixture model
- Summary; further problems

1. Introduction: Estimation theory on a model $\mathcal{P}$

- Suppose $\mathcal{P}$ is a semiparametric model.

1. Introduction: Estimation theory on a model $\mathcal{P}$

- Suppose $\mathcal{P}$ is a semiparametric model.
- Fix $P_0 \in \mathcal{P}$; assume $X_1, \ldots, X_n$ are i.i.d. P_0.

1. Introduction: Estimation theory on a model $\mathcal{P}$

- Suppose $\mathcal{P}$ is a semiparametric model.
- Fix $P_0 \in \mathcal{P}$; assume $X_1, \ldots, X_n$ are i.i.d. P_0.
- Let $\dot{\mathcal{P}} \equiv \dot{\mathcal{P}}(P_0)$ be the tangent space of $\mathcal{P}$ at P_0.

1. Introduction: Estimation theory on a model $\mathcal{P}$

- Suppose $\mathcal{P}$ is a semiparametric model.
- Fix $P_0 \in \mathcal{P}$; assume $X_1, \ldots, X_n$ are i.i.d. P_0.
- Let $\dot{\mathcal{P}} \equiv \dot{\mathcal{P}}(P_0)$ be the tangent space of $\mathcal{P}$ at P_0.
- Assume that $\dot{\mathcal{P}} \subsetneq L_2^0(P_0)$; thus $L_2^0(P_0) = \dot{\mathcal{P}} + \dot{\mathcal{P}}^{\perp}$ where $\dot{\mathcal{P}}^{\perp} \neq \emptyset$.

1. Introduction: Estimation theory on a model $\mathcal{P}$

- Suppose $\mathcal{P}$ is a semiparametric model.
- Fix $P_0 \in \mathcal{P}$; assume $X_1, \ldots, X_n$ are i.i.d. P_0.
- Let $\dot{\mathcal{P}} \equiv \dot{\mathcal{P}}(P_0)$ be the tangent space of $\mathcal{P}$ at P_0.
- Assume that $\dot{\mathcal{P}} \subsetneq L_2^0(P_0)$; thus $L_2^0(P_0) = \dot{\mathcal{P}} + \dot{\mathcal{P}}^\perp$ where $\dot{\mathcal{P}}^\perp \neq \emptyset$.
- Suppose that $\nu = \nu(P)$ is a (differentiable) parameter: i.e. $\nu : \mathcal{P} \to \mathbb{R}$.

- Suppose that $\widehat{\nu}_n$ is an inefficient estimator of $\nu(P)$ at P_0 with influence function ψ: thus

$$\begin{aligned}\sqrt{n}(\hat{\nu}_n - \nu(P_0)) &= \sqrt{n}\mathbb{P}_n\psi + o_p(1)\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi(X_i) + o_p(1)\end{aligned}$$

where $E_0\psi(X_1) = 0$, $E_0\psi^2 < \infty$, and $\psi \notin \dot{\mathcal{P}}$.

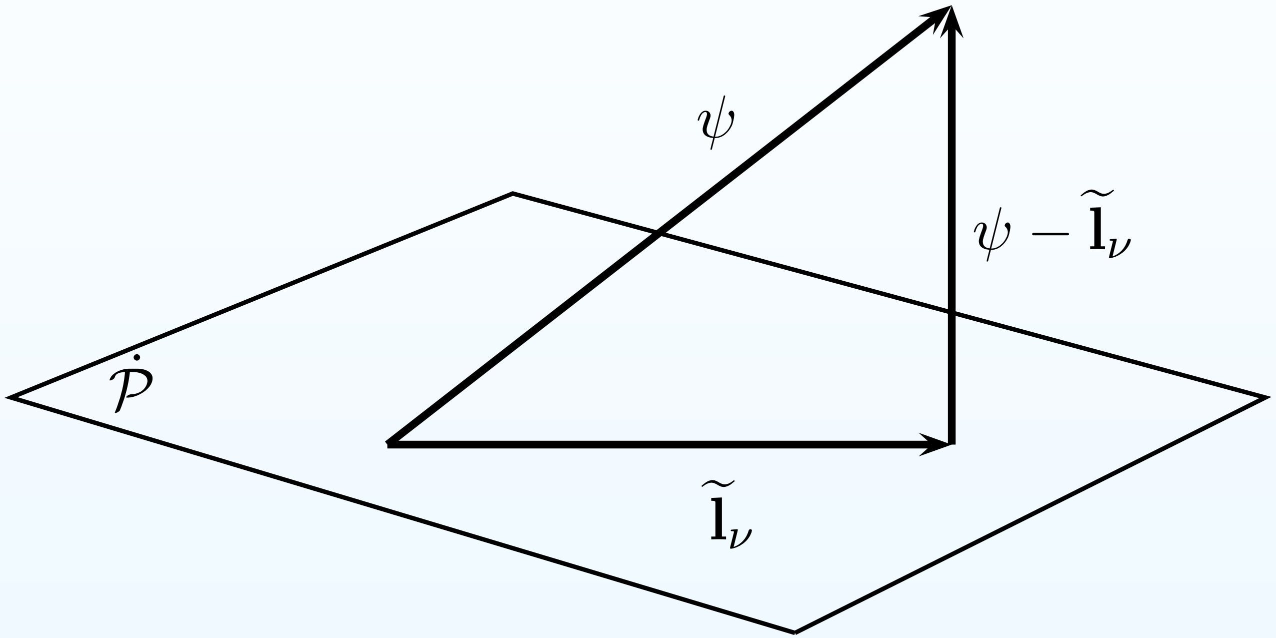
- Suppose that $\widehat{\nu}_n$ is an inefficient estimator of $\nu(P)$ at P_0 with influence function ψ: thus

$$\begin{aligned}\sqrt{n}(\hat{\nu}_n - \nu(P_0)) &= \sqrt{n}\mathbb{P}_n\psi + o_p(1)\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi(X_i) + o_p(1)\end{aligned}$$

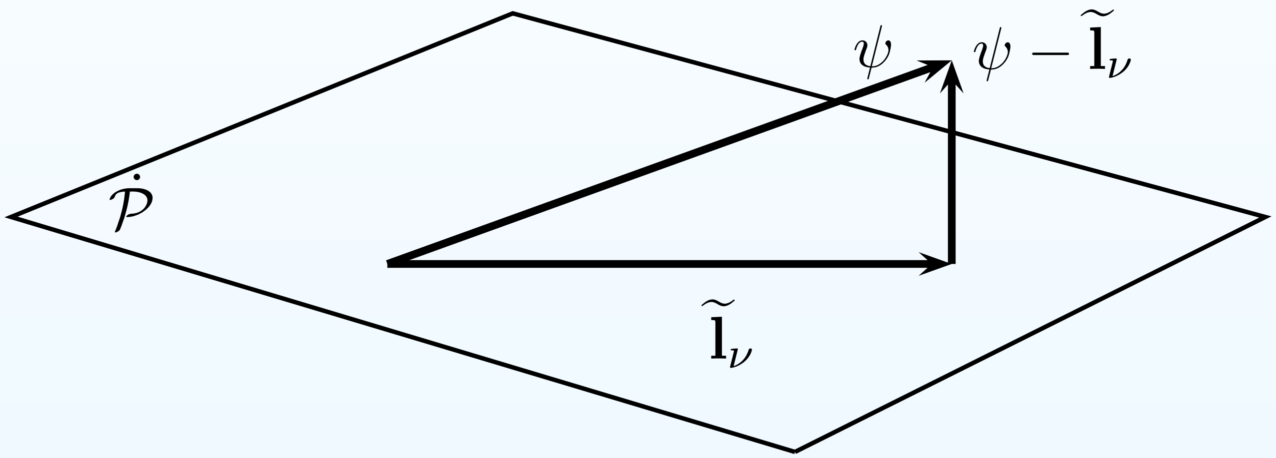
  where $E_0\psi(X_1) = 0$, $E_0\psi^2 < \infty$, and $\psi \notin \dot{\mathcal{P}}$.
- It is well - known that the efficient influence function $\widetilde{\mathrm{l}}_\nu$ for estimation of ν on $\mathcal{P}$ is given by

$$\widetilde{\mathrm{l}}_\nu = \Pi(\psi|\dot{\mathcal{P}}) \equiv \text{orthogonal projection of } \psi \text{ onto } \dot{\mathcal{P}};$$

  see e.g. Bickel, Klaassen, Ritov, and Wellner (1993, 1998), Proposition 1, page 65.



ψ
$\psi - \widetilde{\mathbf{l}}_\nu$
$\dot{\mathcal{P}}$
$\widetilde{\mathbf{l}}_\nu$


ψ
$\psi - \tilde{\mathbf{l}}_\nu$
$\dot{\mathcal{P}}$
$\tilde{\mathbf{l}}_\nu$

- If $\widehat{\nu}_n^{\text{eff}}$ is a locally regular and asymptotically efficient estimator of $\nu(P_0)$, then $\widehat{\nu}_n^{\text{eff}}$ satisfies

$$\begin{aligned}\sqrt{n}(\widehat{\nu}_n^{\text{eff}} - \nu(P_0)) &= \sqrt{n}\mathbb{P}_n\widetilde{\mathbf{l}}_\nu + o_p(1) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widetilde{\mathbf{l}}_\nu(X_i) + o_p(1).\end{aligned}$$

2. Estimation theory off a model $\mathcal{P}$

- George Box (1987): "Essentially, all models are wrong, but some are useful."

2. Estimation theory off a model $\mathcal{P}$

- George Box (1987): "Essentially, all models are wrong, but some are useful."
- Variant 1: "Remember that all models are wrong; the practical question is how wrong do they have to be to not be useful."

2. Estimation theory off a model $\mathcal{P}$

- George Box (1987): “Essentially, all models are wrong, but some are useful.”
- Variant 1: “Remember that all models are wrong; the practical question is how wrong do they have to be to not be useful.”
- Variant 2: “All models are false, but some models are useful.”

- Consider enlarging $\mathcal{P}$ by allowing an additional parametric sub-model $\mathcal{Q} \equiv \{ Q_\eta : \eta \in \mathbb{R}\}$ parametrized by a real parameter η which satisfies $Q_0 = P_0 \in \mathcal{P}$.

- Consider enlarging $\mathcal{P}$ by allowing an additional parametric sub-model $\mathcal{Q} \equiv \{ Q_\eta : \eta \in \mathbb{R}\}$ parametrized by a real parameter η which satisfies $Q_0 = P_0 \in \mathcal{P}$.
- Suppose that $\mathcal{Q}$ is regular with, for simplicity, densities $\{q_\eta\}$ with respect to a dominating measure μ and score function $a \in L_2^0(P_0)$ but $a \notin \dot{\mathcal{P}}$.

- Consider enlarging $\mathcal{P}$ by allowing an additional parametric sub-model $\mathcal{Q} \equiv \{ Q_\eta : \eta \in \mathbb{R}\}$ parametrized by a real parameter η which satisfies $Q_0 = P_0 \in \mathcal{P}$.
- Suppose that $\mathcal{Q}$ is regular with, for simplicity, densities $\{q_\eta\}$ with respect to a dominating measure μ and score function $a \in L_2^0(P_0)$ but $a \notin \dot{\mathcal{P}}$.
- Thus for the sequence of densities $q_n \equiv q_{n^{-1/2}}$,

$$\sqrt{n}\,\{\sqrt{q_n} - \sqrt{p_0}\} \to \frac{1}{2} a \sqrt{p_0} \qquad \text{in} \;\; L_2(\mu).$$

