Gaussian Phase Transitions and Conic Intrinsic Volumes: Steining the Steiner formula

Larry Goldstein, University of Southern California

Nourdin
GIoVAnNi Luxembourg University
Peccati

IMA Minnesota, April 14^{th}, 2015
[arxiv1411.6265]

Ingredients

- Convex Geometry
- Compressed sensing
- Gaussian inequalities

Ingredients

- Convex Geometry: Angular version of the classical Steiner formula for expansion of compact convex sets.
- Compressed sensing
- Gaussian inequalities

Ingredients

- Convex Geometry: Angular version of the classical Steiner formula for expansion of compact convex sets.
- Compressed sensing: Recovery of unknowns under structural assumptions such as sparsity for vectors, or a low rank condition for matrices.
- Gaussian inequalities

Ingredients

- Convex Geometry: Angular version of the classical Steiner formula for expansion of compact convex sets reveal the 'intrinsic volume' distributions associated with closed convex cones.
- Compressed sensing: Recovery of unknowns under structural assumptions such as sparsity for vectors or a low rank condition, for matrices.
- Gaussian inequalities and Stein's Method: For providing finite sample bounds on Gaussian fluctuations.

Concentration Connection

Subject of this work was motivated by the papers by Amelunxen, Lotz, McCoy and Tropp [AMLT13] and McCoy and Tropp [MT14], where the conic intrinsic volume distributions were studied using Poincaré and log Sobolev inequalities.

In particular it was shown that these distributions exhibit concentration phenomenon.

Raises the possibility that one might make use of other Gaussian inequalities that may be lurking about in the background.

Concentration Connection

Subject of this work was motivated by the papers by Amelunxen, Lotz, McCoy and Tropp [AMLT13] and McCoy and Tropp [MT14], where the conic intrinsic volume distributions were studied using Poincaré and log Sobolev inequalities.

In particular it was shown that these distributions exhibit concentration phenomenon.

Raises the possibility that one might make use of other Gaussian inequalities that may be lurking about in the background.

Concentration of Conic Intrisic Volumes

From: Living on the Edge [ALMT13]

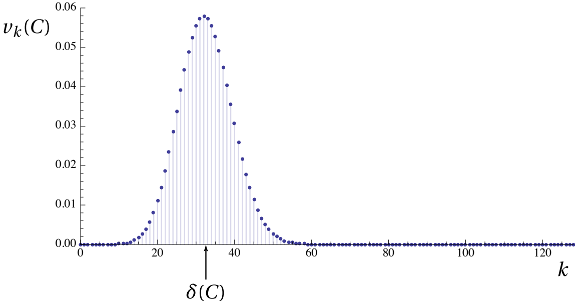

FIGURE 2.2: **Concentration of conic intrinsic volumes.** This plot displays the conic intrinsic volumes $v_k(C)$ of a circular cone $C \subset \mathbb{R}^{128}$ with angle $\pi/6$. The distribution concentrates sharply around the statistical dimension $\delta(C) \approx 32.5$. See Section 3.4 for further discussion of this example.

Convex Geometry

For any closed convex set K, let

$$d(\mathbf{x}, K) = \inf_{\mathbf{y} \in K} ||\mathbf{x} - \mathbf{y}||.$$

Classical Steiner formula (1840) for the expansion of a compact convex set $K \subset \mathbb{R}^d$, with B_j the unit ball in $\mathbb{R}^j$,

$$\mathrm{Vol}_d\{\mathbf{x} : d(\mathbf{x}, K) \leq \lambda\} = \sum_{j=0}^{d} \lambda^{d-j} \mathrm{Vol}(B_{d-j}) \mathcal{V}_j.$$

Intrinsic volumes, $\mathcal{V}_d$ is volume, $2\mathcal{V}_{d-1}$ is surface area ... and $\mathcal{V}_0$ is Euler characteristic.

Convex Geometry

The set $C \subset \mathbb{R}^d$ is a cone if $\tau(\mathbf{x}+\mathbf{y}) \in C$ for all $\{\mathbf{x},\mathbf{y}\} \subset C$ and $\tau > 0$.
With S^{j-1} unit sphere in $\mathbb{R}^j$, one has the angular analog, Hergoltz (1943),

$$\mathrm{Vol}_{d-1}\left\{\mathbf{x} \in S^{d-1} : d^2(\mathbf{x}, C) \ \leq \lambda\right\} = \sum_{j=0}^{d} \beta_{j,d}(\lambda) v_j$$

The numbers $v_0, \ldots, v_d$ are the conic intrinsic volumes, and are non negative and sum to 1.

