Minimizing the Condition Number of a Gram Matrix on the Sphere

Xiaojun Chen

Hong Kong Polytechnic University

Joint work with

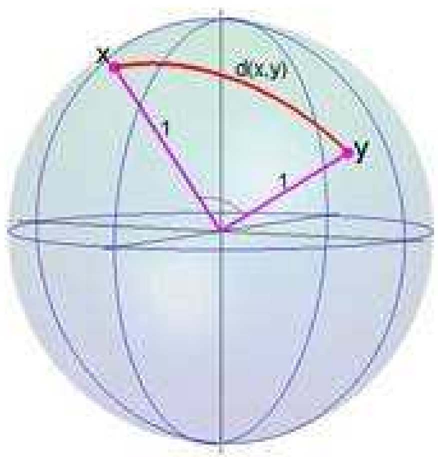
Robert S. Womersley (Univ. of New South Wales, Australia)

Jane Ye (Univ. of Victoria, Canada)

23-27 October, NASC2010, Beijing

Least squares approximation on the sphere

$\mathbb{S}^2 = \{\mathbf{z} \in R^3 \ : \ \|\mathbf{z}\|_2 = 1\,\}$,   Area $|\mathbb{S}^2| = 4\pi$



$\mathbb{P}_t$: the linear space of restrictions of polynomials of degree $\leq t$ in 3 variables to $\mathbb{S}^2$.

dim $\mathbb{P}_t = (t+1)^2$

Gram matrix

$\mathbb{P}_t$ can be spanned by an orthonormal set of real spherical harmonics with degree ℓ and order k,

$$\{\, Y_{\ell k} \,|\, k = 1, \ldots, 2\ell + 1, \ell = 0, 1, \ldots, t\}.$$

Let $X_N = \{\mathbf{z}_1, \ldots, \mathbf{z}_N\} \subset S^2$ be a set of N-points on the sphere.
The Gram matrix is defined as

$$G_t(X_N) = Y(X_N)^T Y(X_N),$$

where $Y(X_N) \in R^{(t+1)^2 \times N}$ and the j-th column of $Y(X_N)$ is given by

$$Y_{\ell k}(\mathbf{z}_j), \qquad k = 1, \ldots, 2\ell + 1, \quad \ell = 0, 1, \ldots, t.$$

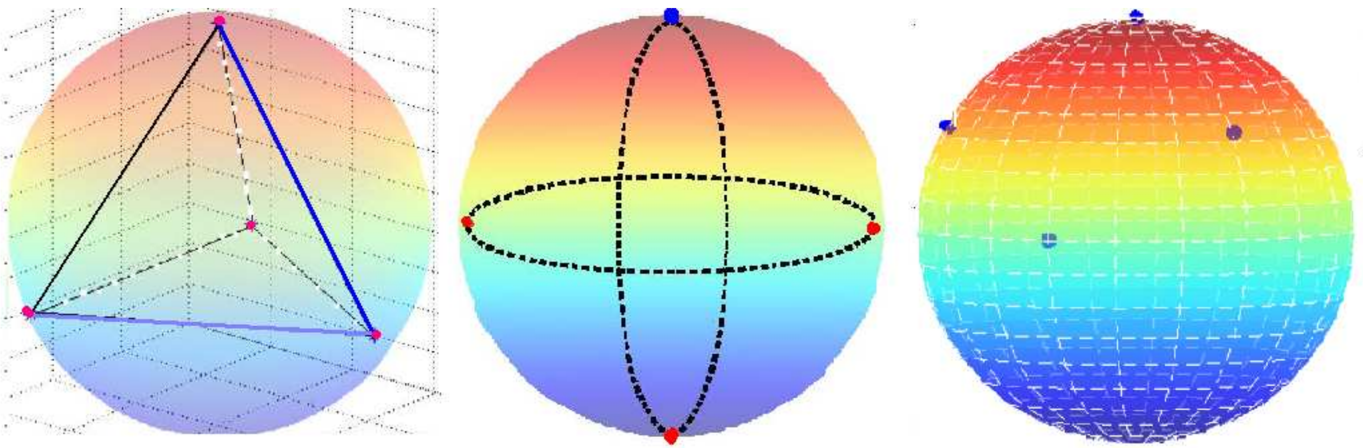
The Gram matrix G_t is a function of a set of N-points X_N.

Distribution of points on the sphere

$t = 1$,   dim$P_1 = 4$, $X_N = \{\mathbf{z}_1, \ldots, \mathbf{z}_4\} \subset S^2$,

$$G_1(X_N) \in R^{4\times 4}$$

$\det(G_1(X_N)) = \frac{1}{\pi^4}$	$\det(G_1(X_N)) = 0$	$\det(G(X_N)) = 0$



Regular tetrahedron: $G_1(X_N) = \frac{1}{\pi} I_4$,   cond$(G_1(X_N)) = 1$.