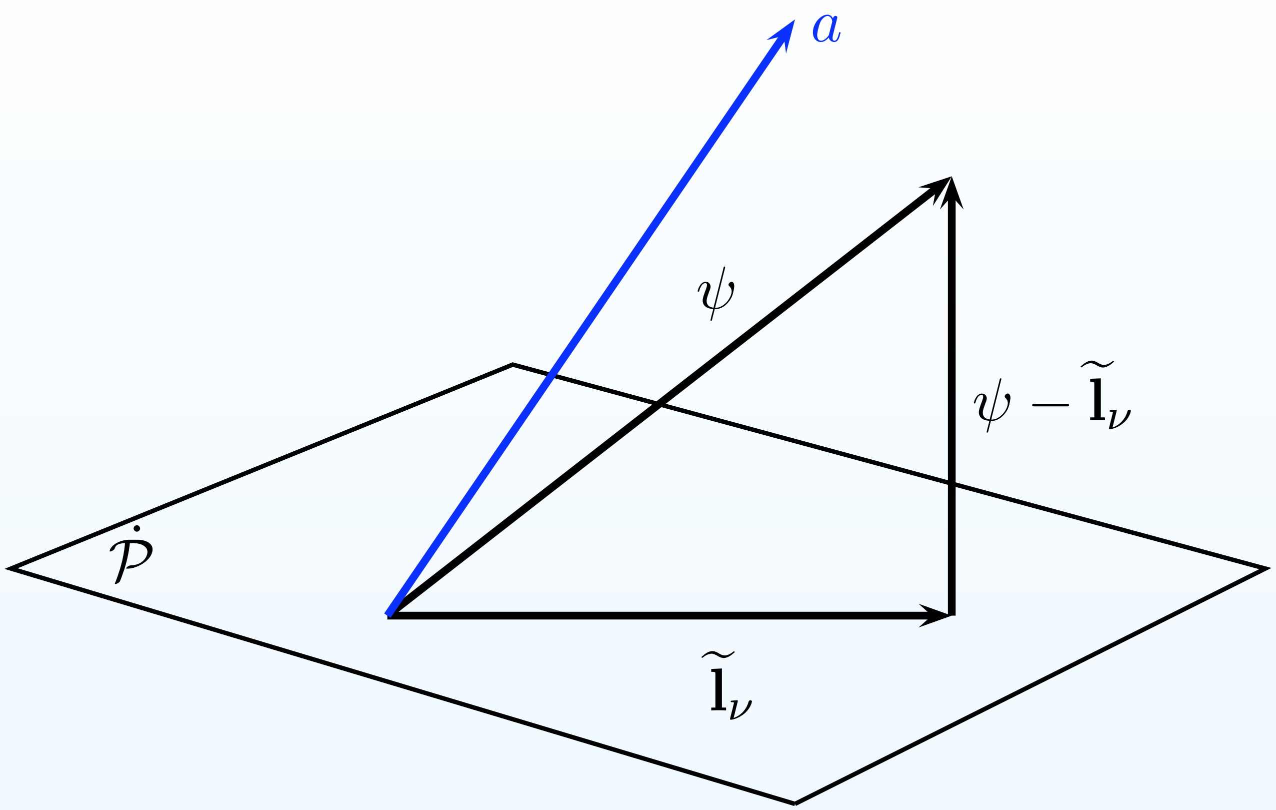
- Given $a \in L_2^0(P_0) \setminus \dot{\mathcal{P}}$, define a one-dimensional parametric submodel

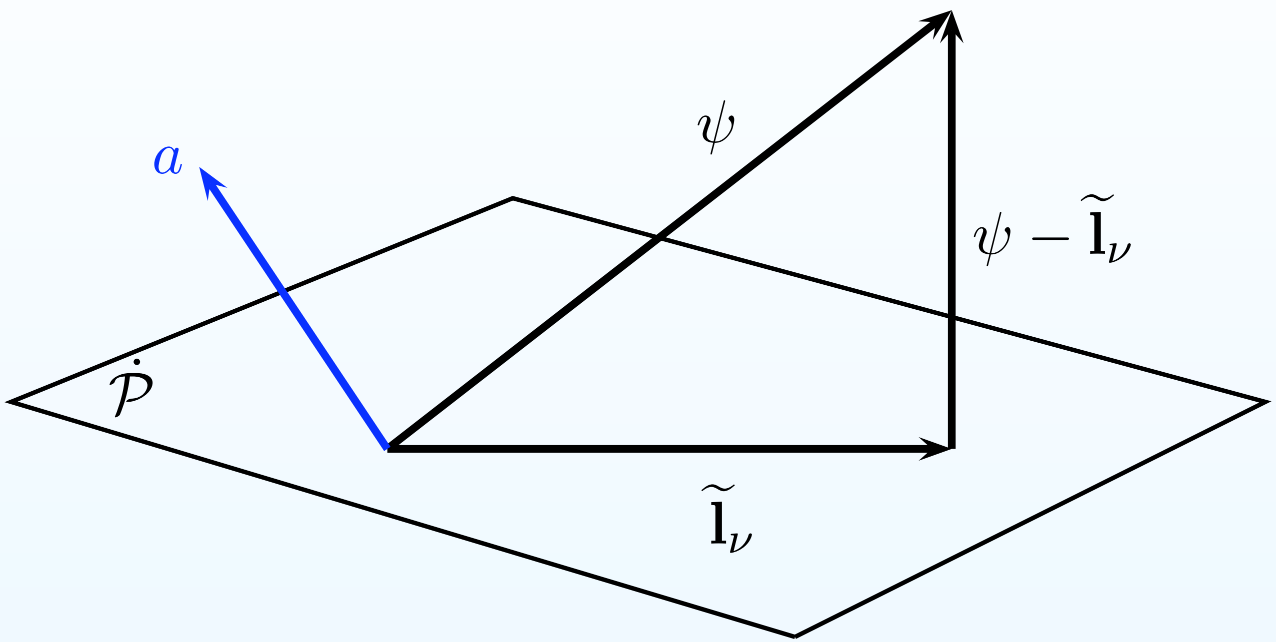
$$\mathcal{Q} = \{Q_\eta : \frac{dQ_\eta}{d\mu} \equiv q_\eta, \ \eta \in \mathbb{R}\}$$

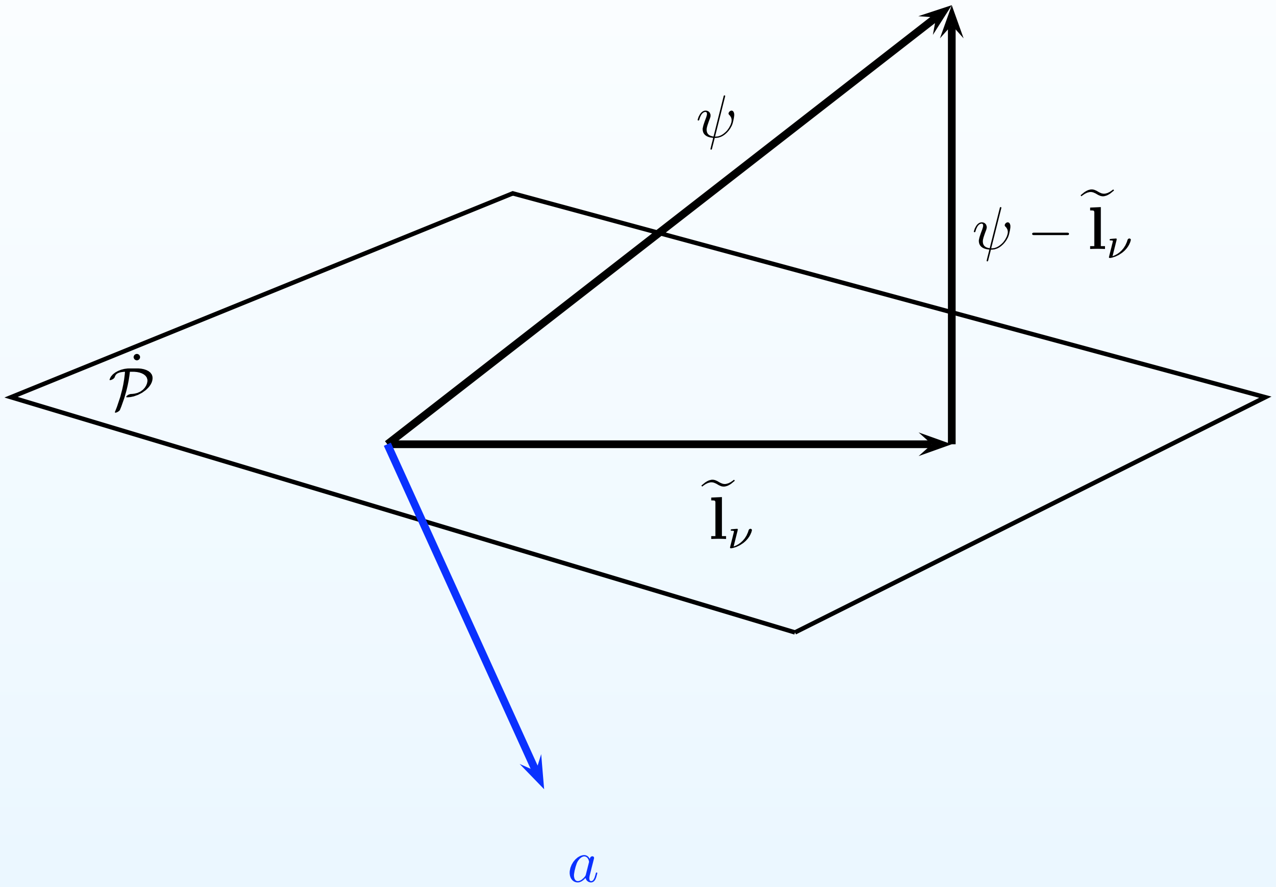
  where, with $m(x) = 2/(1 + e^{-x})$,

$$q_\eta(x) = p_0(x) \frac{m(\eta a(x))}{\int m(\eta a) dP_0}.$$

  Then Q_η has score function a and satisfies $Q_0 = P_0$.


a
ψ
$\psi - \widetilde{\mathbf{l}}_\nu$
$\dot{\mathcal{P}}$
$\widetilde{\mathbf{l}}_\nu$


a
ψ
$\psi - \widetilde{\mathbf{l}}_\nu$
$\dot{\mathcal{P}}$
$\widetilde{\mathbf{l}}_\nu$


ψ
$\psi - \widetilde{\mathbf{l}}_\nu$
$\dot{\mathcal{P}}$
$\widetilde{\mathbf{l}}_\nu$
a

- Now asymptotic linearity of $\widehat{\nu}_n$ and Le Cam's second lemma yields

$$\begin{pmatrix} \sqrt{n}(\widehat{\nu}_n - \nu(P_0)) \\ \log \frac{dQ_n^n}{dP_0^n} \end{pmatrix} \overset{P_0^n}{\longrightarrow}_d N_2\left(\begin{pmatrix} 0 \\ -\sigma^2/2 \end{pmatrix}, \begin{pmatrix} E\psi^2 & E(\psi a) \\ E(\psi a) & \sigma^2 \end{pmatrix} \right)$$

where $\sigma^2 = Ea^2(X)$.

- Hence by Le Cam's third lemma we find that

$$\begin{pmatrix} \sqrt{n}(\widehat{\nu}_n - \nu(P_0)) \\ \log \frac{dQ_n^n}{dP_0^n} \end{pmatrix}$$

$$\overset{Q_n^n}{\rightarrow}_d N_2 \left(\begin{pmatrix} E(\psi a) \\ +\sigma^2/2 \end{pmatrix}, \begin{pmatrix} E\psi^2 & E(\psi a) \\ E(\psi a) & \sigma^2 \end{pmatrix} \right)$$

where $\sigma^2 = Ea^2(X)$.

- Hence by Le Cam's third lemma we find that

$$\begin{pmatrix} \sqrt{n}(\widehat{\nu}_n - \nu(P_0)) \\ \log \frac{dQ_n^n}{dP_0^n} \end{pmatrix}$$
$$\overset{Q_n^n}{\to}_d N_2 \left(\begin{pmatrix} E(\psi a) \\ +\sigma^2/2 \end{pmatrix}, \begin{pmatrix} E\psi^2 & E(\psi a) \\ E(\psi a) & \sigma^2 \end{pmatrix} \right)$$

where $\sigma^2 = Ea^2(X)$.
- This implies that

$$\begin{aligned} \liminf_n E_{Q_n} \left\{ \sqrt{n}(\widehat{\nu}_n - \nu(P_0)) \right\}^2 & \geq E\psi^2 + \{E(\psi a)\}^2 \\ & \equiv \text{AMSE}_{\widehat{\nu}}(a). \end{aligned}$$

where $AMSE_T(a)$ stands for the "Asymptotic Mean Squared Error of the estimator T in the direction a".

- Repeating this argument for the efficient estimator $\widehat{\nu}_n^{\text{eff}}$ with efficient influence function $\widetilde{\mathbf{l}}_\nu$ yields

$$\begin{pmatrix} \sqrt{n}(\widehat{\nu}_n^{\text{eff}} - \nu(P_0)) \\ \log \frac{dQ_n^n}{dP_0^n} \end{pmatrix} \overset{Q_n^n}{\rightarrow}_d N_2\left(\begin{pmatrix} E(\widetilde{\mathbf{l}}_\nu a) \\ +\sigma^2/2 \end{pmatrix}, \begin{pmatrix} E\widetilde{\mathbf{l}}_\nu^2 & E(\widetilde{\mathbf{l}}_\nu a) \\ E(\widetilde{\mathbf{l}}_\nu a) & \sigma^2 \end{pmatrix}\right)$$

where $\sigma^2 = Ea^2(X)$.