Can associate a distribution $\mathcal{L}(V)$ given by $P(V=j) = v_j$ to the cone C, also write V_C.

Recovery of Structured Unknowns

Observe (a small number) m of random linear combinations of an unknown $\mathbf{x}_0 \in \mathbb{R}^d$ (in high dimension),

$\mathbf{z} = A\mathbf{x}_0$ for $A \in \mathbb{R}^{m \times d}$ known, with i.i.d. $\mathcal{N}(0,1)$ entries.

Unknown $\mathbf{x}_0$ lies in the feasible region

$$F = \{\mathbf{x} : A\mathbf{x} = \mathbf{z}\} = \mathrm{Null}(A) + \mathbf{x}_0.$$

With $m < d$ not possible to recover $\mathbf{x}_0$. Say $\mathbf{x}_0$ is sparse, has some small number s of non-zero entries, $\|\mathbf{x}_0\|_0 = s$.

Minimizing $\|\mathbf{x}\|_0$ over $\mathbf{x} \in F$ is computationally infeasible.

Recovery of Structured Unknowns

Observe (a small number) m of random linear combinations of an unknown $\mathbf{x}_0 \in \mathbb{R}^d$ (in high dimension),

$\mathbf{z} = A\mathbf{x}_0$ for $A \in \mathbb{R}^{m\times d}$ known, with i.i.d. $\mathcal{N}(0,1)$ entries.

Unknown $\mathbf{x}_0$ lies in the feasible region

$$F = \{\mathbf{x} : A\mathbf{x} = \mathbf{z}\} = \mathrm{Null}(A) + \mathbf{x}_0.$$

With $m < d$ not possible to recover $\mathbf{x}_0$. Say $\mathbf{x}_0$ is sparse, has some small number s of non-zero entries, $\|\mathbf{x}_0\|_0 = s$.

Minimizing $\|\mathbf{x}\|_0$ over $\mathbf{x} \in F$ is computationally infeasible.

Recovery of Structured Unknowns

Observe (a small number) m of random linear combinations of an unknown $\mathbf{x}_0 \in \mathbb{R}^d$ (in high dimension),

$\mathbf{z} = A\mathbf{x}_0 \quad$ for $A \in \mathbb{R}^{m \times d}$ known, with i.i.d. $\mathcal{N}(0,1)$ entries.

Unknown $\mathbf{x}_0$ lies in the feasible region

$$F = \{\mathbf{x} : A\mathbf{x} = \mathbf{z}\} = \mathrm{Null}(A) + \mathbf{x}_0.$$

With $m < d$ not possible to recover $\mathbf{x}_0$. Say $\mathbf{x}_0$ is sparse, has some small number s of non-zero entries, $\|\mathbf{x}_0\|_0 = s$.

Minimizing $\|\mathbf{x}\|_0$ over $\mathbf{x} \in F$ is computationally infeasible.

Convex Program

Determine a convex function $f(\mathbf{x})$ that promotes the known structure of $\mathbf{x}_0$, and

$$\text{minimize } f(\mathbf{x}) \text{ over } \mathbf{x} \in F.$$

Chandrasekaran, Recht, Parrilo, and Willsky (2012) show how to construct f for some given structure.

To promote sparsity let $f(\mathbf{x}) = \|\mathbf{x}\|_1$, the L^1 norm, i.e. $\sum_{i=1}^{d} |x_i|$.

Candès, Romberg, Tao (2006) and Donoho (2006), Donoho and Tanner (2009).

Why does it work?

Descent Cone

For $f(\mathbf{x}) = \|\mathbf{x}\|_1$ and $\mathbf{x}_0 \in \mathbb{R}^d$, let

$$\mathcal{D}(f, \mathbf{x}_0) = \{\mathbf{y} \in \mathbb{R}^d : \exists \tau > 0, \|\mathbf{x}_0 + \tau \mathbf{y}\|_1 \leq \|\mathbf{x}_0\|_1\}.$$

The set of all directions from which, starting at x_0, do not increase the L^1 norm.