Four sets of points on the sphere

minimimum energy system $\qquad \operatorname{argmin} \sum_{i \neq j}^{N} \frac{1}{\|\mathbf{z}_i - \mathbf{z}_j\|}$

extremal system $\qquad \operatorname{argmax} \det(G_t(X_N))$

spherical design $\qquad \int_{\mathbb{S}^2} p(\mathbf{z}) d\mathbf{z} = \frac{4\pi}{N} \sum_{i=1}^{N} p(\mathbf{z}_i), \quad \forall p \in \mathbb{P}_t$

minimum cond points $\qquad$ optimal solution of problem:

$$\min \frac{\lambda_{\max}(G_t(X_N))}{\lambda_{\min}(G_t(X_N))}$$

Spherical t-Design

Definition 1 (Delsarte-Goethals-Seidel 1977)
A spherical t-design is a set of N points $X_N = \{\mathbf{z}_1, \ldots, \mathbf{z}_4\} \subset S^2$ such that

$$\frac{1}{4\pi} \int_{S^2} p(\mathbf{z}) d\mathbf{z} = \frac{1}{N} \sum_{i=1}^{N} p(\mathbf{z}_i)$$

for every polynomial $p \in \mathtt{P}_t$.

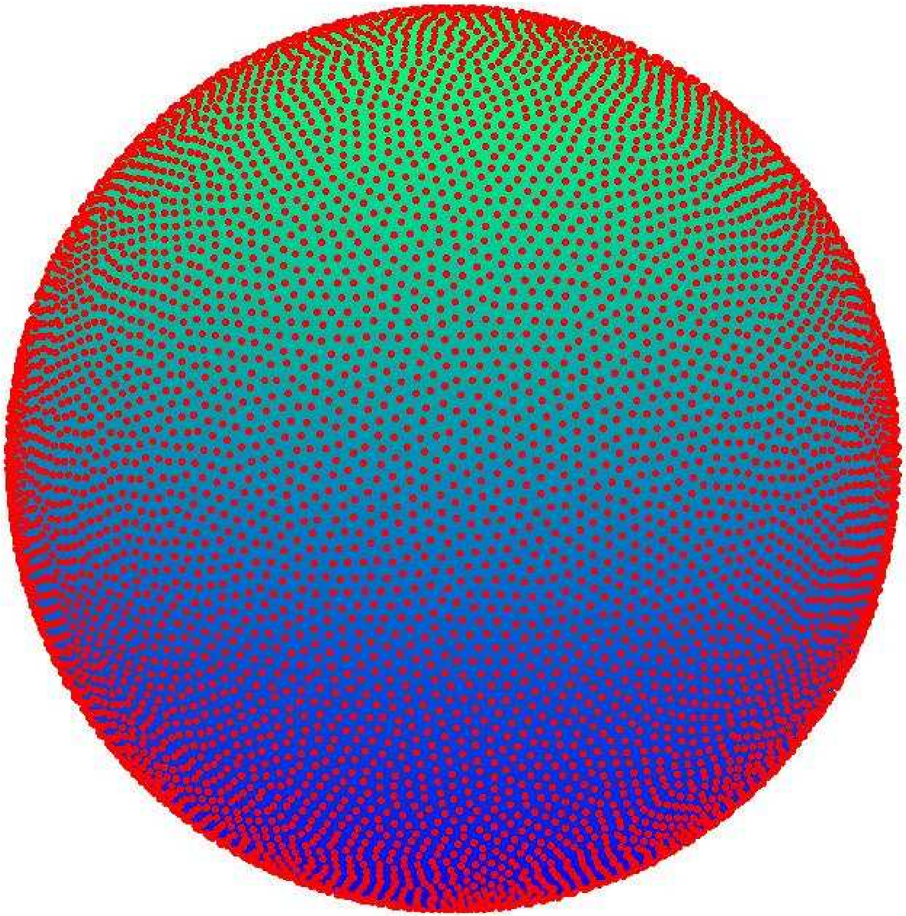
- The average value of $p \in \mathtt{P}_t$ on the whole sphere equals the average value of p on the set.
- The equally weighted cubature rule is exact for all $p \in \mathtt{P}_t$.

No answer to what is the number of points needed to construct a spherical t-design for any $t \geq 1$ **?** Whether $N = O(t^2)$ as $t \to \infty$ **?**

Can we guarantee the existence of spherical t-designs with $(t+1)^2$ points and well-conditioned Gram matrices **?**

Spherical 100-design

Chen-Frommer-Lang, to appear in Numer. Math.