- Repeating this argument for the efficient estimator $\widehat{\nu}_n^{\text{eff}}$ with efficient influence function $\widetilde{\mathbf{l}}_\nu$ yields

$$\begin{pmatrix} \sqrt{n}(\widehat{\nu}_n^{\text{eff}} - \nu(P_0)) \\ \log \frac{dQ_n^n}{dP_0^n} \end{pmatrix}$$

$$\overset{Q_n^n}{\rightarrow}_d N_2 \left(\begin{pmatrix} E(\widetilde{\mathbf{l}}_\nu a) \\ +\sigma^2/2 \end{pmatrix}, \begin{pmatrix} E\widetilde{\mathbf{l}}_\nu^2 & E(\widetilde{\mathbf{l}}_\nu a) \\ E(\widetilde{\mathbf{l}}_\nu a) & \sigma^2 \end{pmatrix} \right)$$

where $\sigma^2 = Ea^2(X)$.

- This implies that

$$\begin{aligned} \liminf_n E_{Q_n} \left\{ \sqrt{n}(\widehat{\nu}_n^{\text{eff}} - \nu(P_0)) \right\}^2 &\geq E\widetilde{\mathbf{l}}_\nu^2 + \{E(\widetilde{\mathbf{l}}_\nu a)\}^2 \\ &\equiv \text{AMSE}_{\widehat{\nu}^{\text{eff}}}(a). \end{aligned}$$

3. Estimation theory off a model $\mathcal{P}$: two questions

- Question 1: When does it hold that

$$\begin{aligned}\text{AMSE}_{\hat{\nu}}(a) &= E(\psi^2) + \{E(\psi a)\}^2 \\ &< E(\widetilde{\mathbf{l}}_\nu^2) + \{E(\widetilde{\mathbf{l}}_\nu a)\}^2 \\ &= \text{AMSE}_{\hat{\nu}\text{eff}}(a)?\end{aligned}$$

3. Estimation theory off a model $\mathcal{P}$: two questions

- Question 1: When does it hold that

$$\begin{aligned} \text{AMSE}_{\hat{\nu}}(a) &= E(\psi^2) + \{E(\psi a)\}^2 \\ &< E(\widetilde{\mathbf{l}}_\nu^2) + \{E(\widetilde{\mathbf{l}}_\nu a)\}^2 \\ &= \text{AMSE}_{\hat{\nu}\text{eff}}(a)? \end{aligned}$$

- Question 2: How large can the difference

$$\text{AMSE}_{\widehat{\nu}_n}(a) - \text{AMSE}_{\widehat{\nu}\text{eff}}(a)$$

be, subject to a bound on Ea^2? That is, can we bound

$$\sup\left\{E\psi^2 + \{E(\psi a)\}^2 - \left(E\widetilde{\mathbf{l}}_\nu^2 + \{E(\widetilde{\mathbf{l}}_\nu a)\}^2\right) :\ a \notin \dot{\mathcal{P}},\ Ea^2 \leq \kappa^2\right\}?$$

- Question 3: How large can the difference

$$\text{AMSE}_{\widehat{\nu}\text{eff}}(a) - \text{AMSE}_{\widehat{\nu}_n}(a)$$

be, subject to a bound on Ea^2? That is, can we bound

$$\sup \left\{ E\widetilde{\mathbf{l}}_\nu^2 + \{E(\widetilde{\mathbf{l}}_\nu a)\}^2 - \left(E\psi^2 + \{E(\psi a)\}^2\right) \, : \; a \notin \dot{\mathcal{P}}, \; Ea^2 \leq \kappa^2 \right\} ?$$

4. An answer to question 1

- **Proposition 1.** Let

$$\widetilde{a} = \Pi\left(a \,\middle|\, \left[\psi, \widetilde{\mathbf{l}}_\nu\right]\right) = c\frac{\psi - \widetilde{\mathbf{l}}_\nu}{E(\psi - \widetilde{\mathbf{l}}_\nu)^2} + d\frac{\widetilde{\mathbf{l}}_\nu}{E(\widetilde{\mathbf{l}}_\nu^2)}.$$

Then

$$\begin{aligned} \text{AMSE}_{\hat{\nu}}(a) &= E(\psi^2) + \{E(\psi a)\}^2 \\ &< E(\widetilde{\mathbf{l}}_\nu^2) + \{E(\widetilde{\mathbf{l}}_\nu a)\}^2 \qquad (1) \\ &= \text{AMSE}_{\hat{\nu}\text{eff}}(a) \end{aligned}$$

if $d^2 > E(\psi - \widetilde{\mathbf{l}}_\nu)^2 \equiv B^2$ and

$$-d - \sqrt{d^2 - B^2} < c < -d + \sqrt{d^2 - B^2}.$$

- **Proof:** Since

$$E(\psi^2) = E(\psi - \widetilde{\mathbf{l}}_\nu)^2 + E(\widetilde{\mathbf{l}}_\nu^2),$$

the inequality (1) is equivalent to

$$E(\psi - \widetilde{\mathbf{l}}_\nu)^2 + \{E(\psi a)\}^2 < \{E(\widetilde{\mathbf{l}}_\nu a)\}^2. \qquad (2)$$

But since

$$E(\widetilde{\mathbf{l}}_\nu a) = E(\widetilde{\mathbf{l}}_\nu \tilde{a}) = d, \qquad \text{and}$$

$$E(\psi a) = E(\psi \tilde{a}) = E\left(\psi - \widetilde{\mathbf{l}}_\nu + \widetilde{\mathbf{l}}_\nu\right)\left(c\frac{\psi - \widetilde{\mathbf{l}}_\nu}{E(\psi - \widetilde{\mathbf{l}}_\nu)^2} + d\frac{\widetilde{\mathbf{l}}_\nu}{E(\widetilde{\mathbf{l}}_\nu^2)}\right)$$

$$= c + d,$$

we can rewrite (2) as:

$$B^2 \equiv E(\psi - \widetilde{\mathbf{l}}_\nu)^2 < d^2 - (d+c)^2 = -2cd - c^2,$$

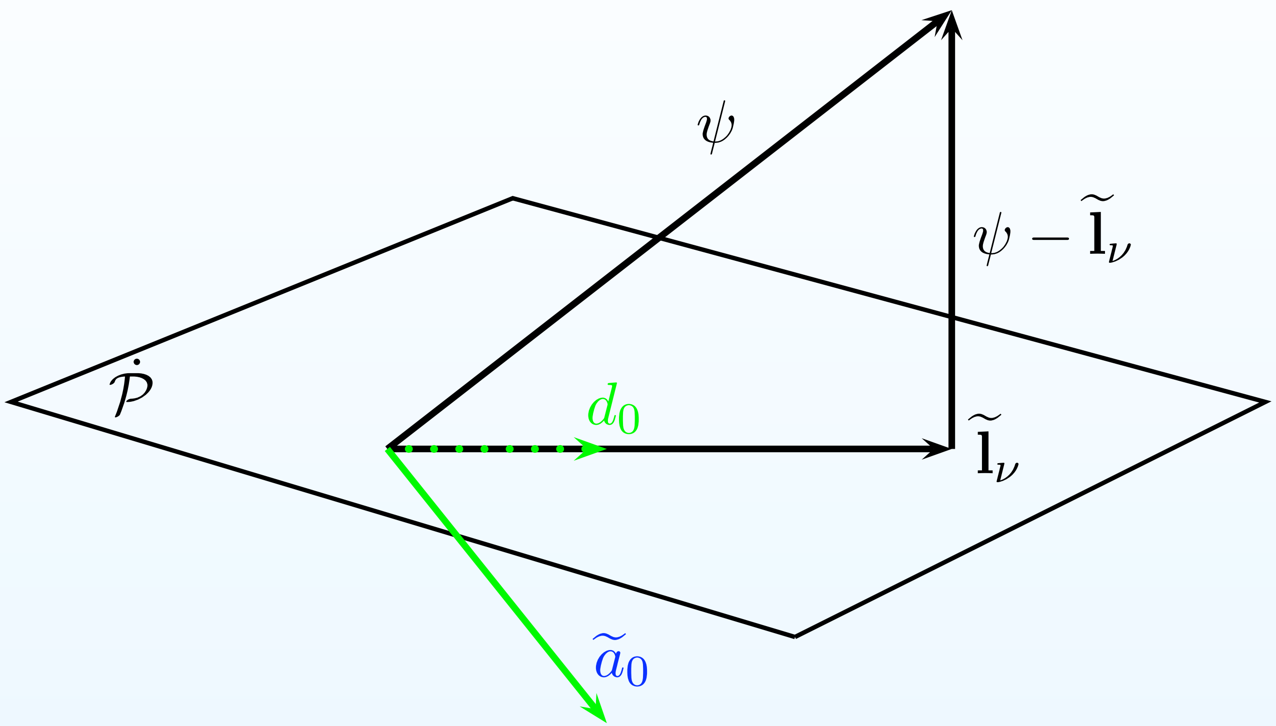
or, equivalently

$$c^2 + 2dc + B^2 < 0.$$

This holds only if $d^2 > B^2$, and

$$-d - \sqrt{d^2 - B^2} < c < -d + \sqrt{d^2 - B^2}. \qquad \square$$

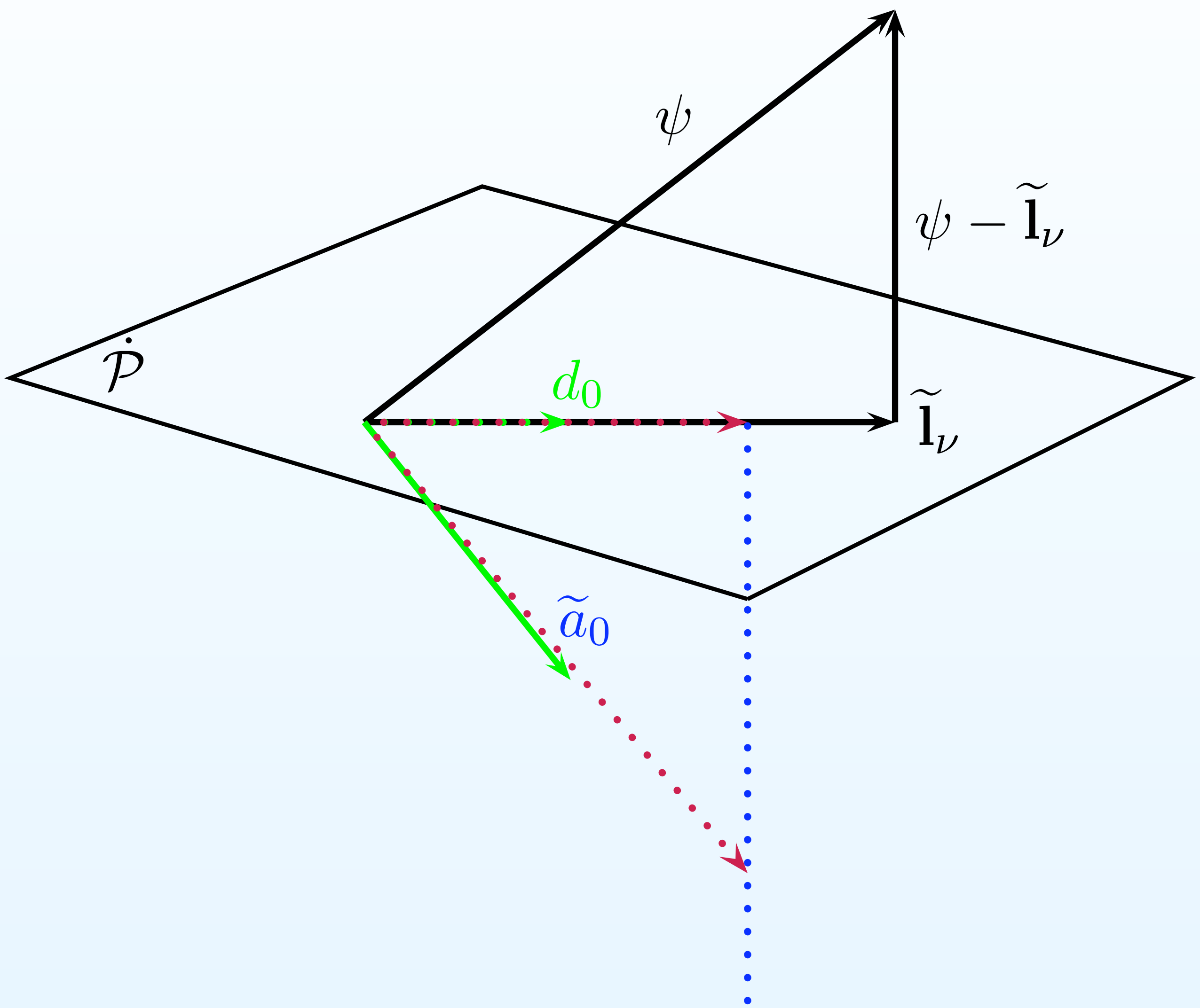
What does this mean geometrically?

