

Robust D-Optimal Design for Generalized Linear Models

Convex-Concave Structure and Efficient Algorithms

Jacob M. Aguirre

Joint work with Dmitrii M. Ostrovskii

School of ISyE, Georgia Tech

2026

Goal

- **Question:** In a GLM, the Fisher information depends on the parameter θ . When is the robust design problem $\max_{\lambda} \min_{\theta \in K} \log \det H(\lambda, \theta)$ convex–concave, and can interior-point methods solve it?
- **Answer:** It is convex–concave for every design **if and only if $\log h$ is convex**, where $h = \varphi''$ is the scalar Fisher curvature. A quantitative strengthening makes one homotopy **self-concordant convex–concave for all $t \geq 0$** , and path following certifies an ε -saddle point.
- **Mechanism:** Cauchy–Binet writes $\log \det H$ as a **log-partition function over bases**; its Hessian in θ is an average of the scalar curvatures $(\log h)''$ plus a covariance.

Outline

1. **Robust D-optimal design:** GLMs, the robust criterion, square designs.
2. **Convex–concave structure:** log-convexity of h , Cauchy–Binet, leverage scores.
3. **Scalar conditions:** the q -VB condition, Pearson's inequality, infinite divisibility.
4. **Path following:** compatibility, the uniform homotopy, global complexity.

Approximate D-optimal design

- **Candidate experiments:** $x_1, \dots, x_N \in \mathbb{R}^d$, the rows of $X \in \mathbb{R}^{N \times d}$, with $\text{rank } X = d$.
- **Approximate design:** weights $\lambda \in \bar{\Delta}$, the probability simplex in $\mathbb{R}^N$; λ_i is the fraction of runs at x_i . We write Δ for its relative interior.
- **Linear model** $y = x^\top \theta + \text{noise}$: the information matrix $\sum_i \lambda_i x_i x_i^\top$ does not depend on θ .
- **D-optimal design** (Kiefer & Wolfowitz 1960; Kiefer 1974):

$$\max_{\lambda \in \bar{\Delta}} \log \det \left(\sum_{i=1}^N \lambda_i x_i x_i^\top \right),$$

a concave problem; it minimizes the volume of the confidence ellipsoid for θ .

GLMs: the information depends on θ

- **Canonical GLM:** the response at x follows a natural exponential family with log-partition function φ and natural parameter $x^\top \theta$.

- **Information matrix:**

$$H(\lambda, \theta) := \sum_{i=1}^N \lambda_i h(x_i^\top \theta) x_i x_i^\top, \quad h := \varphi'' > 0 \text{ (scalar Fisher curvature).}$$

- **Examples:**

- Poisson: $h(t) = e^t$;
 - Bernoulli (logistic regression): $h(t) = [4 \cosh^2(t/2)]^{-1}$;
 - exponential, with rate $t > 0$: $h(t) = t^{-2}$.
- A design optimal at one θ need not stay optimal at another (Ford, Torsney & Wu 1992; Chaloner & Verdinelli 1995; Khuri et al. 2006).

The robust criterion

- **Uncertainty set:** a compact convex K with nonempty interior in the natural parameter domain; $\Theta := \text{int } K$.
- **Robust D-optimal design:**

$$\sup_{\lambda \in \bar{\Delta}} \inf_{\theta \in K} \log \det H(\lambda, \theta).$$

- **Saddle form:** we minimize in λ and maximize in θ

$$\mathcal{F}(\lambda, \theta) := -\log \det H(\lambda, \theta).$$

- $\mathcal{F}(\cdot, \theta)$ is convex since H is affine in λ . **The difficulty is entirely in θ .**
- **Predictor range:** $\mathcal{T}_X(K) := \{x_i^\top \theta : \theta \in K, i = 1, \dots, N\}$. All scalar conditions are imposed there.

Square designs: a scalar test

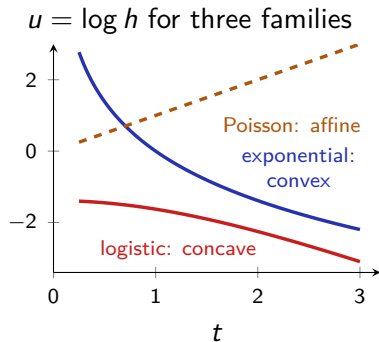
“Square designs reduce the question to the scalar function $\log h$.”