Reformulation I: Parametrization

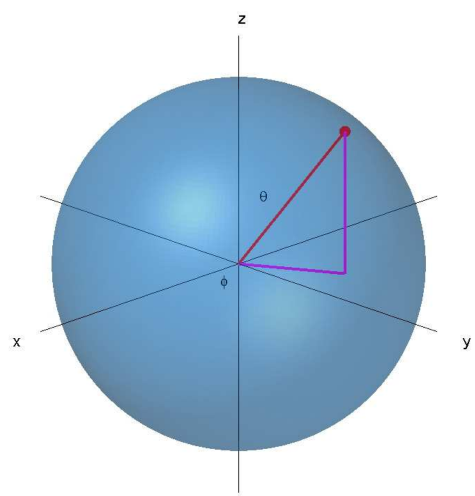
$$N = (t+1)^2, \quad m = 2N - 3, \quad X_N = \{\mathbf{z}_1, \dots, \mathbf{z}_4\} \subset S^2.$$

Represent $\mathbf{z}_i \in X_N \subset S^2$ using polar coordinates with angles θ_i, ϕ_i.

$$\mathbf{z}_i = [\sin(\theta_i)\cos(\phi_i), \quad \sin(\theta_i)\sin(\phi_i), \quad \cos(\theta_i)]^T, \quad \theta_1 = 0, \phi_1 = \phi_2 = 0$$

Fix $\mathbf{z}_1$ on the north pole and $\mathbf{z}_2$ on the zero meridian.

$x_\theta = [\theta_2, \dots, \theta_N]^T, \quad x_\phi = [\phi_3, \dots, \phi_N]^T, x = [x_\theta^T, x_\phi^T]^T \in R^m$



Reformulation II, Gram matrix

Define the Legendre polynomials by the recurrence

$p_0(u) = 1$
$p_1(u) = z$
$\ell p_\ell(z) = (2\ell - 1) u p_{\ell-1}(u) - (\ell - 1) p_{\ell-2}(u)$

for $\ell = 2, \ldots, t, \quad u \in [-1, 1]$.

Define the Jacobi polynomials

$$J_t(u) = \sum_{\ell=0}^{t} (2\ell + 1) p_\ell(u)$$

Define the Gram matrix $G(x) \in R^{N \times N}$

$$G_{i,j}(X_N(x)) = J_t(\mathbf{z}_i(x)^T \mathbf{z}_j(x))$$

Euclidean condition number

$V \in R^{\ell \times n}$ with $\ell \geq n$, and rank$(V) = n$. Let $A = V^T V$

The Euclidean condition number of V is defined by

$$\kappa(V) = \max_{y \neq 0} \frac{\|y\|}{\|Vy\|} \max_{z \neq 0} \frac{\|Vz\|}{\|z\|} = \|V\| \|V^\dagger\| = \sqrt{\kappa(A)} = \frac{\sqrt{\lambda_1(A)}}{\sqrt{\lambda_n(A)}},$$

$V^\dagger = (V^T V)^{-1} V^T$ is the Moore-Penrose generalized inverse of V.

Suppose each entry of $V(x)$ is a continuously differentiable function of $x \in R^m$. Let $A(x) = V(x)^T V(x)$.

We consider

$$\begin{array}{ll} \text{minimize} & \kappa(A(x)) \\ \text{subject to} & x \in \mathcal{X}, \end{array} \qquad (1)$$

where $\mathcal{X}$ is a convex set in R^m.

A small Vandermonde-like matrix

$$V(x) = \begin{pmatrix} 1 & -x \\ 1 & 0 \\ 1 & x \end{pmatrix}, \quad A(x) = V(x)^T V(x) = \begin{pmatrix} 3 & 0 \\ 0 & 2x^2 \end{pmatrix}$$

$$f(x) = \frac{\lambda_1(A(x))}{\lambda_n(A(x))} = \begin{cases} \dfrac{3}{2x^2}, & 0.5 \le x \le \sqrt{1.5} \\ \dfrac{2x^2}{3}, & \sqrt{1.5} \le x \le 1.5. \end{cases}$$

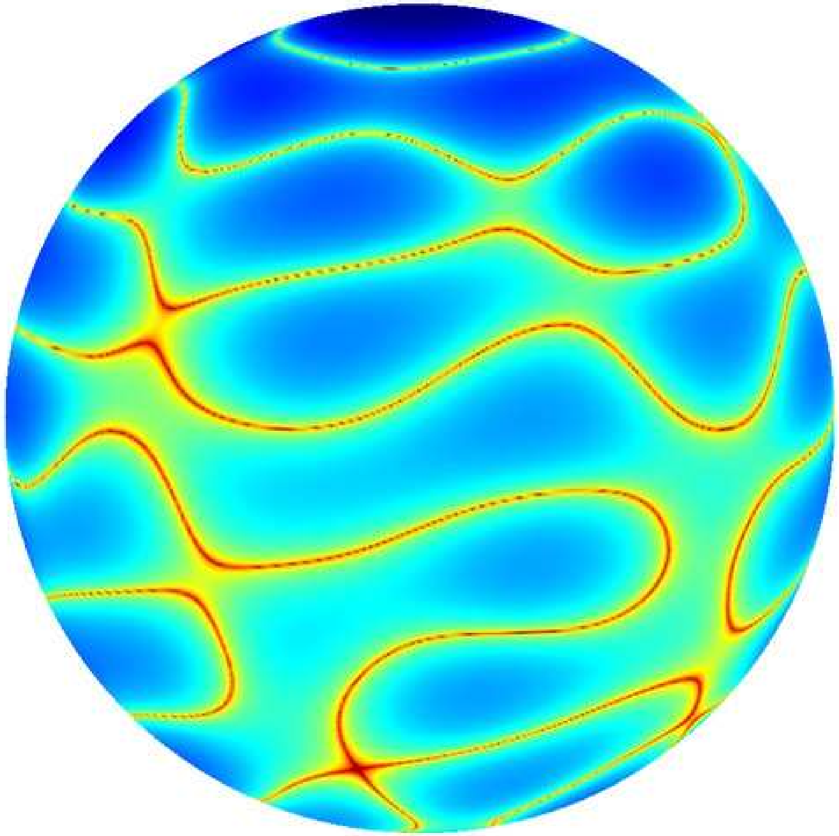
We consider

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in [0.5, 1.5]. \end{array}$$