Convex recovery of unknowns

From: Living on the Edge [ALMT13]

PHASE TRANSITIONS IN RANDOM CONVEX PROGRAMS 7

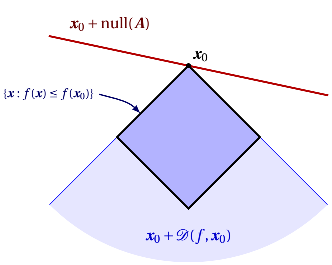

FIGURE 2.3: **The optimality condition for a regularized inverse problem.** The condition for the regularized linear inverse problem (2.4) to succeed requires that the descent cone $\mathscr{D}(f, x_0)$ and the null space $\mathrm{null}(A)$ do not share a ray. **[left]** The regularized linear inverse problem succeeds. **[right]** The regularized linear inverse problem fails.

Convex recovery of sparse unknowns

So, translating by $\mathbf{x}_0$, success if and only if

$$\mathrm{Null}(A) \cap C = \{0\}$$

where C is the descent cone of the L^1 norm at $\mathbf{x}_0$.

We know $\dim(\mathrm{Null}(A)) = d - m$. So if the 'dimension of C' were $\delta(C)$, we would want

$$d - m + \delta(C) \leq d \quad \text{so} \quad m \geq \delta(C).$$

Statistical Dimension of C is $\delta(C) = E[V_C]$. Recovery success probability p sandwiched

$$P(V \leq m - 1) \leq p \leq P(V \leq m).$$

Concentration would imply sharp transition, threshold phenomenon.

Metric Projection

For K any closed convex set, the infimum

$$d(\mathbf{x}, K) = \inf_{\mathbf{y} \in K} \|\mathbf{x} - \mathbf{y}\|,$$

is attained at a unique vector, called the metric projection of $\mathbf{x}$ onto K, denoted by $\Pi_K(\mathbf{x})$.

Gaussian Connection

The cone C is polyhedral if there exists an integer N and vectors $\mathbf{u}_1, \ldots, \mathbf{u}_N$ in $\mathbb{R}^d$ such that

$$C = \bigcap_{i=1}^{N} \{\mathbf{x} \in \mathbb{R}^d : \langle \mathbf{u}_i, \mathbf{x} \rangle \geq 0\}.$$

For $\mathbf{g} \sim \mathcal{N}(0, I_d)$, the conic intrinsic volumes satisfy

$v_j = P\,(\Pi_C(\mathbf{g})$ lies in the relative interior of a j-dimensional face of C)

Consider the orthant $\mathbb{R}^d_+$.

'Master Steiner Formula' [MT14]

Implies that with X_i independent χ_1^2 random variables,

$$\|\Pi_C(\mathbf{g})\|^2 =_d \sum_{j=0}^{V_C} X_i.$$

Letting $G_C = \|\Pi_C(\mathbf{g})\|^2$,

$$G_C =_d \sum_{j=0}^{V_C} X_i = \sum_{j=0}^{V_C} (X_i - 1) + V_C,$$

so recalling $\delta_C = E[V_C]$ and letting $\tau_C^2 = \mathrm{Var}(V_C)$, we see

$$E[G_C] = \delta_C \quad \text{and} \quad \mathrm{Var}(G_C) = 2\delta_C + \tau_C^2$$

Master Steiner Formula [MT14]

Relation

$$G_C =_d \sum_{j=0}^{V_C} X_i \quad \text{for } G_C = \|\Pi_C(\mathbf{g})\|^2$$

yields (strange) moment generating function identity

$$Ee^{tV} = Ee^{\xi_t G} \quad \text{with} \quad \xi_t = \frac{1}{2}\left(1 - e^{-2t}\right).$$

Concentration [MT14]

From

$$Ee^{tV} = Ee^{\xi_t G} \quad \text{with} \quad \xi_t = \frac{1}{2}\left(1 - e^{-2t}\right),$$

and the fact that $G = \|\Pi_C(\mathbf{g})\|^2$, use log Sobolev inequality to obtain bound on its moment generating function.

Translate into bound on moment generating fucntion for V_C to show concentration around δ_C, implying phase transition for exact recovery.

Concentration [MT14]

From

$$Ee^{tV} = Ee^{\xi_t G} \quad \text{with} \quad \xi_t = \frac{1}{2}\left(1 - e^{-2t}\right),$$

and the fact that $G = \|\Pi_C(\mathbf{g})\|^2$, use log Sobolev inequality to obtain bound on its moment generating function.

Translate into bound on moment generating fucntion for V_C to show concentration around δ_C, implying phase transition for exact recovery.

