Asymptotic Normality and Optimality of Stochastic Linearized Augmented Lagrangian Method

Jiajin Li

Sauder School of Business

University of British Columbia

SIAM Conference on Optimization, 2026

Joint work with Feng Ruan (Northwestern).

What this talk is about.

Understand the statistical behavior (asymptotic normality) of stochastic primal–dual algorithms for constrained optimization.

Warm-up: the classical SGD

- **Setup.** Consider a strongly convex and smooth stochastic optimization problem:

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x}) := \mathbb{E}[F(\boldsymbol{x},\xi)]. \tag{P0}$$

- **Algorithm.** The simplest method—stochastic gradient descent (SGD):

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k - \alpha_k \nabla F(\boldsymbol{x}^k, \xi^k).$$

Here α_k is the stepsize and ξ^k denotes an i.i.d. sample.

Warm-up: the classical SGD

- **Setup.** Consider a strongly convex and smooth stochastic optimization problem:

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x}) := \mathbb{E}[F(\boldsymbol{x},\xi)]. \tag{P0}$$

- **Algorithm.** The simplest method—stochastic gradient descent (SGD):

$$\boxed{\boldsymbol{x}^{k+1} = \boldsymbol{x}^k - \alpha_k \nabla F(\boldsymbol{x}^k, \xi^k)}.$$

Here α_k is the stepsize and ξ^k denotes an i.i.d. sample.

Statistical guarantee for SGD

Theorem (Ruppert (1988); Polyak–Juditsky (1992))

If $\alpha_k \propto k^{-\theta}$ *with* $\theta \in (\frac{1}{2}, 1)$ *and standard noise conditions hold, then*

$$\sqrt{k}(\overline{\boldsymbol{x}}_k - \boldsymbol{x}^\star) \xrightarrow{d} \mathcal{N}\Big(0,\ \nabla^2 f(\boldsymbol{x}^\star)^{-1}\, \Sigma\, \nabla^2 f(\boldsymbol{x}^\star)^{-1}\Big),$$

where $\overline{\boldsymbol{x}}_k := \frac{1}{k}\sum_{i=1}^{k} \boldsymbol{x}^i$, *and* $\Sigma := \mathrm{Cov}\big(\nabla F(\boldsymbol{x}, \xi)\big)$.

- Matches the covariance of SAA (sample-average approximation) $\Rightarrow$ same statistical efficiency.
- *Asymptotically optimal* in the Hájek–Le Cam sense (Hájek '70; Le Cam '71).
- Enables online covariance estimation and construction of confidence intervals for $\boldsymbol{x}^\star$.

Statistical guarantee for SGD

Theorem (Ruppert (1988); Polyak–Juditsky (1992))

If $\alpha_k \propto k^{-\theta}$ *with* $\theta \in (\frac{1}{2}, 1)$ *and standard noise conditions hold, then*

$$\sqrt{k}(\overline{\boldsymbol{x}}_k - \boldsymbol{x}^\star) \xrightarrow{d} \mathcal{N}\Big(0,\ \nabla^2 f(\boldsymbol{x}^\star)^{-1}\, \Sigma\, \nabla^2 f(\boldsymbol{x}^\star)^{-1}\Big),$$

where $\overline{\boldsymbol{x}}_k := \frac{1}{k}\sum_{i=1}^{k} \boldsymbol{x}^i$, *and* $\Sigma := \mathrm{Cov}\big(\nabla F(\boldsymbol{x}, \xi)\big)$.

- Matches the covariance of SAA (sample-average approximation) $\Rightarrow$ same statistical efficiency.
- *Asymptotically optimal* in the Hájek–Le Cam sense (Hájek '70; Le Cam '71).
- Enables online covariance estimation and construction of confidence intervals for $\boldsymbol{x}^\star$.

From unconstrained to constrained

We consider

$$\min_{x \in \mathcal{X}} f(x) := \mathbb{E}[F(x, \xi)], \qquad \mathcal{X} := \{ x \in \mathbb{R}^n : h_i(x) \leq 0,\ i \in [m] \}. \tag{P}$$

Assume $F(\cdot, \xi)$ is C^2 in x (a.s.) and each h_i is C^2.