$x^* = \sqrt{1.5}$ is the minimizer. The Clarke generalized gradient at x^* is

$$\partial f(x^*) = \operatorname{conv}\left\{ -2\sqrt{\frac{2}{3}}, \quad 2\sqrt{\frac{2}{3}} \right\} = 2\sqrt{\frac{2}{3}}[-1, 1].$$

An example on the sphere



The condition number for degree $t = 9$, $N = 100$ extremal points. The plot is of the condition number of the Gram matrix G when the point at the north pole is moved to locations over the whole sphere.

Least squares approximation on $[-1,1]$

Let $\{p_j,\ j=0,\ldots,n-1\}$ be a basis for $\mathbb{P}_{n-1}[-1,1]$, the linear space of polynomials of degree $\leq n-1$.
For given ℓ distinct real numbers

$$-1 \leq a_1 < a_2 < \ldots < a_\ell \leq 1$$

we consider the weighted Vandermonde-like matrix

$$V(w,a) = \begin{pmatrix} w_1 p_0(a_1) & w_1 p_1(a_1) & w_1 p_2(a_1) & \ldots & w_1 p_{n-1}(a_1) \\ w_2 p_0(a_2) & w_2 p_1(a_2) & w_2 p_2(a_2) & \ldots & w_2 p_{n-1}(a_2) \\ \vdots & \vdots & \vdots & & \vdots \\ w_\ell p_0(a_\ell) & w_\ell p_1(a_\ell) & w_\ell p_2(a_\ell) & \ldots & w_\ell p_{n-1}(a_\ell) \end{pmatrix}.$$

$$w = (w_1,\ldots,w_\ell)^T \quad \text{and} \quad a = (a_1,\ldots,a_\ell)^T.$$

Generalized gradient

Let $d(x)$ be the multiplicity of $\lambda_1(A(x))$, and $b(x)$ be the multiplicity of $\lambda_n(A(x))$. Let $A(x)$ admit an eigenvalue decomposition

$$A(x) = U(x)\mathrm{diag}(\lambda(A(x)))U(x)^T$$

with $U(x)^T U(x) = I$. Denote

$$U_\alpha = (u_1(x), \ldots, u_{d(x)}(x)), \quad \text{and} \quad U_\beta = (u_{n-b(x)+1}(x), \ldots, u_n(x)).$$

Suppose that rank$(V(x)) = n$. Then f is Clarke regular and the Clarke generalized gradient $\partial f(x)$ is

$$\{g \in R^m : g_k = \frac{1}{\lambda_n(A(x))}\langle U_\alpha^T A_k(x) U_\alpha, P_\alpha\rangle - \frac{\kappa(A(x))}{\lambda_n(A(x))}\langle U_\beta^T A_k(x) U_\beta, P_\beta\rangle$$

$$k = 1, \ldots, m$$

$$\text{where } P_\alpha \in D^+_{d(x)}, \mathrm{tr}(P_\alpha) = 1, P_\beta \in D^+_{b(x)}, \mathrm{tr}(P_\beta) = 1\}.$$

Some properties

1. $f(x) = \kappa(A(x))$ is strongly semismooth on $\mathcal{X}$.

2. Suppose that $V(x)$ is a linear mapping of x on $\mathcal{X}$. Then $\lambda_1(A(x))$ with $A(x) = V(x)^T V(x)$ is a convex function on $\mathcal{X}$.

3. Let B be a fixed $m \times n$ matrix with $m \geq n$ and rank$(B) = n$. Define

$$h(W) = \kappa(B^T W B), \quad W \in S_m^{++}.$$

Then h is quasi-convex and strongly pseudo-convex.

Example: $V(x) = XB$, where $X \in D_m^{++}$ with diagonal elements $x_i, i = 1, \ldots, m$, $A(V(x)) = B^T X^T X B = B^T W B$.

Smoothing function I

We introduce the smoothing function of the condition number as follows:

$$\tilde{f}(x,\mu) = -\frac{\ln(\sum_{i=1}^{n} e^{\lambda_i(A(x))/\mu})}{\ln(\sum_{i=1}^{n} e^{-\lambda_i(A(x))/\mu})}. \qquad (2)$$

In numerical computations, we use an equivalent formula

$$\tilde{f}(x,\mu) = \frac{\lambda_1(A(x)) + \mu\ln(\sum_{i=1}^{n} e^{(\lambda_i(A(x))-\lambda_1(A(x)))/\mu})}{\lambda_n(A(x)) - \mu\ln(\sum_{i=1}^{n} e^{(\lambda_n(A(x))-\lambda_i(A(x)))/\mu})},$$

which is more numerically stable than (2).

