Discretisation sensitivity in gradient-based optimisation methods on continuous Hilbert spaces

Tobias Schwedes,
Department of Mathematics, Imperial College London

Mathematics of Planet Earth

Imperial College London

University of Reading

Problem framework

When solving optimality problems on Hilbert function spaces, existing implementations of optimisation methods typically assume that the problem is posed in Euclidean space. For gradient-based methods, functional derivatives are then inaccurately represented with respect to ℓ^2 instead of the inner product induced by the function space. This error manifests as a mesh dependence of the discretised problem in the iteration number required to solve the problem. Here, the magnitude of the mesh dependence is quantified by an analytical expression for the iteration count of a steepest descent method in a finite element discretised version of a continuous optimisation problem.

Inner product representation of derivatives

Let $J : H \to \mathbb{R}$ be Fréchet differentiable on a Hilbert space $(H, \langle \cdot, \cdot \rangle_H)$. For any $u \in H$, let $J'(u) \in \mathcal{L}(H, \mathbb{R}) = H^*$ state the Fréchet derivative of J at u. According to the Riesz representation theorem, there is a unique $\hat{u} \in H$ such that

$$J'(u)v = \langle \hat{u}, v \rangle_H$$

for all $v \in H$. The map $\mathcal{R}_H : H^* \to H$ that sends $J'(u)$ to its unique representative $\hat{u} \in H$ is called Riesz-map, and $\mathcal{R}_H(J'(u)) = \hat{u}$ is called Riesz-representer of $J'(u)$ in H. Thus, the derivative of J can be represented with respect to the inner product corresponding to the Hilbert space it is defined on. Depending on the space we assume as the domain of J, there are a number of ways to represent J':

$$\begin{aligned} J'(u)\delta u &= \mathcal{R}_{\ell^2}(J(u)) \cdot \delta u \\ &= \langle \mathcal{R}_{L^2}(J(u)), \delta u \rangle_{L^2} \\ &= \langle \mathcal{R}_{H^1}(J(u)), \delta u \rangle_{H^1} \end{aligned}$$

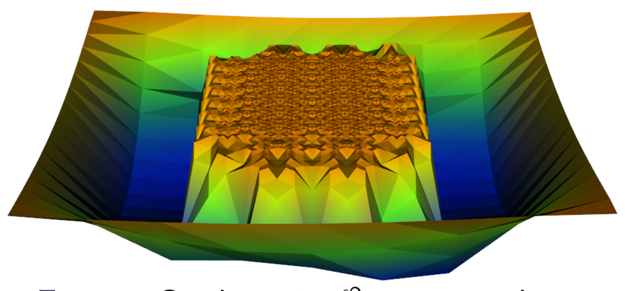


Figure: Gradient in ℓ^2 inner product

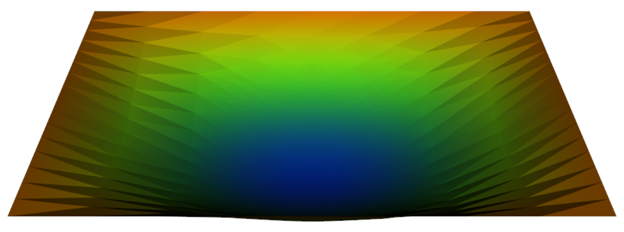


Figure: Gradient in L^2 inner product

Problem definition

To discuss the influence of the choice of the inner product upon the convergence of an optimisation method, a problem as simple as possible, at the same time being generic in this context, is considered:

$$\min_{u \in H} \frac{\alpha}{2}\langle u, u \rangle_H + \beta \langle u, w \rangle_H + \gamma,$$

where $\alpha > 0$, $\beta, \gamma \in \mathbb{R}$ and $w \in H$. Further, $H \subset \{u : \Omega \to \mathbb{R}\}$, with bounded domain $\Omega \subset \mathbb{R}^n$ and $n \in \mathbb{N}$, is a Hilbert function space with inner product $\langle \cdot, \cdot \rangle_H$ that is of some integral type. For instance, H could be a Hilbert Sobolev space, i.e. $H = W^{k,2}$ with $k \in \mathbb{N}$.