- **SAA baseline.** Sample-average approximation admits asymptotic normality and is *asymptotically optimal* (Hájek–Le Cam sense) (Dupacová–Wets '88; Shapiro '89; King–Rockafellar '93).
- **Are practical online first-order methods also optimal?**
  - **Dual Averaging** (Duchi–Ruan '18): optimal *only when* $\mathcal{X}$ is polyhedral.
  - **Projected SGD** (Davis–Drusvyatskiy–Jiang '23): optimal for general $\mathcal{X}$, but lacks finite-time active-set identification and requires stronger noise assumptions.
- Both are **primal-only** and require explicit projection onto $\mathcal{X}$.

From unconstrained to constrained

We consider

$$\min_{x \in \mathcal{X}} f(x) := \mathbb{E}[F(x, \xi)], \qquad \mathcal{X} := \{ x \in \mathbb{R}^n : h_i(x) \leq 0,\ i \in [m] \}. \tag{P}$$

Assume $F(\cdot, \xi)$ is C^2 in x (a.s.) and each h_i is C^2.

- **SAA baseline.** Sample-average approximation admits asymptotic normality and is *asymptotically optimal* (Hájek–Le Cam sense) (Dupacová–Wets '88; Shapiro '89; King–Rockafellar '93).
- **Are practical online first-order methods also optimal?**
  - **Dual Averaging** (Duchi–Ruan '18): optimal *only when* $\mathcal{X}$ is polyhedral.
  - **Projected SGD** (Davis–Drusvyatskiy–Jiang '23): optimal for general $\mathcal{X}$, but lacks finite-time active-set identification and requires stronger noise assumptions.
- Both are **primal-only** and require explicit projection onto $\mathcal{X}$.

From unconstrained to constrained

We consider

$$\min_{x \in \mathcal{X}} f(x) := \mathbb{E}[F(x, \xi)], \qquad \mathcal{X} := \{ x \in \mathbb{R}^n : h_i(x) \leq 0,\ i \in [m] \}. \tag{P}$$

Assume $F(\cdot, \xi)$ is C^2 in x (a.s.) and each h_i is C^2.

- **SAA baseline.** Sample-average approximation admits asymptotic normality and is *asymptotically optimal* (Hájek–Le Cam sense) (Dupacová–Wets '88; Shapiro '89; King–Rockafellar '93).
- **Are practical online first-order methods also optimal?**
  - **Dual Averaging** (Duchi–Ruan '18): optimal *only when* $\mathcal{X}$ is polyhedral.
  - **Projected SGD** (Davis–Drusvyatskiy–Jiang '23): optimal for general $\mathcal{X}$, but lacks finite-time active-set identification and requires stronger noise assumptions.
- Both are **primal-only** and require explicit projection onto $\mathcal{X}$.

Primal–dual methods: what do we want?

Any practical *and* optimal primal–dual method?

- **Asymptotically optimal** for general constrained problems
  (Hájek–Le Cam efficiency; matches information-theoretic lower bounds)
- **Projection-free** onto $\mathcal{X}$ (only easy projection onto a redundant compact X_0 if needed)
- **Joint inference**: construct confidence intervals for both $x^\star$ *and* $\lambda^\star$

Primal–dual methods: what do we want?

Any practical *and* optimal primal–dual method?

- **Asymptotically optimal** for general constrained problems
  (Hájek–Le Cam efficiency; matches information-theoretic lower bounds)
- **Projection-free** onto $\mathcal{X}$ (only easy projection onto a redundant compact X_0 if needed)
- **Joint inference**: construct confidence intervals for both $\boldsymbol{x}^\star$ *and* $\boldsymbol{\lambda}^\star$

Stochastic Linearized Augmented Lagrangian Method (S-LALM)

Define the augmented Lagrangian function (Ben-Tal & Zibulevsky, 1997; Xu, 2021):

$$\mathcal{L}_\rho(\boldsymbol{x}, \boldsymbol{\lambda}) := f(\boldsymbol{x}) \; + \; \frac{\rho}{2} \sum_{i=1}^{m} \left[\max\left\{ h_i(\boldsymbol{x}) + \frac{\lambda_i}{\rho}, \, 0 \right\}^2 - \frac{\lambda_i^2}{\rho^2} \right].$$