Smoothing function II

(i) $\tilde{f}(\cdot, \mu)$ is continuously differentiable for any fixed $\mu > 0$.

(ii) There exists $c > 0$ such that for any $x \in \mathcal{X}$ and $\mu \leq \dfrac{\lambda_n}{2 \ln n}$

$$0 \leq \tilde{f}(x, \mu) - f(x) \leq c\mu. \tag{3}$$

$$\lim_{x \to \bar{x}, \mu \downarrow 0,} \tilde{f}(x, \mu) = f(\bar{x}).$$

(iii) For any $\bar{x} \in \mathcal{X}$, $\{\lim_{x \to \bar{x},\ \mu \downarrow 0} \nabla_x \tilde{f}(x, \mu)\}$ is nonempty and bounded.

$$\{\lim_{x \to \bar{x}, \mu \downarrow 0,} \nabla_x \tilde{f}(x, \mu)\} \subset \partial f(\bar{x}).$$

(iv) For any fixed $\mu > 0$, there exists a constant L_μ such that

$$\|\nabla \tilde{f}(x, \mu) - \nabla \tilde{f}(y, \mu)\| \leq L_\mu \|x - y\|. \tag{4}$$

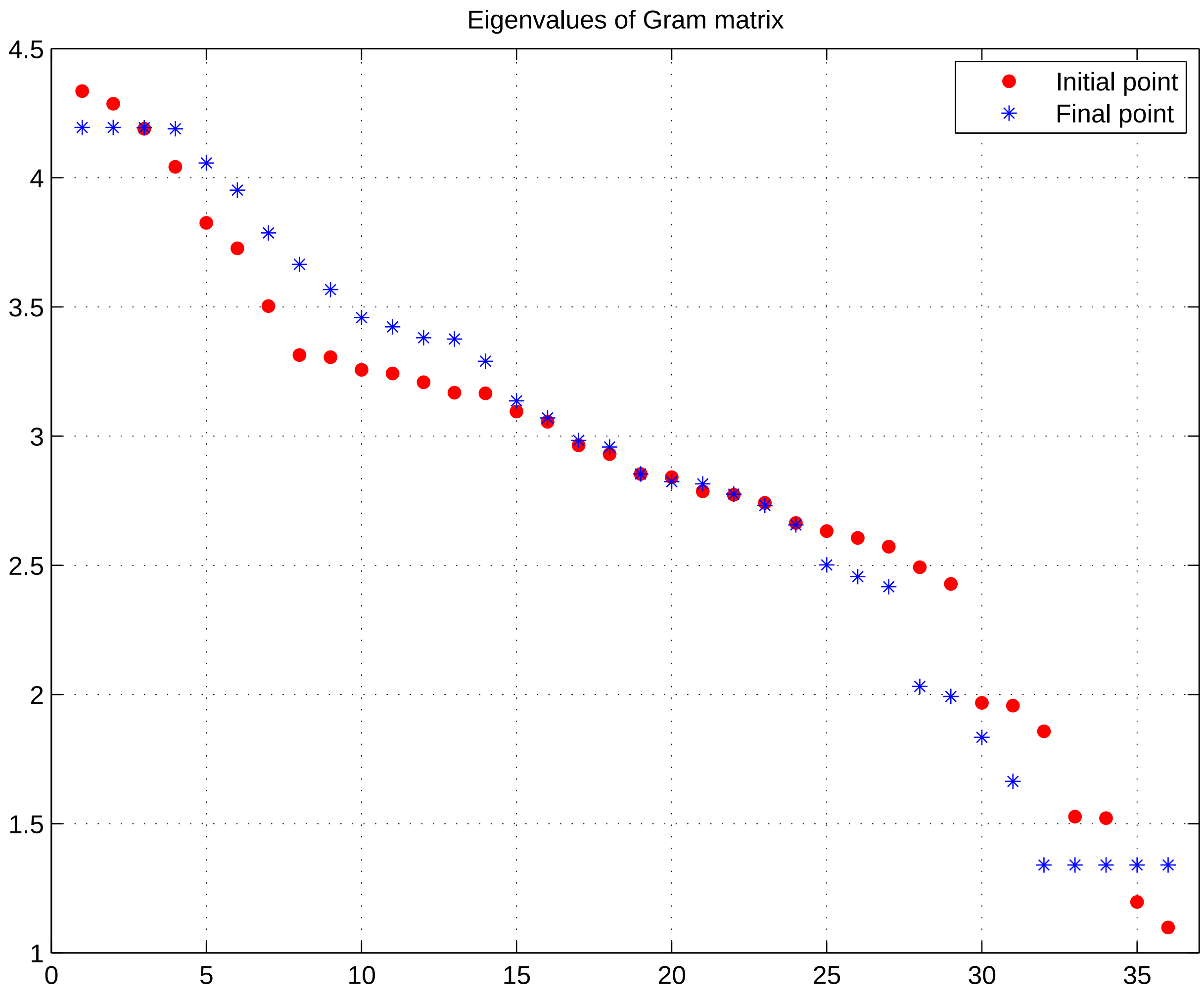
Convergence

Theorem From any staring point $x^0 \in \mathcal{X}$, the sequence $\{x^k\}$ generated by the SPG method is contained in $\mathcal{X}$ and any accumulation point $\bar{x}$ of $\{x^k\}$ is a Clarke stationary point, that is, there is $g \in \partial f(\bar{x})$ such that