- If $N = d$ and X is invertible, then

$$\log \det H(\lambda, \theta) = \sum_{i=1}^d u(x_i^T \theta) + \text{const},$$

where $u := \log h$ and the constant does not depend on θ .

- Concavity of $\mathcal{F}(\lambda, \cdot)$ for all square designs forces u **convex**.
- Can redundancy, $N > d$, destroy this scalar characterization?



Main question

Question: Which scalar property of h makes the robust problem (a) convex–concave for every design, and (b) solvable by interior-point path following with constants that depend only on h ?

- **Robust GLM designs:** Sitter 1992; Dror & Steinberg 2006; Woods et al. 2006; Li, Kang & Deng 2022. Maximin efficiency: Dette 1997; Braess & Dette 2007.
- **Saddle-point interior-point methods:** self-concordant convex–concave (SCCC) functions, Nemirovski 1999.
- **Logarithmic–volumetric potentials:** Vaidya 1996; Anstreicher 1997; Jézéquel, Ostrovskii & Gaillard 2025; Tseng et al. 2025.

Outline

1. **Robust D-optimal design:** GLMs, the robust criterion, square designs.
2. **Convex–concave structure:** log-convexity of h , Cauchy–Binet, leverage scores.
3. **Scalar conditions:** the q -VB condition, Pearson's inequality, infinite divisibility.
4. **Path following:** compatibility, the uniform homotopy, global complexity.

Log-convexity characterizes convex–concavity

Theorem 1

Let J be an open interval with $h > 0$ on J , and $u := \log h$. The following statements are equivalent:

- (a) u is convex on J ;
- (b) $\mathcal{F}$ is convex–concave on $\Delta \times K$ for every full-rank X and every K with $\mathcal{T}_X(K) \subset J$.

- **Sufficiency holds design by design:** if u is convex on every interval $x_i^\top K$, then $\mathcal{F}(\lambda, \cdot)$ is concave on K .
- **Necessity is one-dimensional:** $N = d = 1$ and $x_1 = 1$ imply $\mathcal{F}(1, \theta) = -u(\theta)$.
- **Minimax interchange** (Corollary, via Sion 1958): under (a), $\inf_{\lambda} \sup_{\theta} \mathcal{F} = \sup_{\theta} \inf_{\lambda} \mathcal{F}$.

Proof: $\log \det H$ is a log-partition function

“Redundancy adds curvature in θ ; it cannot destroy convexity.”

- **Bases:** $\mathcal{B}_X$ is the set of d -subsets S of rows with $\det X_S \neq 0$.
- **Cauchy–Binet** and $h = e^u$ imply

$$\det H(\lambda, \theta) = \sum_{S \in \mathcal{B}_X} \det(X_S)^2 \prod_{i \in S} \lambda_i h(x_i^\top \theta) = \sum_{S \in \mathcal{B}_X} e^{w_S(\theta)},$$

$$w_S(\theta) := \log \left(\det(X_S)^2 \prod_{i \in S} \lambda_i \right) + \sum_{i \in S} u(x_i^\top \theta).$$

- **Basis law** $p_{\lambda, \theta}(S) \propto e^{w_S(\theta)}$. Differentiating twice (log-sum-exp), we obtain

$$\nabla_\theta^2 \log \det H(\lambda, \theta) = \mathbb{E}_{\lambda, \theta} \left[\sum_{i \in S} u''(x_i^\top \theta) x_i x_i^\top \right] + \text{Cov}_{\lambda, \theta} \left(\sum_{i \in S} u'(x_i^\top \theta) x_i \right).$$

- $u'' \geq 0$ makes the first term $\succcurlyeq 0$; the covariance is $\succcurlyeq 0$, and it vanishes for square designs.

Leverage scores and logistic regression

“The covariance term cannot rescue logistic regression.”