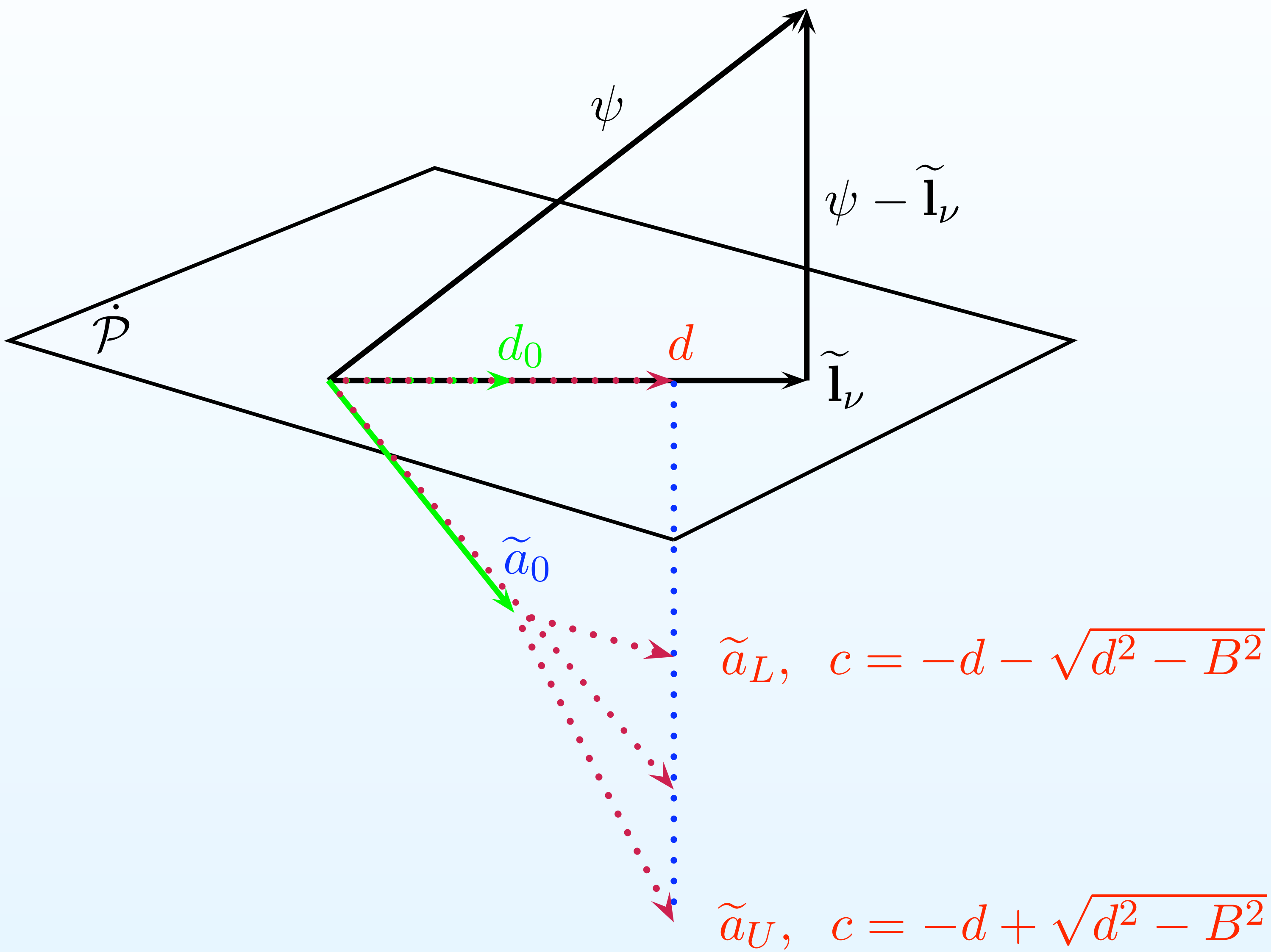


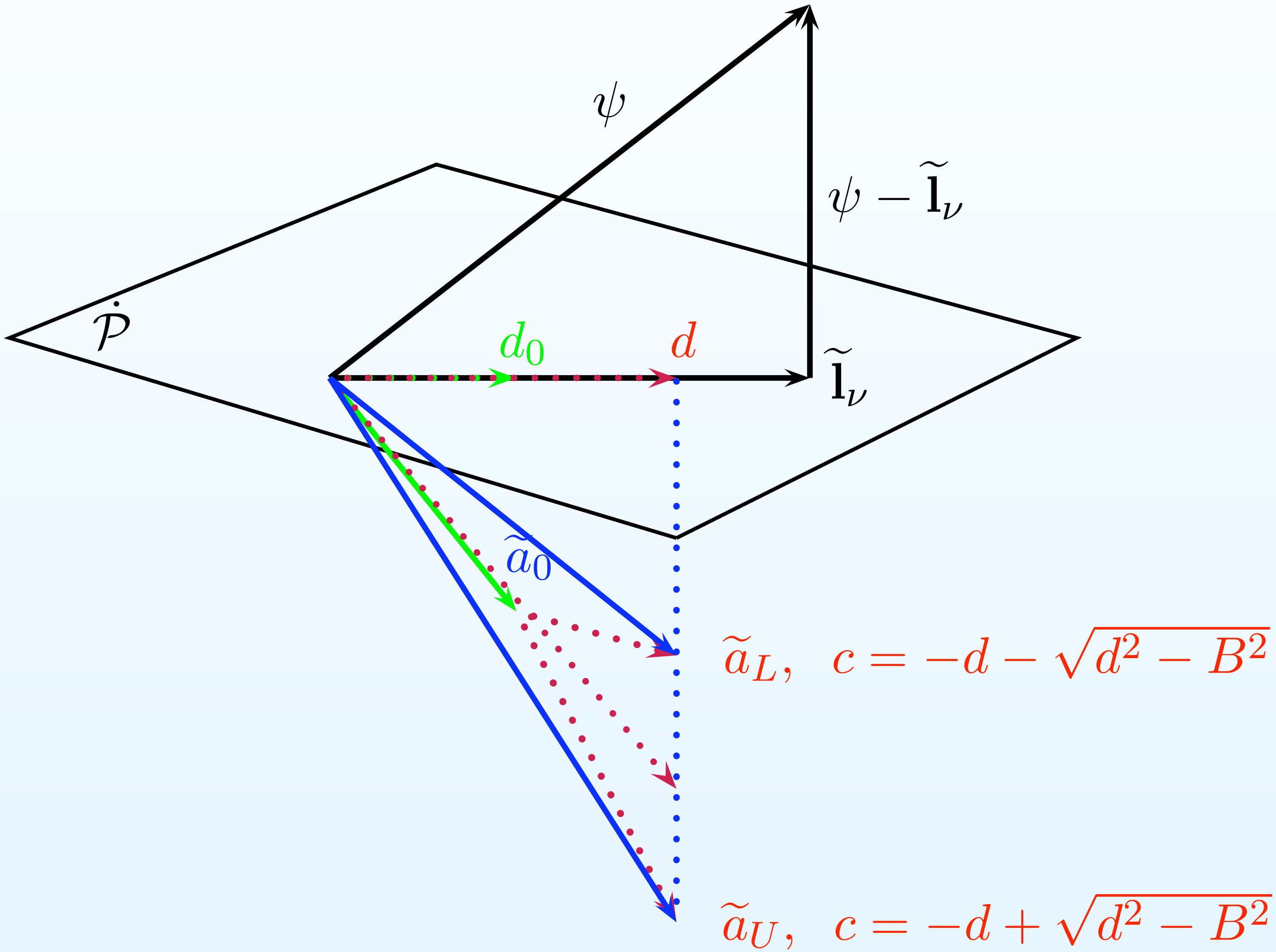
$$d_0^2 \equiv B^2 = E(\psi - \widetilde{\mathbf{l}}_\nu)^2$$

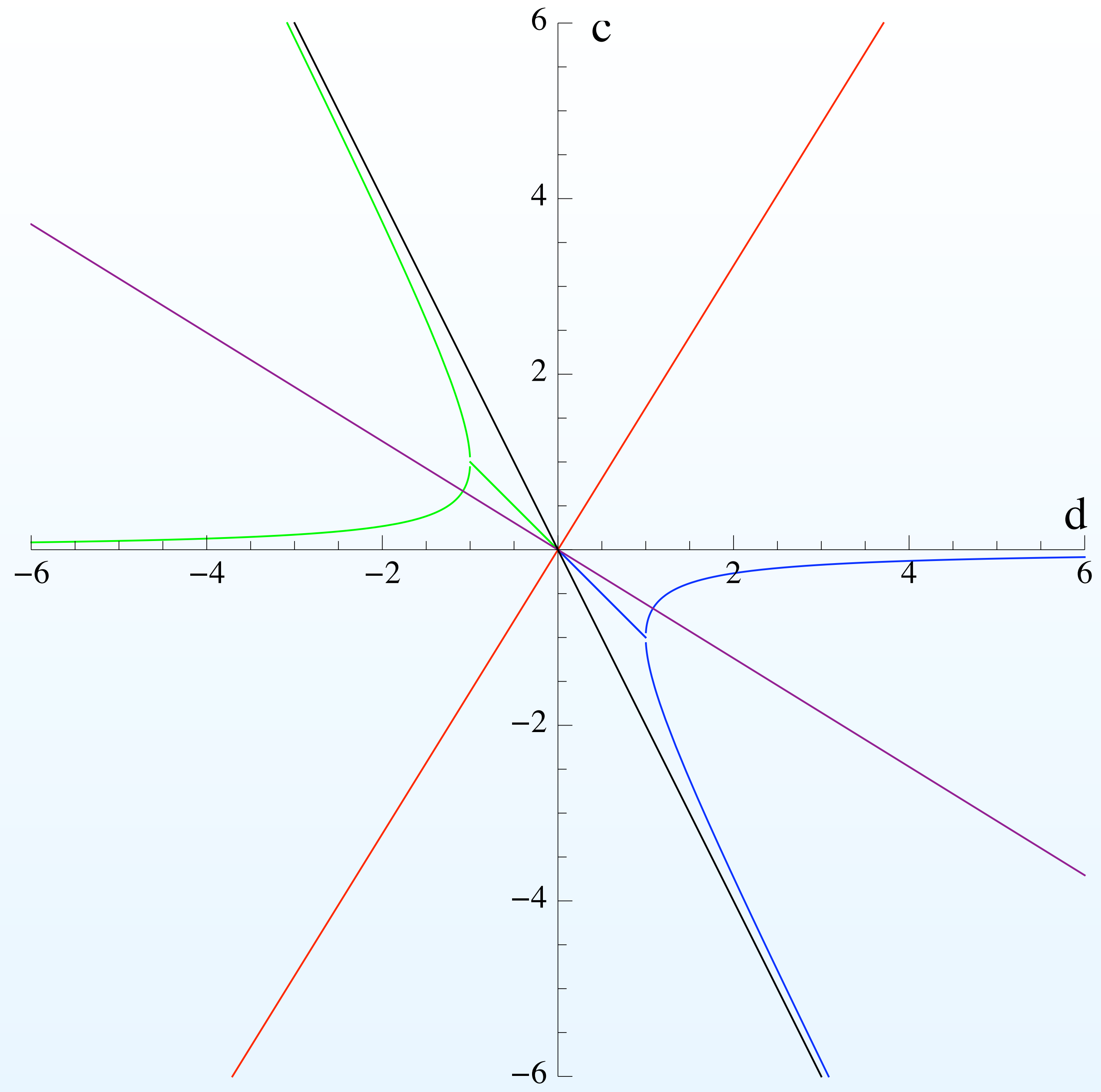
$$\widetilde{a}_0 = -d_0 \frac{\psi - \widetilde{\mathbf{l}}_\nu}{E(\psi - \widetilde{\mathbf{l}}_\nu)^2} + d_0 \frac{\widetilde{\mathbf{l}}_\nu}{E(\widetilde{\mathbf{l}}_\nu)^2}$$