Normal Fluctuations

Standardizing

$$G_C =_d \sum_{j=0}^{V_C} (X_i - 1) + V_C = W_C + V_C$$

yields

$$\frac{G_C - \delta_C}{\sigma_C} =_d \left(\frac{\sqrt{2\delta_C}}{\sigma_C}\right) \frac{W_C}{\sqrt{2\delta_C}} + \left(\frac{\tau_C}{\sigma_C}\right) \frac{V_C - \delta_C}{\tau_C}.$$

Standardized G_C is asymptotically normal and expressions on the right hand side are asymptotically independent. Apply Cramér's theorem.

Normal Fluctuations

Standardizing

$$G_C =_d \sum_{j=0}^{V_C}(X_i - 1) + V_C = W_C + V_C$$

yields

$$\frac{G_C - \delta_C}{\sigma_C} =_d \left(\frac{\sqrt{2\delta_C}}{\sigma_C}\right)\frac{W_C}{\sqrt{2\delta_C}} + \left(\frac{\tau_C}{\sigma_C}\right)\frac{V_C - \delta_C}{\tau_C}.$$

Standardized G_C is asymptotically normal and expressions on the right hand side are asymptotically independent. Apply Cramér's theorem.

Variance Bounds

For first ratio $\sqrt{2\delta_C}/\sigma_C$, by conditional variance formula, and Poincaré inequality

$$\sigma_C^2 = 2\delta_C + \tau_C^2 \quad \text{and} \quad 2\delta_C \leq \sigma_C^2 \leq 4\delta_C.$$

For second ratio τ_C/σ_C, we have $\tau_C^2 \leq \sigma_C^2$ and

$$\frac{\max\{v^2, 4b^2\}}{b} \leq \mathrm{Var}(V_C)$$

where

$$v = \|E\Pi_C(\mathbf{g})\|^2 \quad \text{and} \quad b = \sqrt{2\delta_C/2}$$

Toward the 2^{nd} order Poincaré Inequality

1^{st}: Poincaré Inequality. If $W(\mathbf{g})$ is a smooth real valued function of $\mathbf{g} \sim \mathcal{N}(0, I_d)$, then

$$\mathrm{Var}(W(\mathbf{g})) \leq E\|\nabla W(\mathbf{g})\|^2.$$

e.g.,

$$\sigma_C^2 = \mathrm{Var}\left(\|\Pi_C(\mathbf{g})\|^2\right) \leq E\|2\Pi_C(\mathbf{g})\|^2 = 4\delta_C.$$

Poincaré Inequality

Take $E[W]=0, \varphi$ a smooth function, $\widehat{\mathbf{g}}$ an independent copy of $\mathbf{g}$, and consider

$$E[W\varphi(W)] \quad = \quad E[(W(\mathbf{g})-W(\widehat{\mathbf{g}}))\,\varphi(W(\mathbf{g}))]$$

Poincaré Inequality

Take $E[W] = 0, \varphi$ a smooth function, $\widehat{\mathbf{g}}$ an independent copy of $\mathbf{g}$, and let $\widehat{\mathbf{g}}_t = e^{-t}\mathbf{g} + \sqrt{1 - e^{-2t}}\widehat{\mathbf{g}}$

$$E[W\varphi(W)] = E[(W(\mathbf{g}) - W(\widehat{\mathbf{g}}))\,\varphi(W(\mathbf{g}))]$$

Poincaré Inequality

Take $E[W]=0, \varphi$ a smooth function, $\widehat{\mathbf{g}}$ an independent copy of $\mathbf{g}$, and let $\widehat{\mathbf{g}}_t = e^{-t}\mathbf{g} + \sqrt{1-e^{-2t}}\widehat{\mathbf{g}}$

$$E[W\varphi(W)] \quad = \quad E[(W(\widehat{\mathbf{g}}_0) - W(\widehat{\mathbf{g}}_\infty))\,\varphi(W(\mathbf{g}))]$$

Poincaré Inequality

Take $E[W]=0, \varphi$ a smooth function, $\widehat{\mathbf{g}}$ an independent copy of $\mathbf{g}$, and let $\widehat{\mathbf{g}}_t = e^{-t}\mathbf{g} + \sqrt{1-e^{-2t}}\widehat{\mathbf{g}}$

$$\begin{aligned} E[W\varphi(W)] &= E[(W(\widehat{\mathbf{g}}_0) - W(\widehat{\mathbf{g}}_\infty))\,\varphi(W(\mathbf{g}))] \\ &= -\int_0^\infty \frac{d}{dt}\mathbf{E}[W(\widehat{\mathbf{g}}_t)\varphi(W(\mathbf{g}))]dt \end{aligned}$$