- **Leverage scores:** $\pi_i := \Pr[i \in S] = \lambda_i h(x_i^\top \theta) x_i^\top H^{-1} x_i$, $0 \leq \pi_i \leq 1$, $\sum_i \pi_i = d$.
- **Two-sided bound** (Corollary), with $\eta_i := x_i^\top \theta$:
$$\sum_i \pi_i u''(\eta_i) x_i x_i^\top \preceq \nabla_\theta^2 \log \det H \preceq \sum_i \pi_i \left(u''(\eta_i) + 2(1 - \pi_i) u'(\eta_i)^2 \right) x_i x_i^\top.$$
- **Logistic regression:** if $\mathcal{T}_X(K) \subset (-t_\star, t_\star)$, $t_\star := 2 \operatorname{arcsinh}(1/2) \approx 0.96$, then $u'' + 2(u')^2 < 0$ there, and $\mathcal{F}(\lambda, \cdot)$ is **strictly convex** for every λ and **every** N .

Outline

1. **Robust D-optimal design:** GLMs, the robust criterion, square designs.
2. **Convex–concave structure:** log-convexity of h , Cauchy–Binet, leverage scores.
3. **Scalar conditions:** the q -VB condition, Pearson's inequality, infinite divisibility.
4. **Path following:** compatibility, the uniform homotopy, global complexity.

The q -VB condition

Definition (q -VB condition; cf. Tseng et al. 2025)

For $q \geq 1$, h satisfies the q -VB condition on an interval J if $h''h \geq q(h')^2$ on J . It is *weak* if $q = 1$ and *strong* if $q > 1$.

- **Differential identity:** $h''h - q(h')^2 = h^2(u'' - (q-1)(u')^2)$. For $q = 1$, the weak condition is log-convexity.
- **Power form (Lemma):** for $q > 1$, the q -VB condition holds $\iff h^{1-q}$ is concave.
- **Examples:** $h = t^{-k}$ has $q_{\max} = 1 + 1/k$; Poisson has $q_{\max} = 1$; logistic fails for $q = 1$.
- **Strong conditions are local:** on all of $\mathbb{R}$ only a constant h satisfies one.

Two strengthened scalar assumptions

Assumptions on the predictor range $\mathcal{T}_X(K)$

(A1) $u'' \geq (q-1)(u')^2$ for some $q > 1$.

(A2) $|u'''| \leq 2M_\varphi (u'')^{3/2}$.

- **Different roles:** (A1) bounds the slope of u in its own metric, $|u'|/\sqrt{u''} \leq 1/\sqrt{q-1}$; (A2) bounds the variation of that metric.
- **Neither implies the other:**
 - normal plus exponential, $\varphi(t) = \frac{t^2}{2} - \log(1-t)$: $q = 3/2$, but no finite M_φ ;
 - Poisson: $M_\varphi = 0$, but $q_{\max} = 1$.
- **Power profiles** $h = t^{-k}$: $u = -k \log t$ is k times the log barrier of the half-line, and $q-1 = M_\varphi^2 = 1/k$.
- For simplicity, $M_\varphi \leq 1$ from now on. Then u is a standard self-concordant function with $(u')^2 \leq u''/(q-1)$.

Beyond log-convexity: Pearson's inequality

Proposition (Pearson 1916)

For every canonical family, $u'' \geq -2h$, with equality iff the centered response takes two values.

- **Proof** (Móri, Rohatgi & Székely 1993): the derivatives of h are cumulants of $Z := Y - \mathbb{E}Y$, and

$$(h')^2 = \text{Cov}(Z, Z^2)^2 \stackrel{(\text{Cauchy-Schwarz})}{\leq} \text{Var } Z \cdot \text{Var } Z^2 = h(h'' + 2h^2).$$

- **Curvature correction:** $\log \det H + 2 \sum_i \varphi(x_i^T \theta)$ is concave in λ and convex in θ ; the design problem at each fixed θ is unchanged.
- The coefficient 2 is **sharp** for square logistic designs, and it restores concavity **but not self-concordance**.