Update rules:

$$\boldsymbol{x}^{k+1} \leftarrow \Pi_{\mathcal{X}_0}(\boldsymbol{x}^k - \alpha_k(\nabla_{\boldsymbol{x}} \mathcal{L}_\rho(\boldsymbol{x}^k, \boldsymbol{\lambda}^k) + \xi^k)),$$

$$\boldsymbol{\lambda}^{k+1} \leftarrow \boldsymbol{\lambda}^k + \beta_k \cdot \max\left\{ -\frac{\boldsymbol{\lambda}^k}{\rho}, \, h(\boldsymbol{x}^k) \right\}.$$

- $\mathcal{X}_0$: a **compact redundant constraint set** with $\boldsymbol{x}^\star \in \operatorname{int}(\mathcal{X}_0)$ (e.g., a simple norm ball, easy to project onto).
- The stochastic noise ξ^k enters **only through the primal update**.
- Uses the **simplest gradient update** — no extrapolation or momentum.

Stochastic Linearized Augmented Lagrangian Method (S-LALM)

Define the augmented Lagrangian function (Ben-Tal & Zibulevsky, 1997; Xu, 2021):

$$\mathcal{L}_\rho(\boldsymbol{x}, \boldsymbol{\lambda}) := f(\boldsymbol{x}) \; + \; \frac{\rho}{2} \sum_{i=1}^{m} \left[\max\left\{ h_i(\boldsymbol{x}) + \frac{\lambda_i}{\rho},\, 0 \right\}^2 - \frac{\lambda_i^2}{\rho^2} \right].$$

Update rules:

$$\boxed{\begin{aligned} \boldsymbol{x}^{k+1} &\leftarrow \Pi_{\mathcal{X}_0}\big(\boldsymbol{x}^k - \alpha_k(\nabla_{\boldsymbol{x}} \mathcal{L}_\rho(\boldsymbol{x}^k, \boldsymbol{\lambda}^k) + \xi^k)\big), \\ \boldsymbol{\lambda}^{k+1} &\leftarrow \boldsymbol{\lambda}^k + \beta_k \cdot \max\left\{ -\frac{\boldsymbol{\lambda}^k}{\rho},\, h(\boldsymbol{x}^k) \right\}. \end{aligned}}$$

- $\mathcal{X}_0$: a **compact redundant constraint set** with $\boldsymbol{x}^\star \in \operatorname{int}(\mathcal{X}_0)$ (e.g., a simple norm ball, easy to project onto).
- The stochastic noise ξ^k enters **only through the primal update**.
- Uses the **simplest gradient update** — no extrapolation or momentum.

Stochastic Linearized Augmented Lagrangian Method (S-LALM)

Define the augmented Lagrangian function (Ben-Tal & Zibulevsky, 1997; Xu, 2021):

$$\mathcal{L}_\rho(\boldsymbol{x}, \boldsymbol{\lambda}) := f(\boldsymbol{x}) \; + \; \frac{\rho}{2} \sum_{i=1}^{m} \left[\max\left\{ h_i(\boldsymbol{x}) + \frac{\lambda_i}{\rho},\, 0 \right\}^2 - \frac{\lambda_i^2}{\rho^2} \right].$$

Update rules:

$$\boxed{\begin{aligned} \boldsymbol{x}^{k+1} &\leftarrow \Pi_{\mathcal{X}_0}\big(\boldsymbol{x}^k - \alpha_k(\nabla_{\boldsymbol{x}} \mathcal{L}_\rho(\boldsymbol{x}^k, \boldsymbol{\lambda}^k) + \xi^k)\big), \\ \boldsymbol{\lambda}^{k+1} &\leftarrow \boldsymbol{\lambda}^k + \beta_k \cdot \max\left\{ -\frac{\boldsymbol{\lambda}^k}{\rho},\, h(\boldsymbol{x}^k) \right\}. \end{aligned}}$$

- $\mathcal{X}_0$: a **compact redundant constraint set** with $\boldsymbol{x}^\star \in \mathrm{int}(\mathcal{X}_0)$ (e.g., a simple norm ball, easy to project onto).
- The stochastic noise ξ^k enters **only through the primal update**.
- Uses the **simplest gradient update** — no extrapolation or momentum.

Standard regularity conditions

Intuitively, we impose conditions ensuring that $(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star)$ is **identifiable**.