Since the minimising state u^* is invariant to adding constants to the functional to minimise, the above minimisation problem is equivalent to minimising

$$\frac{\alpha}{2}\langle u, u \rangle_H + \beta \langle u, w \rangle_H + \frac{\alpha}{2}\langle u^*, u^* \rangle_H = \frac{\alpha}{2}\| u - u^* \|_H^2 =: J(u).$$

Finite element discretisation of the problem

In order to compute the minimum of J numerically, we discretise our problem using the finite element method: any $u \in H$ can be approximated in finite dimensions by its so called global interpolant $\mathcal{I}_\mathcal{T}(u) \in V_h \subset H$ given by

$$\mathcal{I}_\mathcal{T}(u)(x) = \sum_{i=1}^{d} u_i \varphi_i(x),$$

where $\dim(V_h) = d$ is the number of degrees of freedom of the underlying discretisation of the domain (mesh). Further, φ_i, $i = 1, \dots, d$, are the so called global basis functions, i.e. linearly independent functions spanning V_h. Thus, the discretised problem can be expressed as

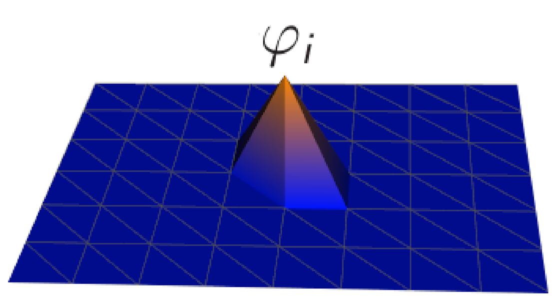


Figure: global basis function for linear Lagrange elements (CG1) on a uniform triangular 2-D mesh.

$$\min_{\mathcal{I}_\mathcal{T}(u)} J(\mathcal{I}_\mathcal{T}(u)) \equiv \min_{\mathbf{u} \in \mathbb{R}^d} \left\{ J(\mathbf{u}) = \frac{\alpha}{2}(\mathbf{u} - \mathbf{u}^*)^T M (\mathbf{u} - \mathbf{u}^*) \right\},$$

where $\mathbf{u} = (u_1, \dots, u_d)$ and $\mathbf{u}^* = (u_1^*, \dots, u_d^*)$ contain the coefficients of the global interpolants of u and u^*, respectively. Further, $M_{ij} = \langle \varphi_i, \varphi_j \rangle_H$ for any $i, j = 1, \dots, d$. Here, we identified $J|_{V_h}$ with the function $\mathbf{u} \mapsto \frac{\alpha}{2}(\mathbf{u} - \mathbf{u}^*)^T M (\mathbf{u} - \mathbf{u}^*)$.

Steepest descent using inner product induced by H

The only non-zero derivatives of J in its continuous formulation are given by

$$\begin{aligned} J'(u)\delta u &= \alpha \langle u - u^*, \delta u \rangle_H, \\ J''(u)\delta u \delta v &= \alpha \langle \delta u, \delta v, \rangle_H \end{aligned}$$

Hence, $\mathcal{R}_H(J'(u)) = \alpha(u - u^*)$. Applying Taylor's theorem for continuous spaces at the steepest descent iterate $u_1 = u_0 - \eta J'(u_0)$ for some arbitrary initial u_0 leads to

$$\begin{aligned} J(u_1) &= J(u_0) - \eta J'(u_0)\mathcal{R}_H(J'(u_0)) + \frac{1}{2}\eta^2 J''(u_0)\mathcal{R}_H^2(J'(u_0)) \\ &= (\frac{1}{2} - \alpha\eta + \frac{\alpha^2}{2}\eta^2)\alpha \| u_0 - u^* \|_H^2 \geq 0, \end{aligned}$$

such that it is easily seen that the minimum is found for the step size $\eta = 1/\alpha$ after only **one** iteration, independently of u_0! It is easily verified that this still holds true in the finite element discretised formulation.

Steepest descent using Euclidean inner product

Based on the finite element discretised version of J, one computes for its derivative,

$$\begin{aligned} J'(\mathbf{u})\delta\mathbf{u} &= (\mathbf{u} - \mathbf{u}^*)^T M \delta\mathbf{u} \\ &= \langle M(\mathbf{u} - \mathbf{u}^*), \delta\mathbf{u} \rangle_{\ell^2}. \end{aligned}$$

Thus, $\mathcal{R}_{\ell^2}(J') = \alpha M(\mathbf{u} - \mathbf{u}^*)$. Note that $\mathcal{R}_{\ell^2}(J')$ differs from $\mathcal{R}_H(J')$ by the the multiplication of the finite element matrix M. Since M reflects the structure of the underlying mesh, i.e. spacial distribution of elements and their sizes, it is expected that the convergence of the steepest descent method using $\mathcal{R}_{\ell^2}(J')$ is mesh-dependent. Applying exact line search, i.e. a step size η_k such that $\partial_{\eta_k} J(\mathbf{u_{k+1}}) = 0$, Kantorovich's lemma (LY08) yields an error estimate in the kth iteration,

$$|J(\mathbf{u}_k) - J(\mathbf{u}^*)| \leq \left[\frac{\kappa(M) - 1}{\kappa(M) + 1}\right]^{2k} J(\mathbf{u}_0),$$

where $\kappa(M)$ denotes the condition number of the matrix M, i.e. $\kappa(M) = \lambda_{\max}^M / \lambda_{\min}^M$, and $\lambda_{\max}^M$ and $\lambda_{\min}^M$ are the largest and smallest eigenvalues of M, respectively.