$$\langle g, x - \bar{x} \rangle \geq 0, \quad \forall x \in \mathcal{X}.$$

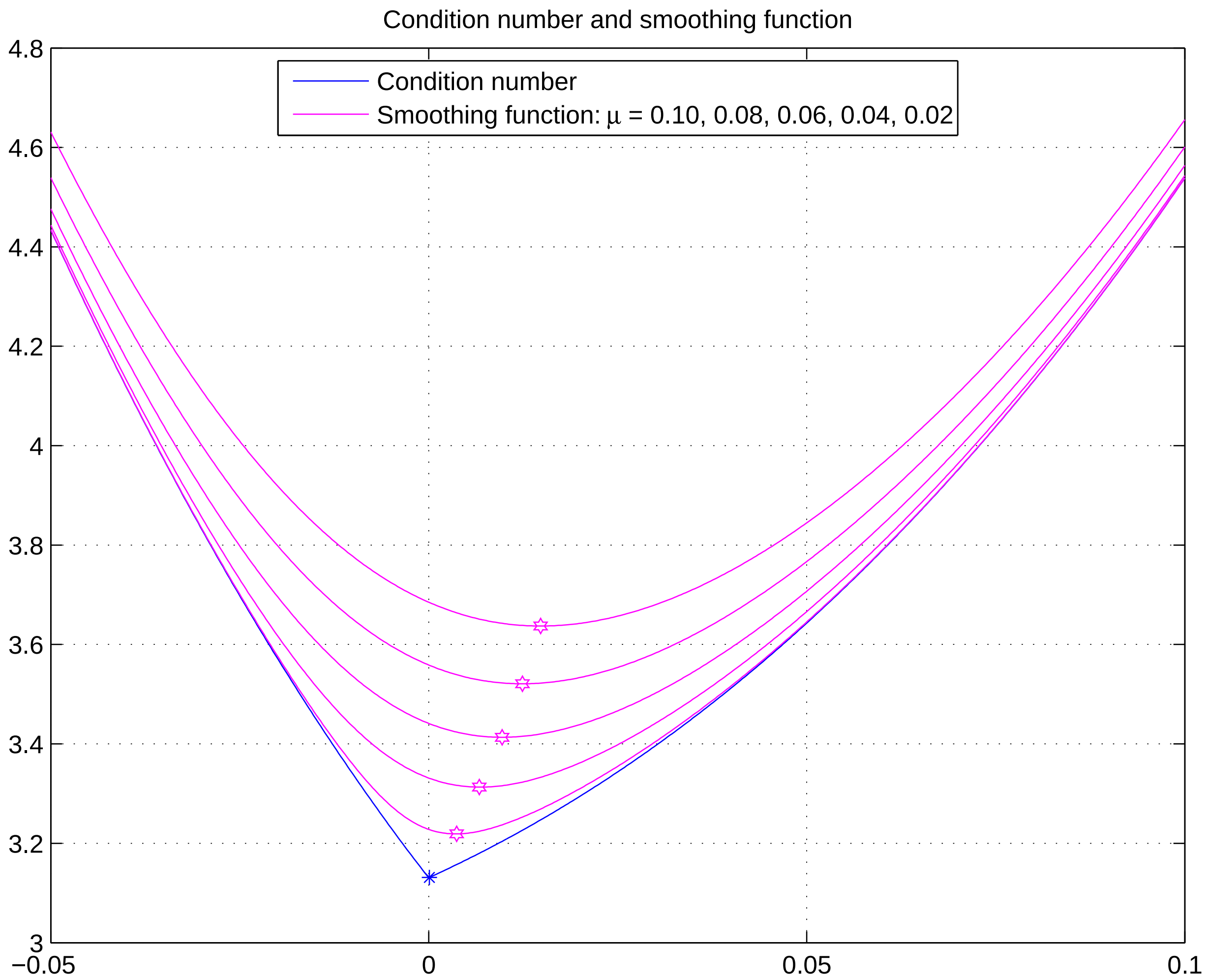
Corollary If the function $f(x)$ is pseudoconvex in a neighborhood $B(\bar{x}) \subset \mathcal{X}$, then the accumulation point is a local optimal solution and if the function $f(x)$ is pseudoconvex on $\mathcal{X}$, then the accumulation point is a global optimal solution over $\mathcal{X}$.