- **Active set**: $\mathcal{I} := \{\, i \in [m] : h_i(\boldsymbol{x}^\star) = 0 \,\}$.
- (LICQ) The gradients $\{\nabla h_i(\boldsymbol{x}^\star)\}_{i \in \mathcal{I}}$ are linearly independent; equivalently, $D := [\, \nabla h_i(\boldsymbol{x}^\star) \,]_{i \in \mathcal{I}}$ has full column rank.
- (SC) Strict complementarity: $\lambda_i^\star > 0$ for all $i \in \mathcal{I}$.
- (SOSC) For all nonzero $d \in \ker(D^\top)$,

$$d^\top \nabla^2_{\boldsymbol{x}\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star)\, d \;>\; 0.$$

Standard regularity conditions

Intuitively, we impose conditions ensuring that $(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star)$ is **identifiable**.

- **Active set**: $\mathcal{I} := \{\, i \in [m] : h_i(\boldsymbol{x}^\star) = 0 \,\}$.
- (LICQ) The gradients $\{\nabla h_i(\boldsymbol{x}^\star)\}_{i \in \mathcal{I}}$ are linearly independent; equivalently, $D := [\, \nabla h_i(\boldsymbol{x}^\star) \,]_{i \in \mathcal{I}}$ has full column rank.
- (SC) Strict complementarity: $\lambda_i^\star > 0$ for all $i \in \mathcal{I}$.
- (SOSC) For all nonzero $d \in \ker(D^\top)$,

$$d^\top \nabla^2_{\boldsymbol{xx}} \mathcal{L}(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star)\, d \;>\; 0.$$

Main theorem

Theorem (Li–Ruan, 2026)

Let $\alpha_k \propto k^{-\theta}$, $\beta_k \propto k^{-\theta}$ *with* $\theta \in (\frac{1}{2}, 1)$. *Assume* (LICQ), (SC), (SOSC) *and standard noise conditions. Then*

$$\sqrt{k}\begin{bmatrix} \bar{\boldsymbol{x}}_k - \boldsymbol{x}^\star \\ \bar{\boldsymbol{\lambda}}_k - \boldsymbol{\lambda}^\star \end{bmatrix} \xrightarrow{d} \mathcal{N}\left(0, \begin{bmatrix} J_p \Sigma J_p^\top & J_p \Sigma J_d^\top \\ J_d \Sigma J_p^\top & J_d \Sigma J_d^\top \end{bmatrix}\right),$$

where

$$J_p := \left(P_T H_0 P_T\right)^\dagger, \qquad J_d := (D^\top D)^{-1} D^\top \Big(I - H_0 \left(P_T H_0 P_T\right)^\dagger\Big),$$

$T := \ker(D^\top)$ *is the tangent space of the active manifold,* $P_T = I - D(D^\top D)^{-1} D^\top$ *is the orthogonal projector, and* H_0 *is the Hessian of the Lagrangian w.r.t.* $\boldsymbol{x}$ *at* $(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star)$.

- The primal covariance $J_p \Sigma J_p^\top$ **matches the information-theoretic lower bound** (Duchi–Ruan '18; Davis–Drusvyatskiy–Jiang '23) — *asymptotically optimal.*

Main theorem

Theorem (Li–Ruan, 2026)

Let $\alpha_k \propto k^{-\theta}$, $\beta_k \propto k^{-\theta}$ *with* $\theta \in (\frac{1}{2}, 1)$. *Assume* (LICQ), (SC), (SOSC) *and standard noise conditions. Then*

$$\sqrt{k}\begin{bmatrix} \bar{\boldsymbol{x}}_k - \boldsymbol{x}^\star \\ \bar{\boldsymbol{\lambda}}_k - \boldsymbol{\lambda}^\star \end{bmatrix} \xrightarrow{d} \mathcal{N}\left(0, \begin{bmatrix} J_p \Sigma J_p^\top & J_p \Sigma J_d^\top \\ J_d \Sigma J_p^\top & J_d \Sigma J_d^\top \end{bmatrix}\right),$$

where

$$J_p := \left(P_T H_0 P_T\right)^\dagger, \qquad J_d := (D^\top D)^{-1} D^\top \Big(I - H_0 \left(P_T H_0 P_T\right)^\dagger\Big),$$

$T := \ker(D^\top)$ *is the tangent space of the active manifold,* $P_T = I - D(D^\top D)^{-1} D^\top$ *is the orthogonal projector, and* H_0 *is the Hessian of the Lagrangian w.r.t.* $\boldsymbol{x}$ *at* $(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star)$.