Infinite divisibility: u is a log-partition function

Lemma (Lévy representation)

The law at η_0 is infinitely divisible, with Gaussian variance σ^2 and Lévy measure Λ , iff

$$h(\eta) = \int e^{(\eta - \eta_0)y} \varrho(dy), \quad \varrho(dy) := \sigma^2 \delta_0(dy) + y^2 \Lambda(dy).$$

- **Auxiliary family** $\tilde{P}_\eta(dy) \propto e^{(\eta - \eta_0)y} \varrho(dy)$, with log-partition function u :

$$u' = \tilde{\mathbb{E}}_\eta Y, \quad u'' = \widetilde{\text{Var}}_\eta Y, \quad u''' = \tilde{\mathbb{E}}_\eta (Y - \tilde{\mathbb{E}}_\eta Y)^3.$$

- **(A1)** bounds the auxiliary mean by its standard deviation; **(A2)** bounds its standardized skewness.
- $u'' = \widetilde{\text{Var}}_\eta Y \geq 0$: **every infinitely divisible family is log-convex**, and Theorem 1 applies.

Sufficient conditions on the representing measure

Proposition

- (i) If ϱ has a log-concave density, then (A2) holds with $M_\varphi = 1$; if $h^{1/k}$ is the Laplace transform of a log-concave density, then with $M_\varphi = 1/\sqrt{k}$.
- (ii) If $\varrho(dy) = r(y)dy$ on $y > 0$ and $0 \leq -y^2(\log r)''(y) \leq \bar{\kappa}$, then (A1) holds with $q = 1 + \frac{1}{\bar{\kappa}+1}$, and (A2) with $M_\varphi = 1$.

- **Compound Poisson, exponential jumps:** $h(t) = 2(1-t)^{-3}$; (i) with $k = 3$ implies the sharp $M_\varphi = 1/\sqrt{3}$, although φ is not self-concordant.
- **(ii) is exact for power profiles**, and it reaches beyond them: $h(t) = [t(t+1)]^{-1}$ has $q_{\max} = 3/2$ and $M_\varphi = 1$.
- A log-concave law need not have a log-concave ϱ : for normal plus exponential, ϱ has an atom at 0.

Sharp constants for standard families

Family	$h(t)$	Domain	$q_{\max}$	M_{φ}
Gamma / exponential	$\propto t^{-2}$	$(0, \infty)$	$3/2$	$1/\sqrt{2}$
Inverse Gaussian	$\propto t^{-3/2}$	$(0, \infty)$	$5/3$	$\sqrt{2/3}$
Compound Poisson–Gamma	$\propto t^{-r-2}$	$(0, \infty)$	$1 + \frac{1}{r+2}$	$\frac{1}{\sqrt{r+2}}$
Spectrally positive stable, $1 < \alpha < 2$	$\propto t^{\alpha-2}$	$(0, \infty)$	$1 + \frac{1}{2-\alpha}$	$\frac{1}{\sqrt{2-\alpha}}$
Non-power infinitely divisible	$[t(t+1)]^{-1}$	$(0, \infty)$	$3/2$	1
Poisson	e^t	$\mathbb{R}$	1	0
Bernoulli	$[4 \cosh^2(t/2)]^{-1}$	$\mathbb{R}$	—	—

- The first four rows are power profiles t^{-k} : $q_{\max} = 1 + 1/k$ and $M_{\varphi} = 1/\sqrt{k}$.
- Poisson** is log-convex but fails (A1); **Bernoulli** fails log-convexity.

Outline

1. **Robust D-optimal design:** GLMs, the robust criterion, square designs.
2. **Convex–concave structure:** log-convexity of h , Cauchy–Binet, leverage scores.
3. **Scalar conditions:** the q -VB condition, Pearson's inequality, infinite divisibility.
4. **Path following:** compatibility, the uniform homotopy, global complexity.

Self-concordant convex-concave functions

- **Saddle metric** of a convex-concave f at $w = (x, y)$:

$$S_f(w) := \nabla_{xx}^2 f(w) \oplus \left(-\nabla_{yy}^2 f(w) \right), \quad \|\xi\|_{f,w}^2 := \xi^T S_f(w) \xi.$$

Definition (Nemirovski 1999)

f is M -SCCC if $|D^3 f(w)[\xi, \xi, \xi]| \leq 2M \|\xi\|_{f,w}^3$; *standard* if $M = 1$; an *SCCC barrier* if $f(\cdot, y)$ and $-f(x, \cdot)$ are barriers.