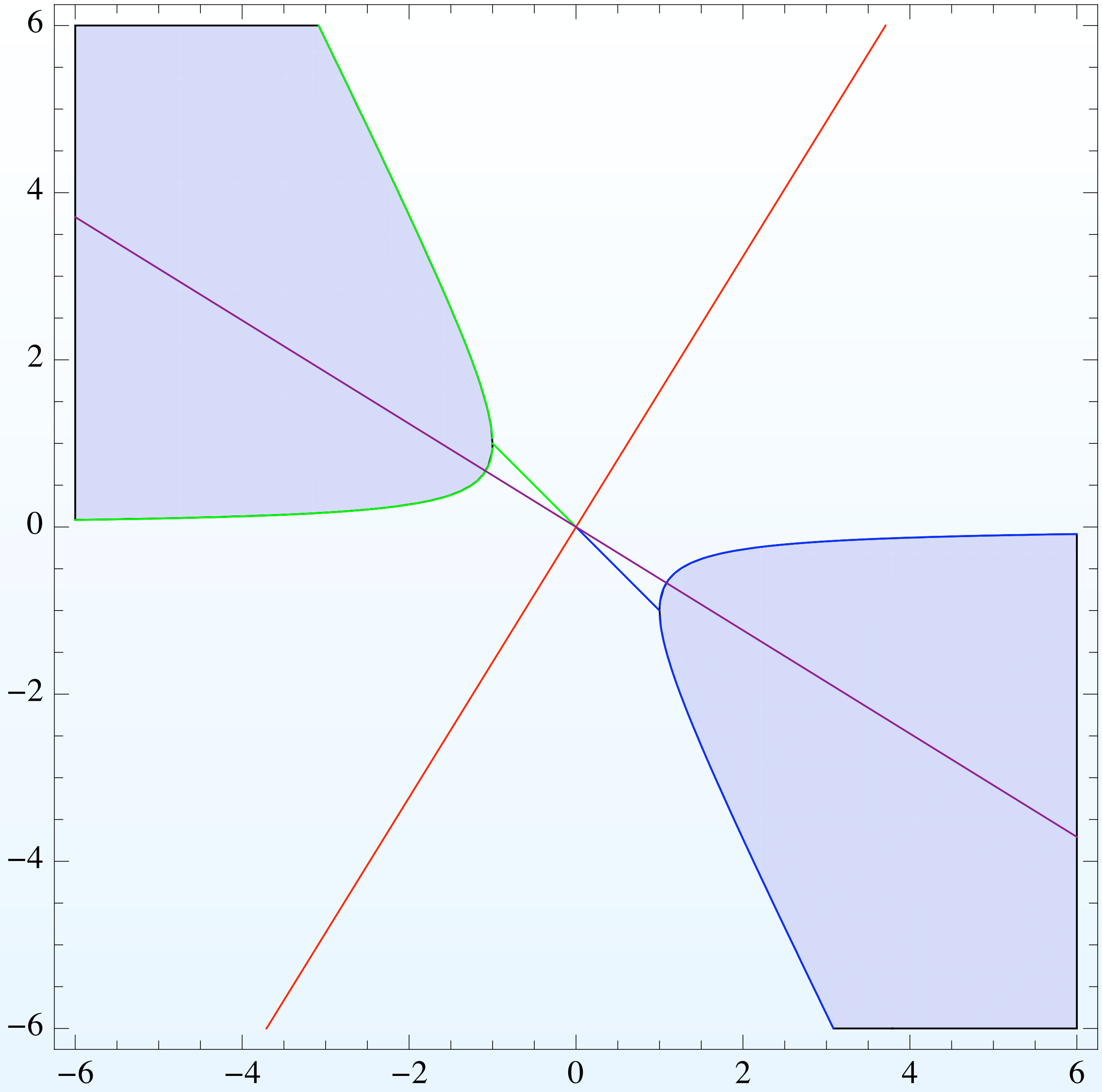
$$\widetilde{a}_0 \perp \psi$$


ψ
$\psi - \widetilde{\mathbf{l}}_\nu$
$\dot{\mathcal{P}}$
d_0
$\widetilde{\mathbf{l}}_\nu$
$\widetilde{a}_0$


ψ
$\psi - \widetilde{\mathbf{l}}_\nu$
$\dot{\mathcal{P}}$
d_0
d
$\widetilde{\mathbf{l}}_\nu$
$\widetilde{a}_0$
$\widetilde{a}_L, \quad c = -d - \sqrt{d^2 - B^2}$
$\widetilde{a}_U, \quad c = -d + \sqrt{d^2 - B^2}$


ψ
$\psi - \widetilde{\mathbf{l}}_\nu$
$\dot{\mathcal{P}}$
d_0
d
$\widetilde{\mathbf{l}}_\nu$
$\widetilde{a}_0$
$\widetilde{a}_L, \quad c = -d - \sqrt{d^2 - B^2}$
$\widetilde{a}_U, \quad c = -d + \sqrt{d^2 - B^2}$


c
d
6
4
2
−2
−4
−6
−6
−4
−2
2
4
6


6
4
2
0
−2
−4
−6
−6
−4
−2
0
2
4
6

5. An answer to question 2

- **"Proposition 2.F"** (F for false! or first attempt)

$$\sup\left\{E\psi^2 + \{E(\psi a)\}^2 - \left(E\widetilde{\mathbf{l}}_\nu^2 + \{E(\widetilde{\mathbf{l}}_\nu a)\}^2\right) : \ a \notin \dot{\mathcal{P}}, \ Ea^2 \le \kappa^2\right\}$$

$$\le \frac{1}{I_\nu} r \left\{1 + \kappa^2\left(1 + \frac{1}{r+1}\right)\right\}$$

$$\le B^2\{1 + 2\kappa^2\}, \qquad r \equiv E(\psi - \widetilde{l}_\nu)^2 / E\widetilde{l}_\nu^2 = I_\nu B^2.$$

5. An answer to question 2

- **"Proposition 2.F"** (F for false! or first attempt)

$$\sup\left\{E\psi^2 + \{E(\psi a)\}^2 - \left(E\tilde{\mathbf{l}}_\nu^2 + \{E(\tilde{\mathbf{l}}_\nu a)\}^2\right) : \ a \notin \dot{\mathcal{P}}, \ Ea^2 \le \kappa^2\right\}$$
$$\le \frac{1}{I_\nu} r \left\{1 + \kappa^2 \left(1 + \frac{1}{r+1}\right)\right\}$$
$$\le B^2\{1 + 2\kappa^2\}, \qquad r \equiv E(\psi - \tilde{l}_\nu)^2 / E\tilde{l}_\nu^2 = I_\nu B^2.$$

- **"Proof."** Since $E\psi^2 = E\tilde{\mathbf{l}}_\nu^2 + E(\psi - \tilde{\mathbf{l}}_\nu)^2$,

$$E\psi^2 + \{E(\psi a)\}^2 - \left(E\tilde{\mathbf{l}}_\nu^2 + \{E(\tilde{\mathbf{l}}_\nu a)\}^2\right)$$
$$= E(\psi - \tilde{\mathbf{l}}_\nu)^2 + \{E(\psi a)\}^2 - \{E(\tilde{\mathbf{l}}_\nu a)\}^2$$
$$\le E(\psi - \tilde{\mathbf{l}}_\nu)^2 + E(\psi^2)E(a^2) - \{E(\tilde{\mathbf{l}}_\nu a)\}^2$$

where the inequality follows by Cauchy-Schwarz,

- **"Proof", cont'd:** and equality holds if $a = C\psi$ for some C. Taking $a = C\psi$, the right side in the last display becomes

$$E(\psi - \tilde{\mathbf{l}}_\nu)^2 + E(\psi^2)E(a^2) - \{E(\tilde{\mathbf{l}}_\nu a)\}^2$$

$$= E(\psi - \tilde{\mathbf{l}}_\nu)^2 + C^2 \left\{ \{E(\psi^2)\}^2 - \{E(\tilde{\mathbf{l}}_\nu \psi)\}^2 \right\}$$

$$= E(\psi - \tilde{\mathbf{l}}_\nu)^2 + C^2 \left\{ \{E(\psi^2)\}^2 - \{E(\tilde{\mathbf{l}}_\nu^2)\}^2 \right\}$$

since $\psi - \tilde{\mathbf{l}}_\nu \perp \tilde{\mathbf{l}}_\nu$

$$\leq E(\psi - \tilde{\mathbf{l}}_\nu)^2 + \frac{\kappa^2}{E\psi^2} \left\{ \{E(\psi^2)\}^2 - \{E(\tilde{\mathbf{l}}_\nu^2)\}^2 \right\}$$

since $Ea^2 = C^2 E\psi^2 \leq \kappa^2$ if $C^2 \leq \kappa^2 / E\psi^2$

$$= E(\psi - \tilde{\mathbf{l}}_\nu)^2 + \frac{\kappa^2}{E\psi^2} \left\{ E(\psi^2) - E(\tilde{\mathbf{l}}_\nu^2) \right\} \left\{ E(\psi^2) + E(\tilde{\mathbf{l}}_\nu^2) \right\}$$

$$= E(\psi - \tilde{\mathbf{l}}_\nu)^2 \left\{ 1 + \kappa^2 \left\{ 1 + \frac{E(\tilde{\mathbf{l}}_\nu^2)}{E(\psi^2)} \right\} \right\} .$$

- **Proposition 2.** An answer to question 2

$$\sup\left\{E\psi^2 + \{E(\psi a)\}^2 - \left(E\widetilde{\mathbf{l}}_\nu^2 + \{E(\widetilde{\mathbf{l}}_\nu a)\}^2\right) :\ a \notin \dot{\mathcal{P}},\ Ea^2 \le \kappa^2\right\}$$

$$= \frac{1}{I_\nu} r \left\{1 + \kappa^2 \frac{1 + (2+r)s_+}{2 + r s_+}\right\}$$

$$\sim \frac{1}{I_\nu} r \left\{1 + \kappa^2 \frac{2+r}{1+r}\right\}, \quad \text{as } r \to \infty$$

where

$$s_+ \equiv s_+(r) \equiv \sqrt{\frac{1}{4} + \frac{1}{r}} + \frac{1}{2}.$$

- **Proposition 3.** An answer to question 3

$$\sup\left\{E\widetilde{\mathbf{l}}_\nu^2 + \{E(\widetilde{\mathbf{l}}_\nu a)\}^2 - \left(E\psi^2 + \{E(\psi a)\}^2\right) : \ a \notin \dot{\mathcal{P}}, \ Ea^2 \le \kappa^2\right\}$$
$$= \frac{r}{I_\nu}\left\{\kappa^2 \frac{(2+r)s_- - 1}{2 - rs_-} - 1\right\} > 0$$

if

$$\kappa^2 > \frac{2 - rs_-}{(2+r)s_- - 1} > 0$$

where

$$s_- \equiv s_-(r) \equiv \sqrt{\frac{1}{4} + \frac{1}{r}} - \frac{1}{2}.$$

- **Sketch of Proofs for Propositions 2 and 3.**
  First consider the case that

$$a = \tilde{a} = c\frac{\psi - \widetilde{l}_\nu}{E(\psi - \widetilde{l}_\nu)^2} + d\frac{\widetilde{l}_\nu}{E(\widetilde{l}_\nu^2)}.$$

In this case we seek to maximize (Proposition 2) or minimize (Proposition 3)