Poincaré Inequality

Take $E[W] = 0, \varphi$ a smooth function, $\widehat{\mathbf{g}}$ an independent copy of $\mathbf{g}$, and let $\widehat{\mathbf{g}}_t = e^{-t}\mathbf{g} + \sqrt{1-e^{-2t}}\widehat{\mathbf{g}}$

$$\begin{aligned} E[W\varphi(W)] &= E[(W(\mathbf{g}_0) - W(\widehat{\mathbf{g}}_\infty))\,\varphi(W(\mathbf{g}))] \\ &= -\int_0^\infty \frac{d}{dt}\mathbf{E}[W(\widehat{\mathbf{g}}_t)\varphi(W(\mathbf{g}))]dt \\ &= \cdots = E[T\varphi'(W(\mathbf{g}))], \end{aligned}$$

$$T = \int_0^\infty e^{-t}\langle \nabla W(\mathbf{g}), \widehat{E}(\nabla W(\widehat{\mathbf{g}}_t))\rangle dt.$$

Take $\varphi(x) = x$, obtain $\mathrm{Var}(W) = E[T]$, use Cauchy Schwarz.

Stein's Method

Stein's method is based on an equation characterizing the target distribution.

Stein's lemma: $Z \sim \mathcal{N}(0,1)$ if and only if

$$E[Z\varphi(Z)] = E[\varphi'(Z)]$$

for all absolutely continuous functions for which these expectations exist.

For test function h, solve the Stein equation

$$\varphi'(w) - w\varphi(w) = h(w) - Eh(Z)$$

for φ, evaluate at W, variable whose distribution is to be approximated, and bound expectation

$$Eh(W) - Eh(Z) = E[\varphi'(W) - W\varphi(W)]$$

Stein's Method

Stein's method is based on an equation characterizing the target distribution.

Stein's lemma: $Z \sim \mathcal{N}(0,1)$ if and only if

$$E[Z\varphi(Z)] = E[\varphi'(Z)]$$

for all absolutely continuous functions for which these expectations exist.

For test function h, solve the Stein equation

$$\varphi'(w) - w\varphi(w) = h(w) - Eh(Z)$$

for φ, evaluate at W, variable whose distribution is to be approximated, and bound expectation

$$Eh(W) - Eh(Z) = E[\varphi'(W) - W\varphi(W)]$$

Stein's Method

Stein's lemma: $Z \sim \mathcal{N}(0,1)$ if and only if

$$E[Z\varphi(Z)] = E[\varphi'(Z)].$$

Suppose for mean zero, variance one variable W one can find T such that

$$E[W\varphi(W)] = E[T\varphi'(W)], \quad \text{and note } E[T] = 1.$$

For a continuous test function $h : \mathbb{R} \to [0,1]$, solution of the Stein equation φ,

$$\begin{aligned} |Eh(W) - Eh(Z)| = |E[\varphi'(W) - W\varphi(W)]| &= |E[\varphi'(W)(1-T)]| \\ &\le \|\varphi'\| E|T-1| \le 2\sqrt{\mathrm{Var}(T)}. \end{aligned}$$

Taking sup over all such h, obtain bound in total variation.

Stein's Method

Stein's lemma: $Z \sim \mathcal{N}(0,1)$ if and only if

$$E[Z\varphi(Z)] = E[\varphi'(Z)].$$

Suppose for mean zero, variance one variable W one can find T such that

$$E[W\varphi(W)] = E[T\varphi'(W)], \quad \text{and note } E[T] = 1.$$

For a continuous test function $h : \mathbb{R} \to [0,1]$, solution of the Stein equation φ,

$$\begin{aligned} |Eh(W) - Eh(Z)| = |E[\varphi'(W) - W\varphi(W)]| &= |E[\varphi'(W)(1-T)]| \\ &\le \|\varphi'\| E|T-1| \le 2\sqrt{\mathrm{Var}(T)}. \end{aligned}$$

Taking sup over all such h, obtain bound in total variation.