- **Closure:** αf is $(M/\sqrt{\alpha})$ -SCCC; $f + g$ is $\max\{M_f, M_g\}$ -SCCC.
- **Design barrier:** $B_\Delta(\lambda) := -\sum_i \log \lambda_i$, standard, with parameter N .
- **Target:** one constant γ such that $\gamma[t\mathcal{F} + B_\Delta - \mathcal{B}_\Theta]$ is standard SCCC **for every** $t \geq 0$, for a parameter barrier $\mathcal{B}_\Theta$.

The parameter barrier

- **Log-curvature potential:** $U_X(\theta) := \sum_i u(x_i^\top \theta)$, with $\nabla^2 U_X(\theta) = X^\top \text{diag}(u''(x_i^\top \theta))X$.

Proposition

Under (A1) and (A2), U_X is standard self-concordant, and for $\theta \in \Theta$, $\bar{\theta} \in K$,

$$|DU_X(\theta)[v]|^2 \leq \frac{N}{q-1} D^2 U_X(\theta)[v, v], \quad DU_X(\theta)[\bar{\theta} - \theta] \leq \frac{N}{q-1}.$$

- **Second bound:** tangent inequality of the concave $h^{1-q} = e^{-(q-1)u} > 0$ (power form).
- **Combined barrier:** with a self-concordant barrier G for Θ with parameter ν_G ,

$$\mathcal{B}_\Theta := U_X + G \text{ has parameter } \nu_\Theta := \frac{N}{q-1} + \nu_G.$$

Volumetric trace calculus

“The q -VB condition is a trace inequality in disguise.”

- $\mathcal{F}(\lambda, \cdot) = -\log \det \nabla^2 L_\lambda$ with $L_\lambda(\theta) := \sum_i \lambda_i \varphi(x_i^\top \theta)$: -2 times a **volumetric barrier**.
- **Pieces** $B_i := H^{-1/2}(\lambda_i h(\eta_i) x_i x_i^\top) H^{-1/2} \succcurlyeq 0$ sum to I . Along v , with $z_i := x_i^\top v$,

$$S := \sum_i u'(\eta_i) z_i B_i, \quad T := \sum_i \frac{h''}{h}(\eta_i) z_i^2 B_i, \quad U := \sum_i \frac{h'''}{h}(\eta_i) z_i^3 B_i$$

are the first three derivatives of $H^{-1/2} H(\lambda, \theta + sv) H^{-1/2}$ at $s = 0$.

Proposition

$-D_{\theta\theta}^2 \mathcal{F}[v, v] = \text{tr } T - \text{tr } S^2$ and $D_{\theta\theta\theta}^3 \mathcal{F}[v, v, v] = -\text{tr } U + 3 \text{tr}(TS) - 2 \text{tr } S^3$.
Under the q -VB condition, $\text{tr } T \geq q \text{tr } S^2$; hence $-D_{\theta\theta}^2 \mathcal{F}[v, v] \geq (q - 1) \text{tr } S^2$.

- **Idea:** $\frac{h''}{h} \geq q(u')^2$ termwise, and $\sum_i a_i^2 B_i \succcurlyeq (\sum_i a_i B_i)^2$ for all $a \in \mathbb{R}^N$ since $\sum_i B_i = I$.

Compatibility of $\mathcal{F}$ with the product barrier

“One power of the barrier norm survives the path weight.”

Proposition (compatibility)

Under (A1) and (A2), let $B := B_\Delta + B_\Theta$ and $\|\xi\|_{B,w}^2 := \xi^\top \nabla^2 B(w) \xi$. Then

$$|D^3 \mathcal{F}(w)[\xi, \xi, \xi]| \leq \beta_{\text{hom}} \|\xi\|_{\mathcal{F},w}^2 \|\xi\|_{B,w},$$

with an explicit β_{hom} that depends only on q and M_φ .

- **Why this form:** **squared** saddle norm of $\mathcal{F}$, **one power** of the barrier norm (Nemirovski 1999). Replacing $\mathcal{F}$ by $t\mathcal{F}$ multiplies both sides by t .
- By contrast, the self-concordance modulus of $\mathcal{F} + \tau B_\Delta - \rho U_X - \sigma G$ deteriorates as $\tau, \rho, \sigma \rightarrow 0$.