Eigenvalues of a Gram matrix on the sphere



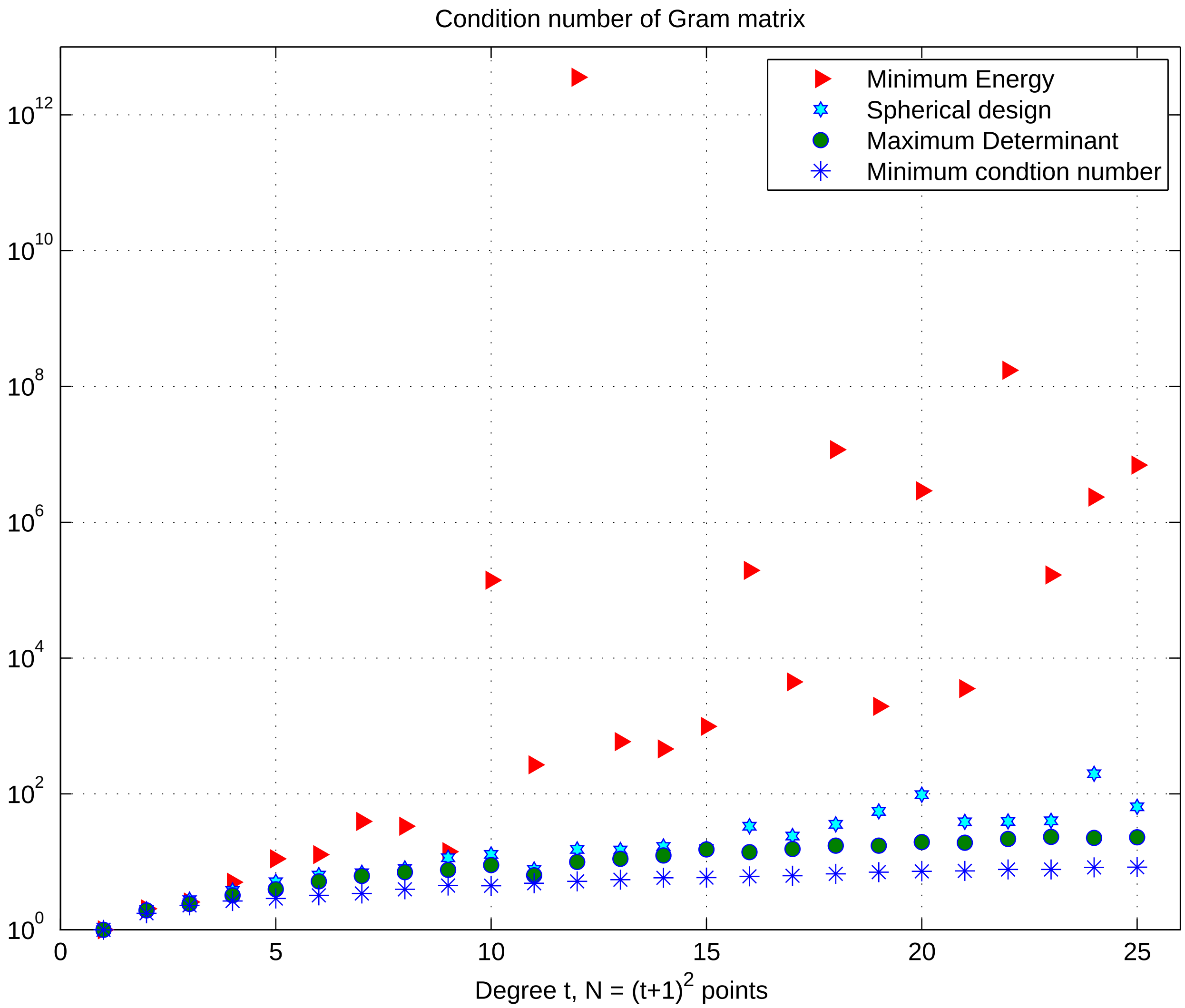
The eigenvalues of 36×36 Gram matrix for degree 5, with 36 points on the sphere.

Smoothing function



For the same Gram matrix, $f(x^* + \alpha \nabla_x \tilde{f}(x^*, 0.0766))$, $\alpha \in [-0.05, 0.1]$.

Condition number on the four sets of points



References

1. X. Chen and R.S. Womersley, Existence of solutions to systems of underdetermined equations and spherical designs, SIAM J. Numer. Anal. 44(2006) 2326-2341.
2. X. Chen, A. Frommer and B. Lang, Computational existence proofs for spherical t-designs, to appear in Numer. Math.

$$\mathtt{www.math.uni-wuppertal.de/SciComp/SphericalTDesigns}$$

3. C. An, X. Chen, I.H. Sloan and R.S. Womersley, Well-conditioned spherical designs on 2-sphere, to appear in SIAM J. Numer. Anal.
4. X. Chen, R.S. Womersely and J. Ye, Minimizing the condition number of a Gram matrix, to appear in SIAM J. Optim.