- The primal covariance $J_p \Sigma J_p^\top$ **matches the information-theoretic lower bound** (Duchi–Ruan '18; Davis–Drusvyatskiy–Jiang '23) — *asymptotically optimal.*

Proof sketch

- (★★★ Almost sure convergence)

$$x^k \xrightarrow{\text{a.s.}} x^\star, \qquad \lambda^k \xrightarrow{\text{a.s.}} \lambda^\star.$$

- (★ Finite-time active-set identification) With probability one, there exists a finite (random) index $K < \infty$ such that, for all $k \geq K$,

$$\{\, i \in [m] : \rho\, h_i(x^k) + \lambda_i^k > 0 \,\} \;=\; \mathcal{I}.$$

- (★★ Local convergence rate) As $k \to \infty$,

$$\frac{1}{\sqrt{k}} \sum_{i=1}^{k} \|x^i - x^\star\|^2 \;\longrightarrow\; 0, \qquad \frac{1}{\sqrt{k}} \sum_{i=1}^{k} \|\lambda^i - \lambda^\star\|^2 \;\longrightarrow\; 0 \quad \text{a.s.}$$

Key challenge

Primal–dual asymmetry: there is *no explicit curvature* on the dual side.

Proof sketch

- (★★★ Almost sure convergence)

$$\boxed{x^k \xrightarrow{\text{a.s.}} x^\star, \qquad \lambda^k \xrightarrow{\text{a.s.}} \lambda^\star.}$$

- (★ Finite-time active-set identification) With probability one, there exists a finite (random) index $K < \infty$ such that, for all $k \geq K$,

$$\boxed{\{\, i \in [m] : \rho\, h_i(x^k) + \lambda_i^k > 0 \,\} \;=\; \mathcal{I}.}$$

- (★★ Local convergence rate) As $k \to \infty$,

$$\frac{1}{\sqrt{k}} \sum_{i=1}^{k} \|x^i - x^\star\|^2 \;\longrightarrow\; 0, \qquad \frac{1}{\sqrt{k}} \sum_{i=1}^{k} \|\lambda^i - \lambda^\star\|^2 \;\longrightarrow\; 0 \quad \text{a.s.}$$

Key challenge

Primal–dual asymmetry: there is *no explicit curvature* on the dual side.

Proof sketch

- (★★★ Almost sure convergence)

$$\boxed{x^k \xrightarrow{\text{a.s.}} x^\star, \qquad \lambda^k \xrightarrow{\text{a.s.}} \lambda^\star.}$$

- (★ Finite-time active-set identification) With probability one, there exists a finite (random) index $K < \infty$ such that, for all $k \geq K$,

$$\boxed{\{\, i \in [m] : \rho\, h_i(x^k) + \lambda_i^k > 0 \,\} \;=\; \mathcal{I}.}$$

- (★★ Local convergence rate) As $k \to \infty$,

$$\boxed{\frac{1}{\sqrt{k}} \sum_{i=1}^{k} \|x^i - x^\star\|^2 \;\longrightarrow\; 0, \qquad \frac{1}{\sqrt{k}} \sum_{i=1}^{k} \|\lambda^i - \lambda^\star\|^2 \;\longrightarrow\; 0 \quad \text{a.s.}}$$

Key challenge

Primal–dual asymmetry: there is *no explicit curvature* on the dual side.

Proof sketch

- (★★★ Almost sure convergence)

$$\boxed{x^k \xrightarrow{\text{a.s.}} x^\star, \qquad \lambda^k \xrightarrow{\text{a.s.}} \lambda^\star.}$$

- (★ Finite-time active-set identification) With probability one, there exists a finite (random) index $K < \infty$ such that, for all $k \geq K$,

$$\boxed{\{\, i \in [m] : \rho\, h_i(x^k) + \lambda_i^k > 0 \,\} \;=\; \mathcal{I}.}$$

- (★★ Local convergence rate) As $k \to \infty$,

$$\boxed{\frac{1}{\sqrt{k}} \sum_{i=1}^{k} \|x^i - x^\star\|^2 \;\longrightarrow\; 0, \qquad \frac{1}{\sqrt{k}} \sum_{i=1}^{k} \|\lambda^i - \lambda^\star\|^2 \;\longrightarrow\; 0 \quad \text{a.s.}}$$

Key challenge

Primal–dual asymmetry: there is *no explicit curvature* on the dual side.