The uniform homotopy

Theorem 2

Under (A1) and (A2), let $\gamma_\varphi := \left(\frac{\beta_{\text{hom}}+2}{2}\right)^2$. For every $t \geq 0$,

$$f_t(\lambda, \theta) := \gamma_\varphi \left[t \mathcal{F}(\lambda, \theta) + B_\Delta(\lambda) - \mathcal{B}_\Theta(\theta) \right]$$

is a nondegenerate standard SCCC barrier on $\Delta \times \Theta$. Its saddle point (λ_t, θ_t) is unique, and $\lambda_0 = \frac{1}{N}\mathbf{1}$, $\theta_0 = \operatorname{argmin} \mathcal{B}_\Theta$.

- **One normalization:** γ_φ depends only on q and M_φ , not on t , N , d , X or K .
- **Proof:** compatibility and self-concordance of B imply

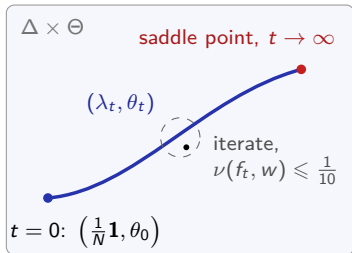
$$|D^3(f_t/\gamma_\varphi)[\xi, \xi, \xi]| \leq \beta_{\text{hom}} t \|\xi\|_{\mathcal{F}}^2 \|\xi\|_B + 2 \|\xi\|_B^3 \leq (\beta_{\text{hom}} + 2) \underbrace{\left(t \|\xi\|_{\mathcal{F}}^2 + \|\xi\|_B^2 \right)}_{= \|\xi\|_{f_t/\gamma_\varphi}^2}^{3/2}.$$

The saddle gap along the path

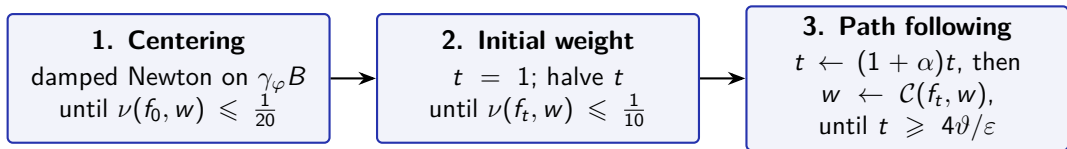
Theorem (gap at the center)

The saddle gap $\text{Gap}_{\mathcal{F}}(\lambda, \theta) := \sup_{\bar{\theta} \in K} \mathcal{F}(\lambda, \bar{\theta}) - \inf_{\bar{\lambda} \in \bar{\Delta}} \mathcal{F}(\bar{\lambda}, \theta)$ at the center of $\mathcal{F} + \tau B_{\Delta} - \rho U_X - \sigma G$ is at most $\tau N + \frac{\rho N}{q-1} + \sigma \nu_G$. Along the path, $\text{Gap}_{\mathcal{F}}(\lambda_t, \theta_t) \leq \frac{N + \nu_{\Theta}}{t}$.

- **Proof in λ :** optimality and convexity imply $\mathcal{F}(\lambda_c, \theta_c) - \mathcal{F}(\bar{\lambda}, \theta_c) \leq \tau DB_{\Delta}(\lambda_c)[\bar{\lambda} - \lambda_c] \leq \tau N$.
- **In θ :** $DU_X[\bar{\theta} - \theta] \leq \frac{N}{q-1}$, and the oscillation of U_X over K never enters.
- **Certificate** (Nemirovski 1999): proximity $\nu(f_t, w) := \|\nabla f_t(w)\|_{S_{f_t}(w)^{-1}} \leq \frac{1}{10}$ implies $\text{Gap}_{\mathcal{F}}(w) \leq \frac{4\vartheta}{t}$, $\vartheta := \max\{N, \nu_{\Theta}\}$.



The path-following method



- **Input:** $\varepsilon > 0$ and any $w^0 \in \Delta \times \Theta$. **Step:** $\alpha := 1/\sqrt{\gamma_\varphi \vartheta}$.
- **Corrector $\mathcal{C}$:** the Saddle Newton method (Nemirovski 1999), full Newton steps for f_t tilted by a linear term whose weight is driven to 0.

Theorem (correctness)

From every $w^0 \in \Delta \times \Theta$, the method is well defined, terminates, and returns (λ, θ) with $\text{Gap}_{\mathcal{F}}(\lambda, \theta) \leq \varepsilon$.