$$c^2 + 2cd$$

subject to the constraint

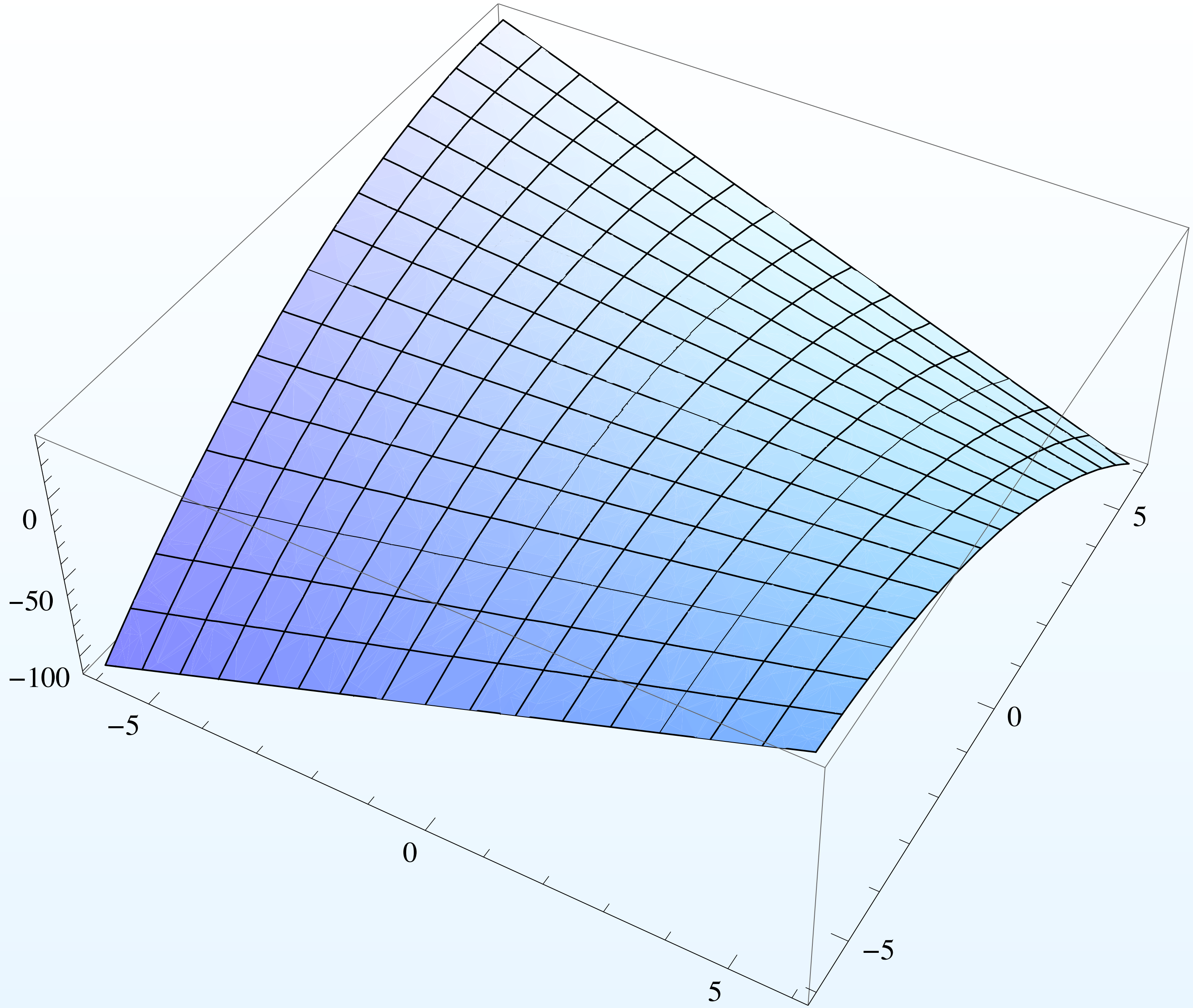
$$\begin{aligned} \kappa^2 \geq E(\tilde{a}^2) &= \frac{c^2}{E(\psi - \widetilde{l}_\nu)^2} + \frac{d^2}{E(\widetilde{l}_\nu^2)} \\ &= \frac{c^2}{B^2} + I_\nu d^2 = \frac{1}{B^2}(c^2 + rd^2). \end{aligned}$$

Carry this out by maximizing (or minimizing)

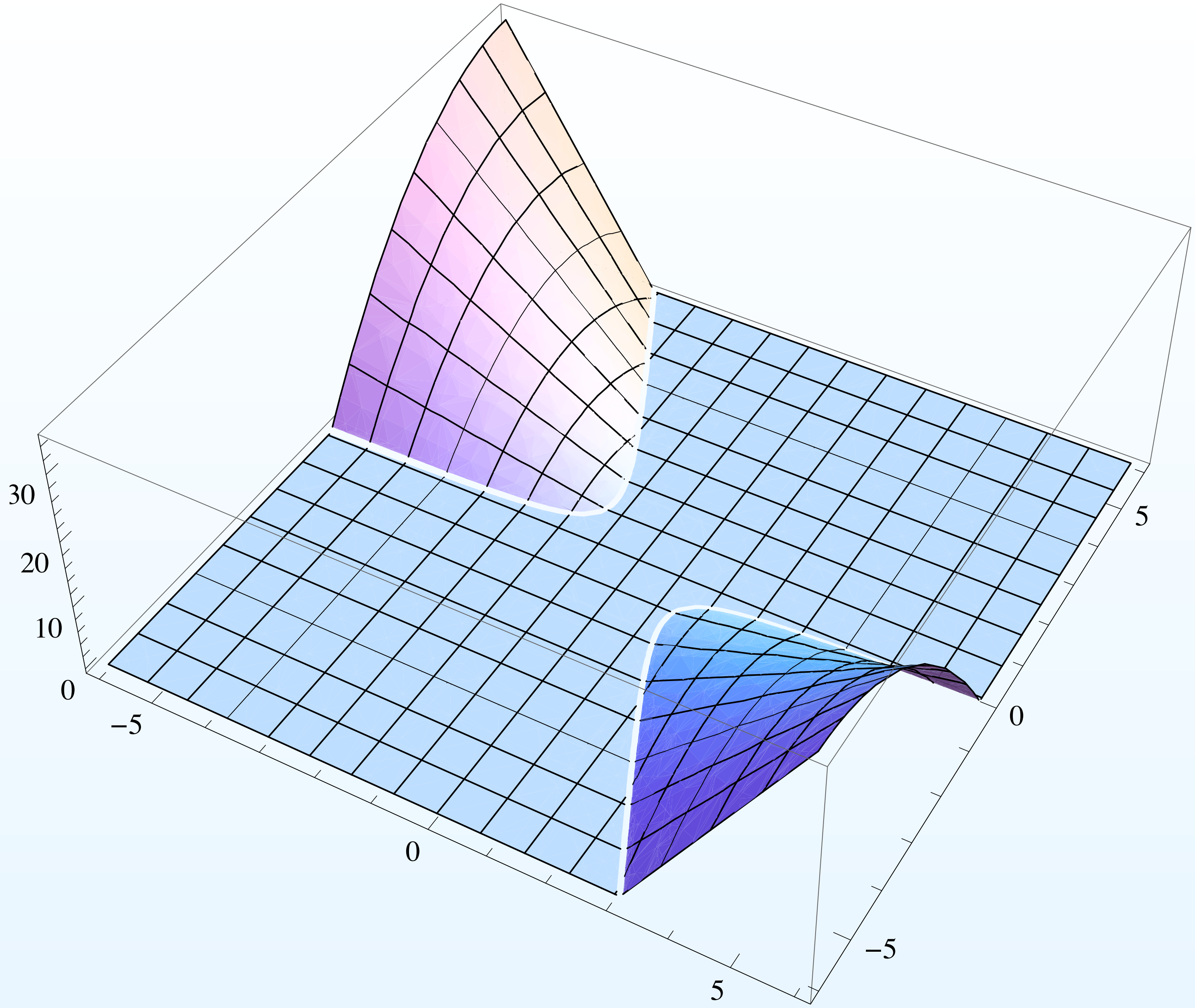
$$f(c, d, \lambda) = c^2 + 2dc + \lambda \left(\kappa^2 - \frac{c^2}{B^2} - I_\nu d^2 \right).$$

Then show that the solution for general $a = \widetilde{a} + (a - \widetilde{a})$ is achieved when $a = \widetilde{a} \in [\psi, \widetilde{l_\nu}]$. $\square$

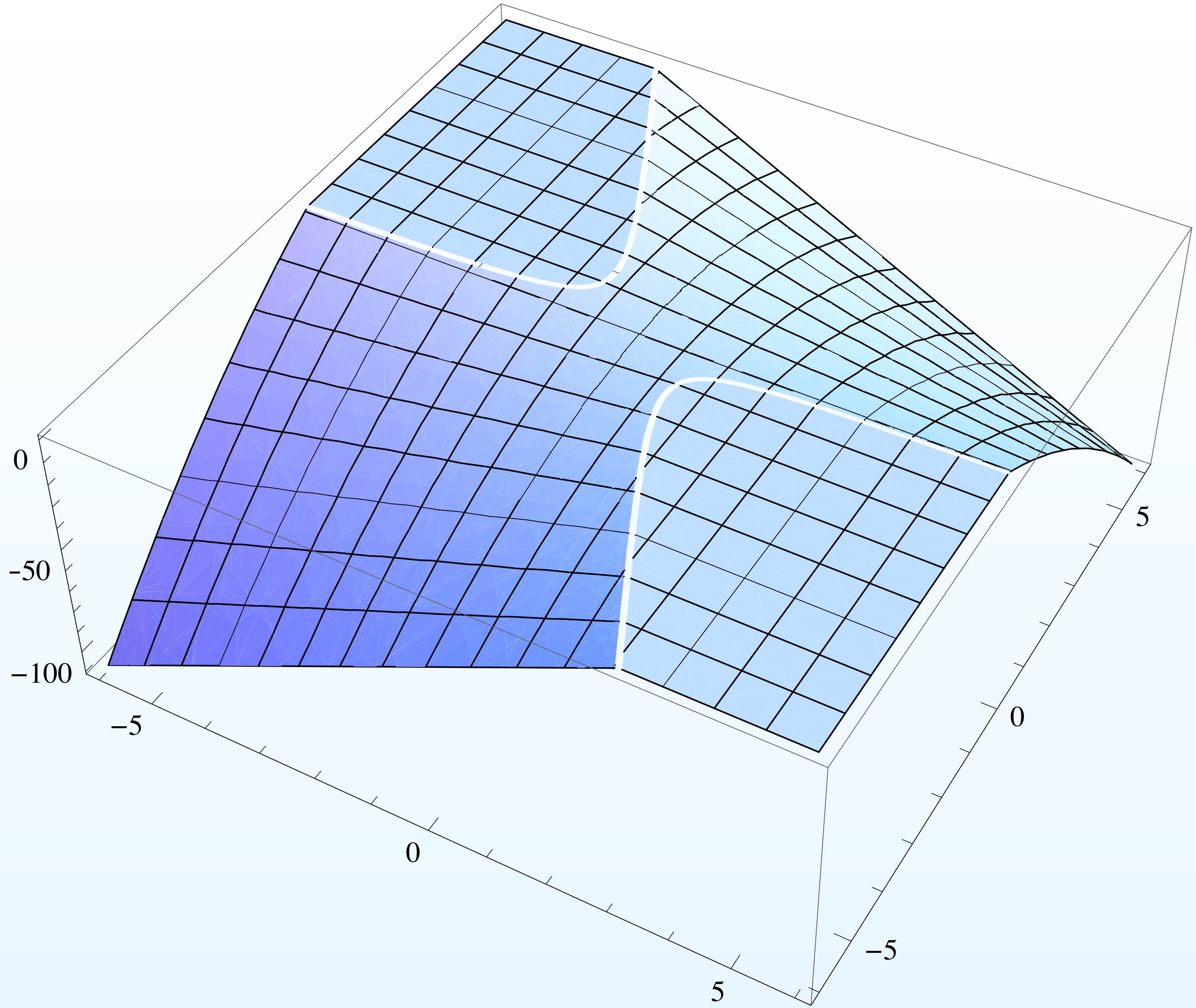
Geometric view?