Proof sketch

- (★★★ Almost sure convergence)

$$\boxed{x^k \xrightarrow{\text{a.s.}} x^\star, \qquad \lambda^k \xrightarrow{\text{a.s.}} \lambda^\star.}$$

- (★ Finite-time active-set identification) With probability one, there exists a finite (random) index $K < \infty$ such that, for all $k \geq K$,

$$\boxed{\{\, i \in [m] : \rho\, h_i(x^k) + \lambda_i^k > 0 \,\} \;=\; \mathcal{I}.}$$

- (★★ Local convergence rate) As $k \to \infty$,

$$\boxed{\frac{1}{\sqrt{k}} \sum_{i=1}^{k} \|x^i - x^\star\|^2 \;\longrightarrow\; 0, \qquad \frac{1}{\sqrt{k}} \sum_{i=1}^{k} \|\lambda^i - \lambda^\star\|^2 \;\longrightarrow\; 0 \quad \text{a.s.}}$$

Key challenge

Primal–dual asymmetry: there is *no explicit curvature* on the dual side.

Almost sure convergence of S-LALM

Theorem (Li–Ruan, 2026)

Let $\{(x^k, \lambda^k)\}_{k\geq 0}$ be generated by S-LALM under the (SOSC) *condition and standard noise assumptions. Then, almost surely, every cluster point of $\{(x^k, \lambda^k)\}$ is a KKT point of Problem* (P).

- The only regularity assumption required is **(SOSC)**.
- (SOSC) implies the **quadratic growth property** of the nonsmooth augmented Lagrangian:

$$\mathcal{L}_\rho(x, \lambda^\star) - \mathcal{L}_\rho(x^\star, \lambda^\star) \;\geq\; \varepsilon \|x - x^\star\|^2, \qquad \forall\, x \in \mathcal{X}_0.$$

- This result is of **independent interest**—it establishes almost sure convergence under minimal assumptions.

Almost sure convergence of S-LALM

Theorem (Li–Ruan, 2026)

Let $\{(x^k, \lambda^k)\}_{k\geq 0}$ be generated by S-LALM under the (SOSC) *condition and standard noise assumptions. Then, almost surely, every cluster point of $\{(x^k, \lambda^k)\}$ is a KKT point of Problem* (P).

- The only regularity assumption required is **(SOSC)**.
- (SOSC) implies the **quadratic growth property** of the nonsmooth augmented Lagrangian:

$$\mathcal{L}_\rho(x, \lambda^\star) - \mathcal{L}_\rho(x^\star, \lambda^\star) \;\geq\; \varepsilon \|x - x^\star\|^2, \qquad \forall\, x \in \mathcal{X}_0.$$

- This result is of **independent interest**—it establishes almost sure convergence under minimal assumptions.

Example

Example

Suppose that $\xi \sim \mathcal{N}(0, I_3)$. We consider

$$\begin{aligned} \min_{\boldsymbol{x} \in \mathcal{X}_0} \quad & \mathbb{E}[(-e_3 + \xi)^\top \boldsymbol{x}] \\ \text{s.t.} \quad & 2\sqrt{3}(\|\boldsymbol{x} - e_1\|^2 - 4) \leq 0, \\ & 2\sqrt{3}(\|\boldsymbol{x} + e_1\|^2 - 4) \leq 0, \end{aligned}$$

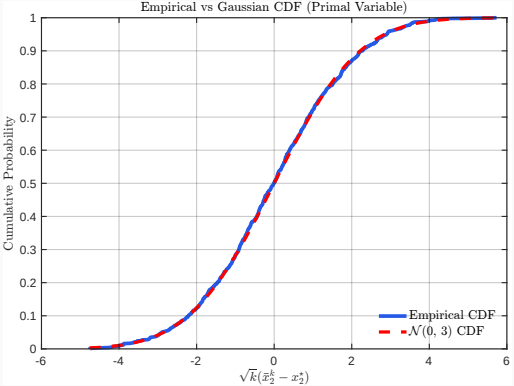
where $\mathcal{X}_0 = \{\boldsymbol{x} \in \mathbb{R}^3 : \|\boldsymbol{x}\|_\infty \leq 3\}$.