Global complexity

Theorem 3

Under (A1) and (A2), from any $w^0 \in \Delta \times \Theta$ the method returns a pair with $\text{Gap}_{\mathcal{F}} \leq \varepsilon$ after

$$O_{q,M_\varphi} \left((N + \nu_G) \log(1 + \mathcal{R}_{\text{in}}) + \sqrt{N + \nu_G} \left[1 + \log_+ \frac{(N + \nu_G) \sqrt{d}}{\varepsilon} \right] \right)$$

Newton steps, updates of t and proximity tests in total.

- **Centering** costs the first term, where $\mathcal{R}_{\text{in}}$ is the radius of $\Delta \times \Theta$ in the local norm of B at w^0 ; **path following** costs the second.
- Each Newton step costs $O((N + d)^3)$ arithmetic operations.
- If $\nu_G = O(d)$, the total is $O(N \log[(1 + \mathcal{R}_{\text{in}})(1 + N^{3/2}/\varepsilon)])$: **linear in N** .

Summary and open problems

- **Structure** (Thm. 1): convex–concave for every design $\iff \log h$ **convex**.
- **Families**: Pearson's correction with sharp coefficient 2; infinite divisibility turns (A1)–(A2) into moment inequalities.
- **Algorithm** (Thms. 2–3): uniform SCCC homotopy, certified ε -saddle point.
- **Open problems**:
 - corrected logistic: complexity independent of the predictor range;
 - fixed redundant designs: concavity in θ without convexity of $\log h$;
 - other criteria: A -optimality, Kiefer's Φ_p family.

Thank you!

References

- J. Kiefer, J. Wolfowitz. The equivalence of two extremum problems. *Canad. J. Math.* (1960).
- J. Kiefer. General equivalence theory for optimum designs (approximate theory). *Ann. Statist.* (1974).
- I. Ford, B. Torsney, C. F. J. Wu. The use of a canonical form in the construction of locally optimal designs for non-linear problems. *J. R. Stat. Soc. B* (1992).
- K. Chaloner, I. Verdinelli. Bayesian experimental design: a review. *Statist. Sci.* (1995).
- A. I. Khuri, B. Mukherjee, B. K. Sinha, M. Ghosh. Design issues for generalized linear models: a review. *Statist. Sci.* (2006).
- R. R. Sitter. Robust designs for binary data. *Biometrics* (1992).
- H. A. Dror, D. M. Steinberg. Robust experimental design for multivariate generalized linear models. *Technometrics* (2006).
- D. C. Woods, S. M. Lewis, J. A. Eccleston, K. G. Russell. Designs for generalized linear models with several variables and model uncertainty. *Technometrics* (2006).
- Y. Li, L. Kang, X. Deng. A maximin Φ_p -efficient design for multivariate generalized linear models. *Statist. Sinica* (2022).
- H. Dette. Designing experiments with respect to 'standardized' optimality criteria. *J. R. Stat. Soc. B* (1997).
- D. Braess, H. Dette. On the number of support points of maximin and Bayesian optimal designs. *Ann. Statist.* (2007).
- A. Nemirovski. On self-concordant convex-concave functions. *Optim. Methods Softw.* (1999).
- P. M. Vaidya. A new algorithm for minimizing convex functions over convex sets. *Math. Program.* (1996).
- K. M. Anstreicher. On Vaidya's volumetric cutting plane method for convex programming. *Math. Oper. Res.* (1997).
- R. Jézéquel, D. M. Ostrovskii, P. Gaillard. Efficient and near-optimal online portfolio selection. *Math. Oper. Res.* (2025).
- W.-F. Tseng, K.-C. Chen, Z.-H. Xiao, Y.-H. Li. Online learning of quantum states with logarithmic loss via VB-FTRL. *ALT* (2025).
- K. Pearson. Mathematical contributions to the theory of evolution. XIX. *Phil. Trans. R. Soc. A* (1916).
- T. F. Móri, V. K. Rohatgi, G. J. Székely. On multivariate skewness and kurtosis. *Theory Probab. Appl.* (1993).
- M. Sion. On general minimax theorems. *Pacific J. Math.* (1958).