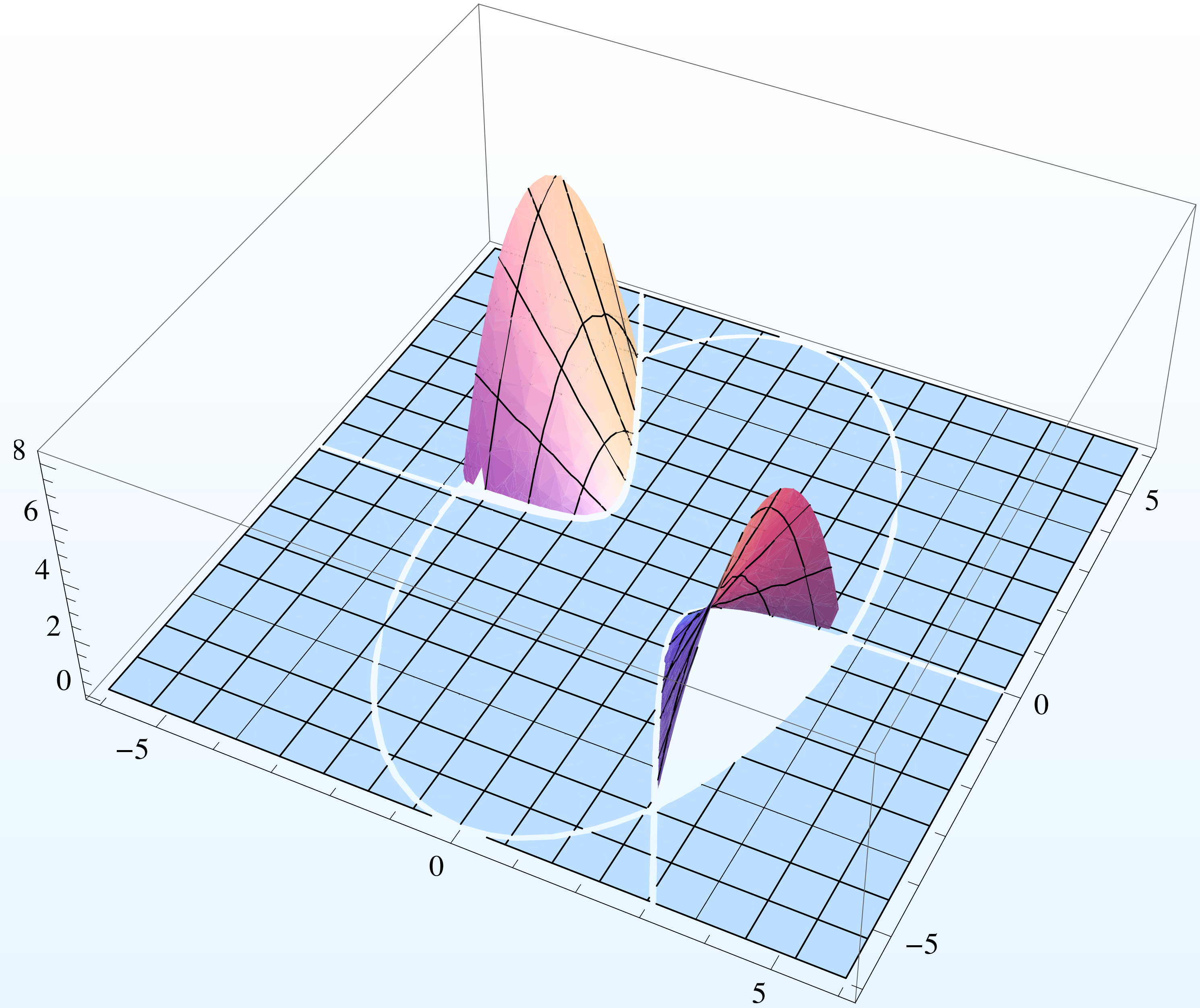
$$-(1 + c^2 + 2cd)$$



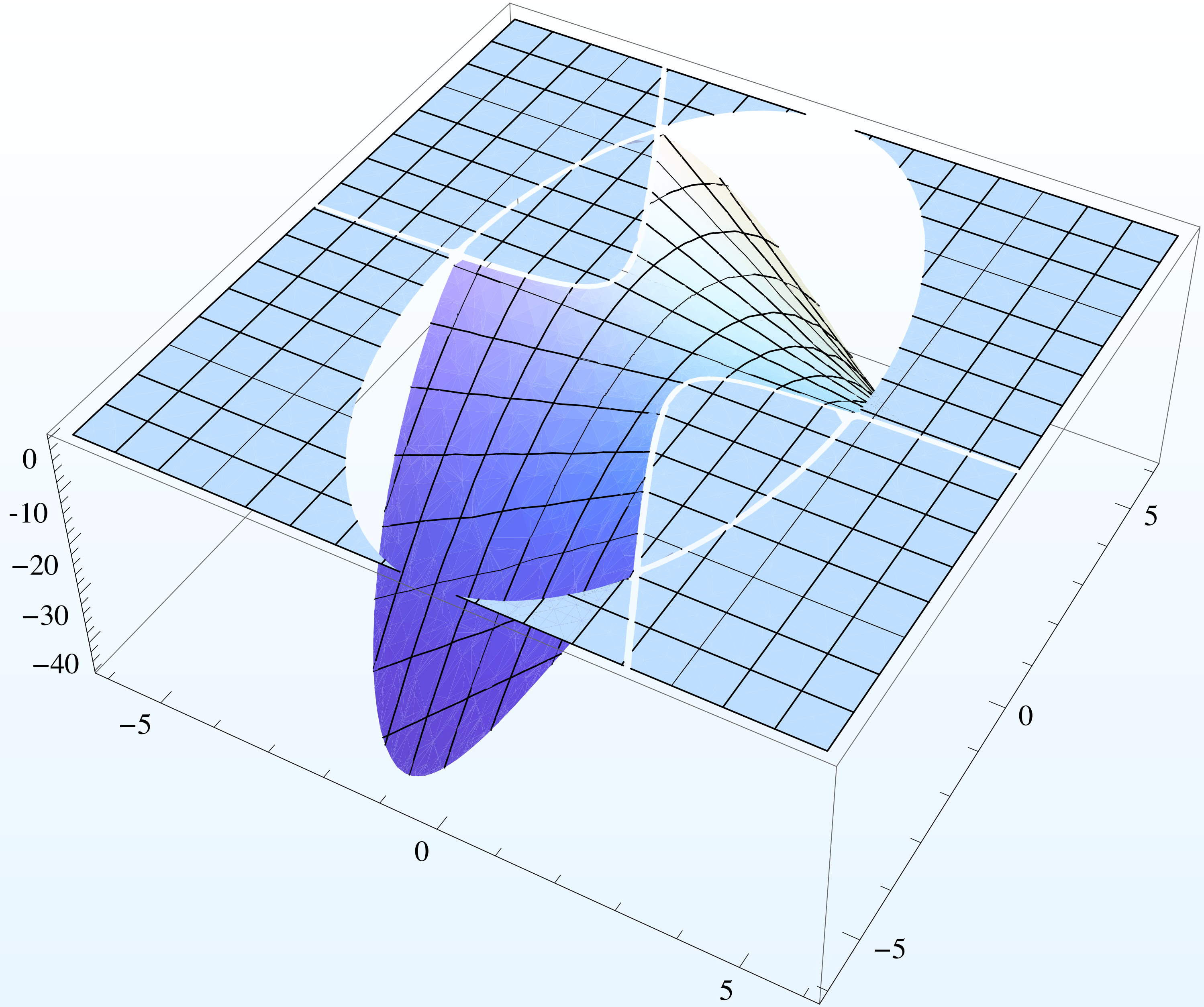
Positive part of $-(1 + c^2 + 2cd)$



Negative part of $-(1 + c^2 + 2cd)$



Positive part of $-(1 + c^2 + 2cd)$ restricted to $\kappa^2 \geq (c^2 + rd^2)/B^2$; $r = 3$, $\kappa^2 = 6$



Negative part of $-(1 + c^2 + 2cd)$ restricted to $\kappa^2 \geq (c^2 + rd^2)/B^2$; $r = 3$, $\kappa^2 = 6$

6. Examples

Example 1. Symmetric location model

- **Model:**

$$\mathcal{P} = \{P_{\nu,f} : \frac{dP_{\nu,f}}{d\lambda}(x) = f(x-\nu),\ \nu \in \mathbb{R},\ f \text{ symmetric at } 0, I_f < \infty\}.$$

6. Examples

Example 1. Symmetric location model

- **Model:**

$$\mathcal{P} = \{P_{\nu,f} : \frac{dP_{\nu,f}}{d\lambda}(x) = f(x-\nu),\ \nu \in \mathbb{R},\ f \text{ symmetric at } 0, I_f < \infty\}.$$

- **Tangent space:** At $P_0 = P_{0,f_0}$ the tangent space is

$$\dot{\mathcal{P}} = \{[-f_0'/f_0] + \text{all even functions } h \in L_2(P_0)\}.$$

6. Examples

Example 1. Symmetric location model

- **Model:**

$$\mathcal{P} = \{P_{\nu,f} : \frac{dP_{\nu,f}}{d\lambda}(x) = f(x-\nu),\ \nu \in \mathbb{R},\ f \text{ symmetric at } 0, I_f < \infty\}.$$

- **Tangent space:** At $P_0 = P_{0,f_0}$ the tangent space is

$$\dot{\mathcal{P}} = \{[-f_0'/f_0] + \text{all even functions } h \in L_2(P_0)\}.$$

- **Inefficient estimator:** Let $\hat{\nu}_n$ be the M-estimator corresponding to the logistic density; i.e. the solution of

$$\mathbb{P}_n \tilde{\psi}(X-\nu) = 0,$$

where g is the logistic density, $g(x) = e^{-x}/(1+e^{-x})^2$, and

$$\tilde{\psi}(x) = -\frac{g'(x)}{g(x)} = \frac{1-e^{-x}}{1+e^{-x}}.$$

- **Influence function ψ:** $\hat{\nu}_n$ is asymptotically linear with influence function

$$\psi(x) = \frac{\tilde{\psi}(x)}{\tilde{\Psi}'(0)}, \qquad \text{where} \quad \tilde{\Psi}(\nu) = P_0 \tilde{\psi}(X - \nu),$$

- **Influence function ψ:** $\hat{\nu}_n$ is asymptotically linear with influence function

$$\psi(x) = \frac{\tilde{\psi}(x)}{\tilde{\Psi}'(0)}, \qquad \text{where} \quad \tilde{\Psi}(\nu) = P_0 \tilde{\psi}(X - \nu),$$

- **Efficient influence function:**

$$\widetilde{\mathbf{l}}_\nu(x) = -\frac{1}{I_{f_0}} \cdot \frac{f_0'(x)}{f_0(x)}$$

where $I_f = \int (f'(x)/f(x))^2 f(x) dx$.

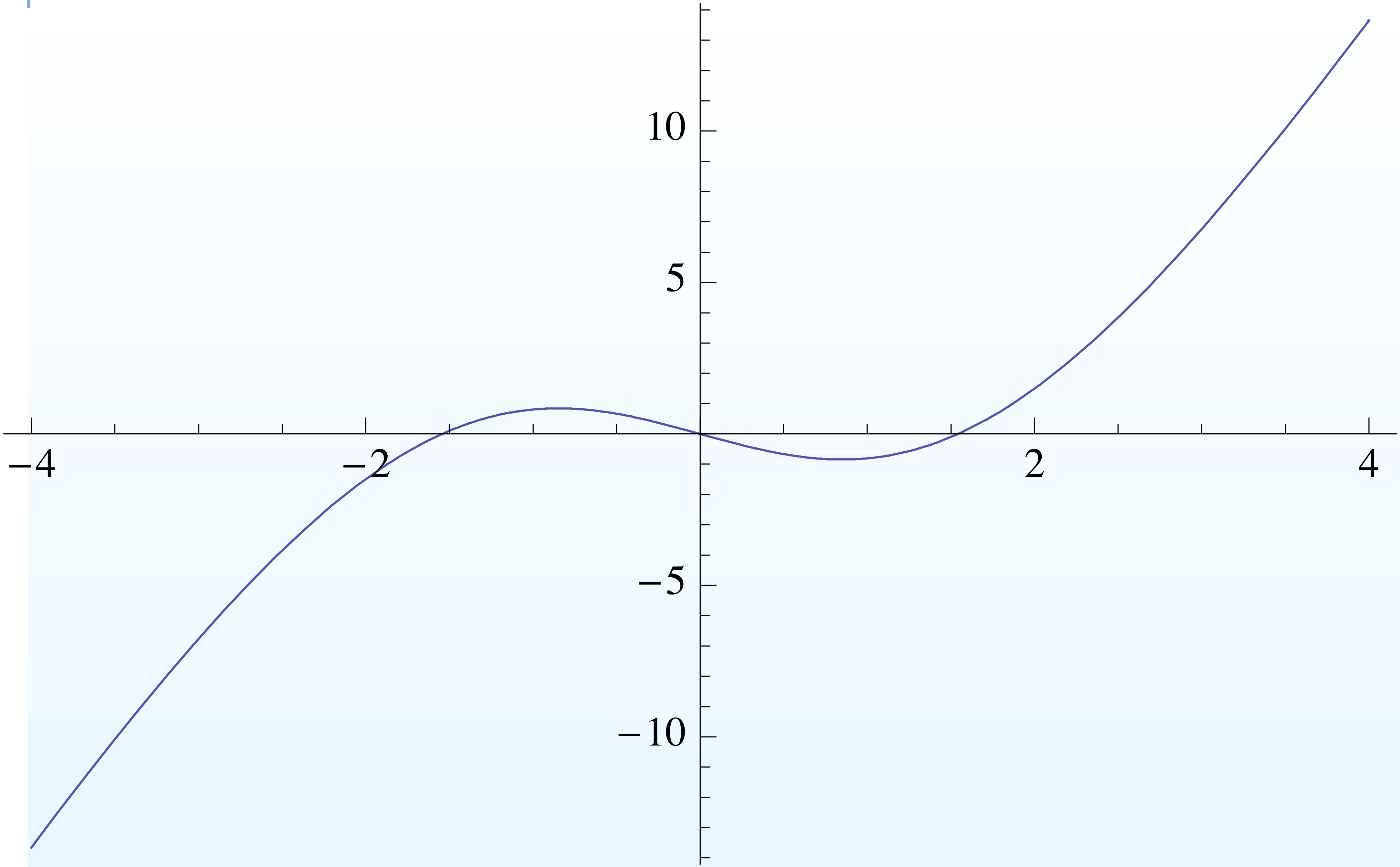
- For example, suppose $f_0 = \phi$, the standard Normal density given by $\phi(x) = (2\pi)^{-1/2}\exp(-x^2/2)$. Then
  - $\widetilde{\mathbf{l}}_\nu(x) = x$
  - $E(\widetilde{\mathbf{l}}_\nu^2) = 1$
  - $E(\psi - \widetilde{\mathbf{l}}_\nu)^2 \equiv B^2 = .01609...$
  - $c = -d_0 = -B = -\sqrt{.01609...} = -.126846$ yields $\tilde{a}_0$ given by

$$
\begin{aligned}
\tilde{a}_0(x) &= -d_0 \frac{\psi(x) - \widetilde{\mathbf{l}}_\nu(x)}{E(\psi - \widetilde{\mathbf{l}}_\nu)^2} + d_0 \frac{\widetilde{\mathbf{l}}_\nu(x)}{E(\widetilde{\mathbf{l}}_\nu^2)} \\
&= -(\psi(x) - x) + d_0 x.
\end{aligned}
$$