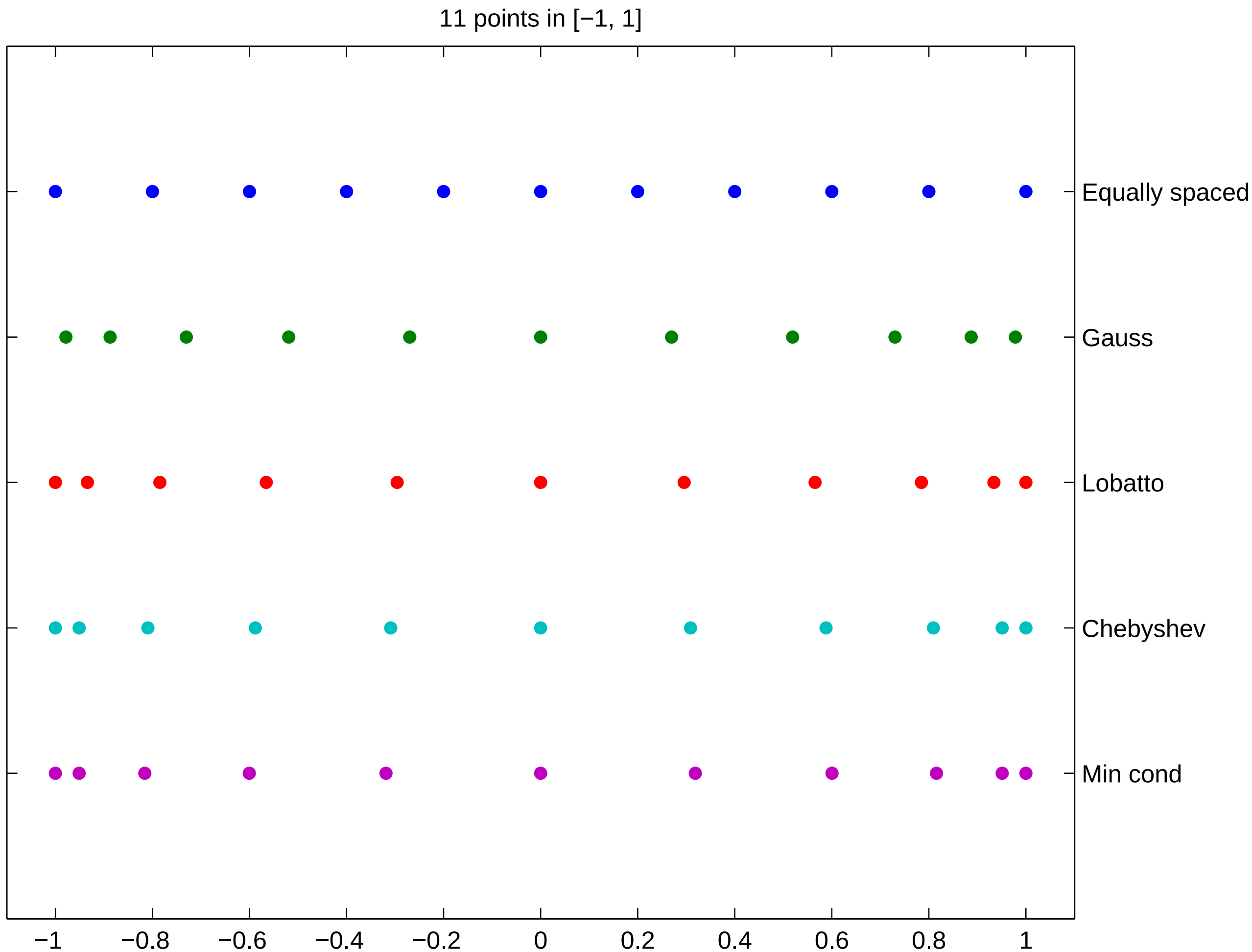
Thank you very much

Five sets of points in the interval $[-1, 1]$

equally spaced points	$a_i = -1 + \frac{2(i-1)}{\ell - 1}, \quad i = 1, \ldots, \ell.$
Gauss points	$\int_{-1}^{1} p(\tau) d\tau = \sum_{i=0}^{n-1} \alpha_i p(a_i), \quad \forall p \in \mathbb{P}_{2n}$
Gauss Lobatto points	$a = \operatorname{argmax} \det(V(x)^T V(x)).$
Chebyshev points	$a_i = \cos \frac{\pi(2i-1)}{2\ell}, \quad i = 1, \ldots, \ell.$
minimum cond points	$a =$ optimal solution of (1) .

Here $\alpha_i, i = 0, \ldots, n-1$ are the integral values of the Lagrange interpolation polynomials on $[-1, 1]$.

Distribution of points



condition number and determinant

	condition number	determinant
equally spaced points	1.946479e+008	5.755277e-022
Gauss points	1.767123e+007	4.616572e-020
Lobatto points	9.606328e+006	7.968101e-019
Chebyshev points	8.307060e+006	6.310887e-019
min cond points	8.176691e+006	5.826573e-019

Values of the condition number and determinant at equally spaced points, Gauss Points, Gauss Lobatto points, Chebyshev points and minimum cond points