2^{nd} order Poincaré inequality

Recall from proof of Poincaré inequality, where $E[W] = 0$ and $W = W(\mathbf{g})$,

$$E[W\varphi(W)] = E[T\varphi'(W)] \text{ with } T = \int_0^\infty e^{-t}\langle \nabla W(\mathbf{g}), \widehat{E}(\nabla W(\widehat{\mathbf{g}}_t))\rangle dt.$$

Chatterjee (2009), for this T and $\mathrm{Var}(W) = 1$ we have

$$d_{\mathrm{TV}}(W, Z) \leq 2\sqrt{\mathrm{Var}(T)}.$$

Variance might be difficult to evaluate directly, so use the Poincaré inequality.

2^{nd} order Poincaré inequality

Recall from proof of Poincaré inequality, where $E[W] = 0$ and $W = W(\mathbf{g})$,

$$E[W\varphi(W)] = E[T\varphi'(W)] \text{ with } T = \int_0^\infty e^{-t}\langle \nabla W(\mathbf{g}), \widehat{E}(\nabla W(\widehat{\mathbf{g}}_t))\rangle dt.$$

Chatterjee (2009), for this T and $\mathrm{Var}(W) = 1$ we have

$$d_{\mathrm{TV}}(W, Z) \leq 2\sqrt{\mathrm{Var}(T)}.$$

Variance might be difficult to evaluate directly, so use the Poincaré inequality.

Gaussian Projections on Convex Cones

Want to obtain limiting normal law, with bounds, for $G_C = \|\Pi_C(\mathbf{g})\|^2$ in order to infer same for V_C.

$$\nabla\|\Pi_C(x)\|^2 = 2\Pi_C(x) \quad \text{hence} \quad T = 4\int_0^\infty e^{-t}\langle \Pi_C(\mathbf{g}), \widehat{E}\Pi_C(\widehat{\mathbf{g}}_t)\rangle dt$$

Use the Poincaré inequality to bound $\mathrm{Var}(T)$, and that $\sigma_C^2 = 2\delta_C + \tau_C^2$, to obtain

$$d_{\mathrm{TV}}(G_C, Z) \leq \frac{16\sqrt{\delta(C)}}{\sigma^2} \leq \frac{8}{\sqrt{\delta(C)}}.$$

Berry-Esseen bound for conic intrinsic volumes

If C is a closed convex cone with $\delta_C = E[V_C]$ and $\tau_C^2 = \mathrm{Var}(V_C) > 0$, then with $\alpha = \tau_C^2/\delta_C$, for $\delta_C \geq 8$

$$\sup_{u \in \mathbb{R}} \left| P\left[\frac{V_C - \delta_C}{\tau_C} \leq u\right] - P[Z \leq u] \right| \leq h(\delta_C) + \frac{48}{\sqrt{\alpha \log^+ \left(\alpha\sqrt{2}\delta_C\right)}}$$

where

$$h(\delta_C) = \frac{1}{72}\left(\frac{\log \delta_C}{\delta_C^{3/16}}\right)^{5/2}$$

For m about $\delta_C + t\tau_C$, obtain bound on

$$\sup_{t \in \mathbb{R}} \left| P\left\{\mathbf{x}_0 \text{ is recovered}\right\} - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-u^2/2} du \right|$$

Berry-Esseen bound for conic intrinsic volumes

If C is a closed convex cone with $\delta_C = E[V_C]$ and $\tau_C^2 = \mathrm{Var}(V_C) > 0$, then with $\alpha = \tau_C^2/\delta_C$, for $\delta_C \geq 8$

$$\sup_{u\in\mathbb{R}} \left| P\left[\frac{V_C - \delta_C}{\tau_C} \leq u\right] - P[Z \leq u] \right| \leq h(\delta_C) + \frac{48}{\sqrt{\alpha \log^+ (\alpha\sqrt{2}\delta_C)}}$$

where

$$h(\delta_C) = \frac{1}{72}\left(\frac{\log \delta_C}{\delta_C^{3/16}}\right)^{5/2}$$

For m about $\delta_C + t\tau_C$, obtain bound on

$$\sup_{t\in\mathbb{R}} \left| P\{\mathbf{x}_0 \text{ is recovered}\} - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t} e^{-u^2/2} du \right|$$

Ingredients

- Convex Geometry
- Compressed sensing
- Gaussian inequalities and Stein's Method
  - Poincaré inequality – bound a variance by an expectation.
  - Log Sobolev inequality
  - 2^{nd} order Poincaré inequality – bound a total variation by a variance, which is bounded by an expectation.