  -

$$AMSE_{\hat{\nu}}(a_0) = E(\psi^2) + \{E(\psi a_0)\}^2 = E\psi^2 = 1.01609...$$

$$AMSE_{\hat{\nu}\text{eff}}(a_0) = E(\widetilde{\mathbf{l}}_\nu^2) + \{E(\widetilde{\mathbf{l}}_\nu a_0)\}^2 = 1 + d_0^2 = 1.01609....$$



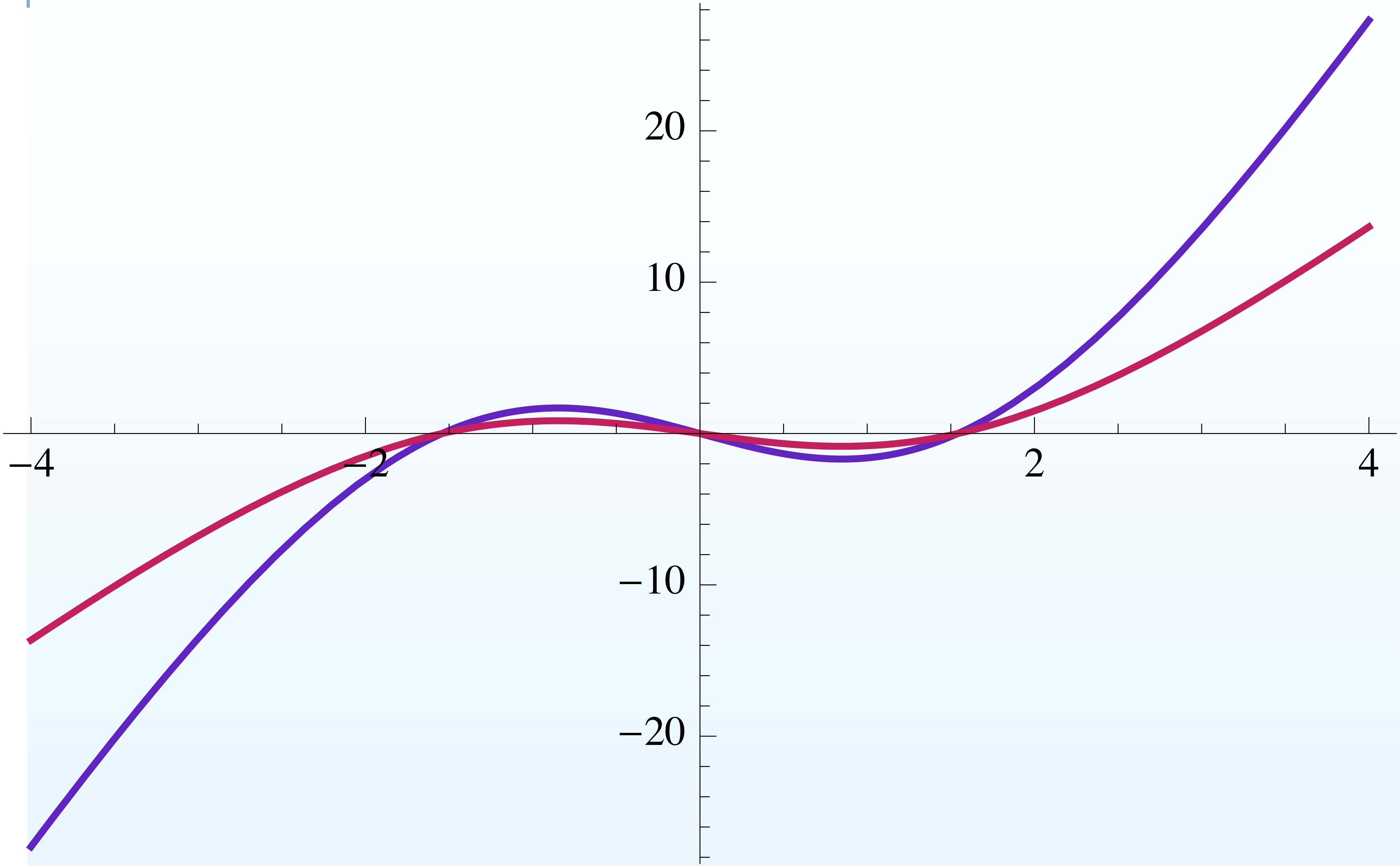
$\widetilde{a}_0$, symmetric location model, $f_0 = \phi$

- For example, suppose $f_0 = \phi$, the standard Normal density given by $\phi(x) = (2\pi)^{-1/2} \exp(-x^2/2)$. Then
  - $\widetilde{\mathbf{l}}_\nu(x) = x$, $E(\widetilde{\mathbf{l}}_\nu^2) = 1$, $E(\psi - \widetilde{\mathbf{l}}_\nu)^2 \equiv B^2 = .01609...$
  - If $d^2 > d_0^2 \equiv B^2$, and $c = -d$, then $\tilde{a}_{-d,d}$ is given by

$$\begin{aligned} \tilde{a}_{-d,d}(x) &= -d\frac{\psi(x) - \widetilde{\mathbf{l}}_\nu(x)}{E(\psi - \widetilde{\mathbf{l}}_\nu)^2} + d\frac{\widetilde{\mathbf{l}}_\nu(x)}{E(\widetilde{\mathbf{l}}_\nu^2)} \\ &= -(\psi(x) - x) + d_0 x. \end{aligned}$$

  -

$$\begin{aligned} AMSE_{\hat{\nu}}(a_{-d,d}) &= E(\psi^2) + \{0\}^2 \\ &= E\psi^2 = 1 + d_0^2 = 1.01609... \\ AMSE_{\hat{\nu}\text{eff}}(a_{-d,d}) &= E(\widetilde{\mathbf{l}}_\nu^2) + \{d\}^2 \\ &= 1 + d^2 > 1 + d_0^2 = 1.01609.... \end{aligned}$$



$\widetilde{a}_0$ (red), $\widetilde{a}_{-d,d}$ (blue), $d^2 = 2B^2$, symmetric location model, $f_0 = \phi$

Example 2. Paired exponential mixture model

- **Model:**

$$\mathcal{P} = \{P_{\nu,G} : \ \frac{dP_{(\nu,G)}}{d\lambda} = p_{(\nu,G)} \ : \ \nu > 0, \ G \text{ a d.f.}\},$$

where

$$p_{(\nu,G)}(x,y) = \int_0^\infty \lambda \exp(-\lambda x) \cdot \nu\lambda \exp(-\nu\lambda y) dG(\lambda)$$

Example 2. Paired exponential mixture model

- **Model:**

$$\mathcal{P} = \{P_{\nu,G} : \quad \frac{dP_{(\nu,G)}}{d\lambda} = p_{(\nu,G)} : \ \nu > 0, \ G \text{ a d.f.}\},$$

where

$$p_{(\nu,G)}(x,y) = \int_0^\infty \lambda \exp(-\lambda x) \cdot \nu\lambda \exp(-\nu\lambda y) dG(\lambda)$$

- **Inefficient estimator:** Let $\hat{\nu}_n$ to be the MLE based on the (pairwise) maximal invariants, (X_i/Y_i), $i = 1, \ldots, n$.

Example 2. Paired exponential mixture model

- **Model:**

$$\mathcal{P} = \{P_{\nu,G} : \frac{dP_{(\nu,G)}}{d\lambda} = p_{(\nu,G)} : \nu > 0, \ G \text{ a d.f.}\},$$

where

$$p_{(\nu,G)}(x,y) = \int_0^\infty \lambda \exp(-\lambda x) \cdot \nu\lambda \exp(-\nu\lambda y) dG(\lambda)$$

- **Inefficient estimator:** Let $\hat{\nu}_n$ to be the MLE based on the (pairwise) maximal invariants, (X_i/Y_i), $i = 1, \ldots, n$.
- **Influence function** ψ**:** $\hat{\nu}_n$ is asymptotically linear with influence function ψ given by

$$\psi(x,y) \quad = \quad 3\nu \left\{ \frac{x - \nu y}{x + \nu y} \right\} \equiv 3\nu^2 \mathbf{i}_\nu^{\text{inv}}(x/y)$$

- $E\psi^2(X,Y) = 3\nu^2$

- $E\psi^2(X,Y) = 3\nu^2$
- **Efficient score function for ν:**

$$\mathbf{l}^*_\nu(x,y) \quad = \quad \frac{1}{2\nu}\left(\frac{x-\nu y}{x+\nu y}\right)\left\{2 - 1 - s\frac{p'_S}{p_S}(s)\right\}$$

where p_S is the density of $S \equiv (X + \nu Y)$ given by

$$p_S(s,G) = s\int_0^\infty \lambda^2 e^{-\lambda s} dG(\lambda).$$

- $E\psi^2(X,Y) = 3\nu^2$
- **Efficient score function for ν:**

$$\mathbf{l}^*_\nu(x,y) \quad = \quad \frac{1}{2\nu}\left(\frac{x-\nu y}{x+\nu y}\right)\left\{2 - 1 - s\frac{p'_S}{p_S}(s)\right\}$$

where p_S is the density of $S \equiv (X+\nu Y)$ given by

$$p_S(s,G) = s\int_0^\infty \lambda^2 e^{-\lambda s} dG(\lambda).$$

- **Information for ν in $\mathcal{P}$:**

$$E(\mathbf{l}^*_\nu(X,Y)^2) = \frac{1}{3\nu^2} + \frac{1}{12\nu^2} I_{scale}(p_S).$$

- $E\psi^2(X,Y) = 3\nu^2$
- **Efficient score function for ν:**

$$\mathbf{l}^*_\nu(x,y) = \frac{1}{2\nu}\left(\frac{x-\nu y}{x+\nu y}\right)\left\{2 - 1 - s\frac{p'_S}{p_S}(s)\right\}$$

where p_S is the density of $S \equiv (X+\nu Y)$ given by

$$p_S(s,G) = s\int_0^\infty \lambda^2 e^{-\lambda s}\,dG(\lambda).$$

- **Information for ν in $\mathcal{P}$:**

$$E(\mathbf{l}^*_\nu(X,Y)^2) = \frac{1}{3\nu^2} + \frac{1}{12\nu^2} I_{scale}(p_S).$$

- **Variance bound:**

$$E(\widetilde{\mathbf{l}}_\nu)^2 = \left(\frac{1}{3\nu^2} + \frac{1}{12\nu^2} I_{scale}(p_S)\right)^{-1} = 3\nu^2\left(1 + \frac{1}{4} I_{scale}(p_S)\right)^{-1}.$$

- **Inefficiency of $\widehat{\nu}$:**

$$
\begin{aligned}
B^2 &= E(\psi - \widetilde{\mathbf{l}}_\nu)^2 = E\psi^2 - E\widetilde{l}_\nu^2 \\
&= 3\nu^2 \left\{ 1 - \frac{1}{1 + I_{scale}(p_S)/4} \right\} \\
&= 3\nu^2 \left\{ \frac{I_{scale}(p_S)/4}{1 + I_{scale}(p_S)/4} \right\} \\
&\equiv 3\nu^2 \left\{ \frac{b}{1+b} \right\}.
\end{aligned}
$$

- **Inefficiency of $\widehat{\nu}$:**

$$
\begin{aligned}
B^2 &= E(\psi - \widetilde{\mathbf{l}}_\nu)^2 = E\psi^2 - E\widetilde{l}_\nu^2 \\
&= 3\nu^2 \left\{ 1 - \frac{1}{1 + I_{scale}(p_S)/4} \right\} \\
&= 3\nu^2 \left\{ \frac{I_{scale}(p_S)/4}{1 + I_{scale}(p_S)/4} \right\} \\
&\equiv 3\nu^2 \left\{ \frac{b}{1+b} \right\}.
\end{aligned}
$$

- $\widehat{\nu}$ is asymptotically more efficient than $\widehat{\nu}^{\text{eff}}$ along $\{Q_n\}$ when $d > d_0 = B$ and $a = a_{-d,d}$:

$$
a_{-d,d}(x,y) = \frac{d}{2\nu} \frac{1+b}{b} \left(\frac{x - \nu y}{x + \nu y} \right) \mathbf{i}_{scale}(s).
$$