$$\sqrt{k} \begin{bmatrix} \bar{\boldsymbol{x}}^k - \boldsymbol{x}^\star \\ \bar{\boldsymbol{\lambda}}^k - \boldsymbol{\lambda}^\star \end{bmatrix} \xrightarrow{d} \mathcal{N}\left(0, \begin{bmatrix} \operatorname{diag}(0,3,0) & 0 \\ 0 & \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix} \end{bmatrix}\right).$$

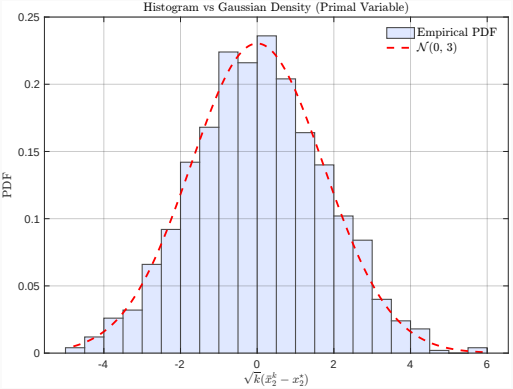
- Hyperparameters set up:

$$\rho = 0.2, \qquad \alpha_k = \frac{0.1}{k^{0.51}}, \qquad \beta_k = \frac{0.05}{k^{0.51}}.$$

Primal Variable

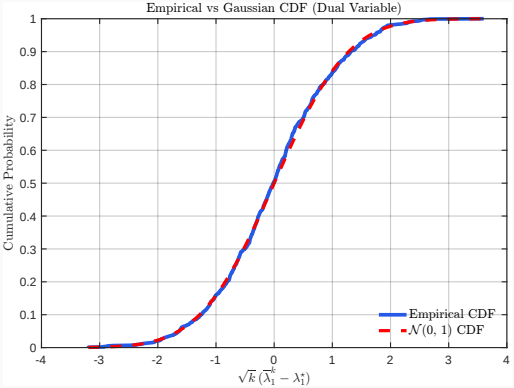


(a) Empirical CDF of the primal variable

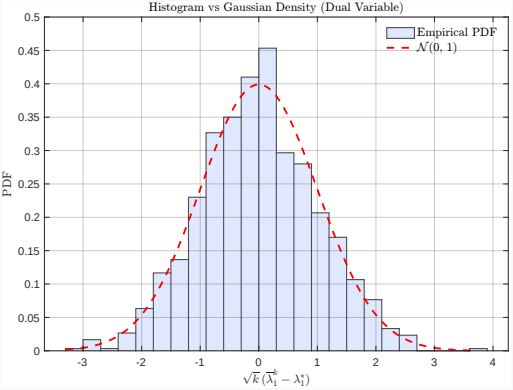


(b) Empirical PDF of the primal variable

Dual Variable

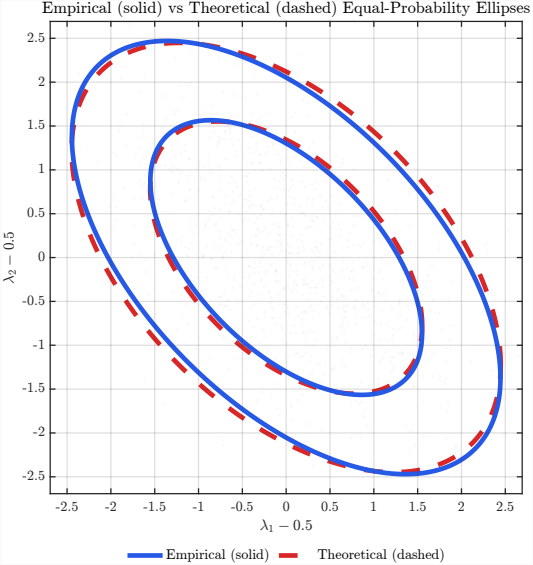


(a) Empirical CDF of the dual variable



(b) Empirical PDF of the dual variable

Covariance of Dual Variables



Summary

- S-LALM gives a projection-light online alternative to SAA.
- Averaging yields asymptotic normality with optimal primal covariance.
- The same limit theory gives joint inference for $(x^\star, \lambda^\star)$.

Thank you for your listening!

Any questions?