An iteration count estimate

A slight generalisation of Fried's result in (Fri72) states an upper bound of $\kappa(M)$ based on a constant C_{FEM} that is inherent to the finite element discretisation. It holds that

$$\kappa(M) \leq C_{\mathrm{FEM}} \cdot p_{\max} \left(\frac{h_{\max}}{h_{\min}}\right)^n,$$

where $p_{\max}$ is the maximum number of elements around any nodal point, and $h_{\max}$, $h_{\min}$ denote the largest, smallest element length in a single direction in the discretised domain with dimension $n \in \mathbb{N}$, respectively. Hence, $h_{\max}/h_{\min}$ is a measure for the non-uniformity in the mesh. Together with the estimate for $|J(\mathbf{u}_k) - J(\mathbf{u}^*)|$ being smaller than ε, an estimate k for the iteration count can be derived as

$$k \geq -\frac{1}{4}\log\left(\varepsilon / J(\mathbf{u}_0)\right) \cdot \left(p_{\max} C_{\mathrm{FEM}} \left(\frac{h_{\max}}{h_{\min}}\right)^n + 1\right) =: \hat{k}.$$

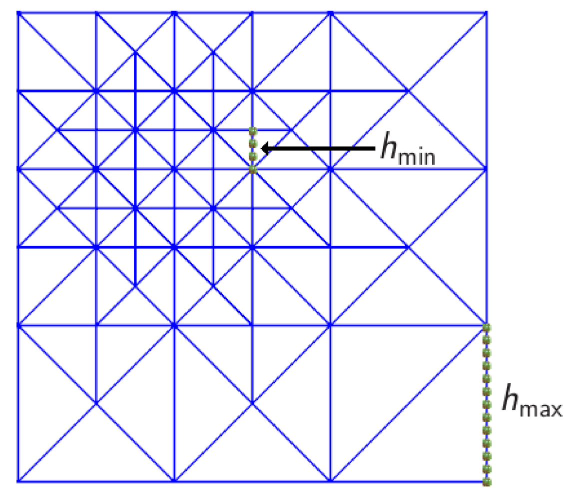
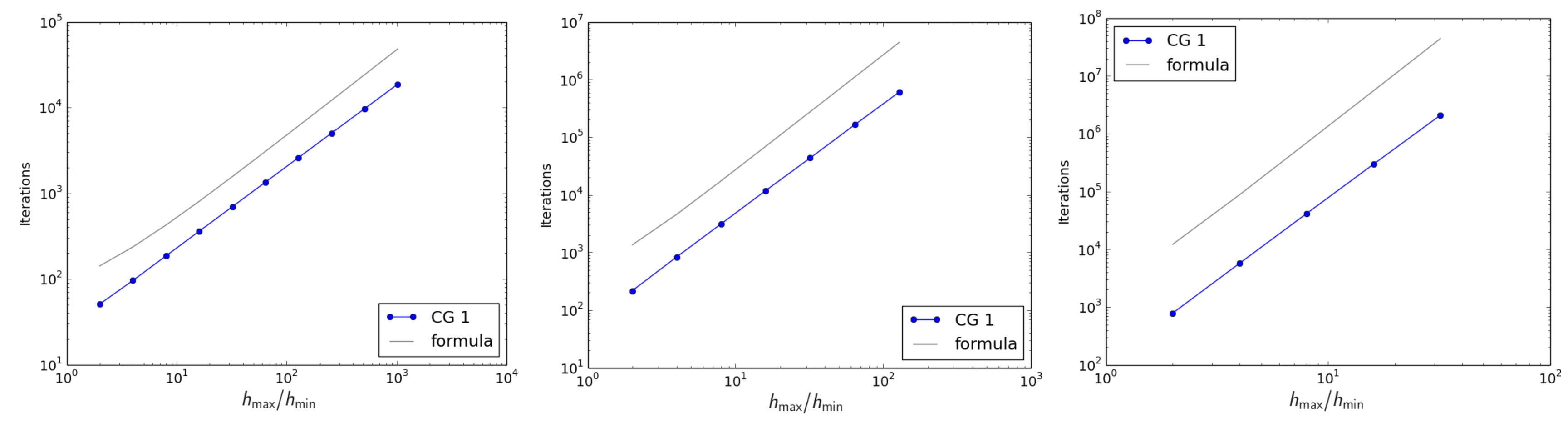


Figure: non-uniform triangular 2-D mesh with $h_{\max}$, $h_{\min}$ displayed.

The formula says that after $\hat{k}$ iterations the absolute error in $J(\mathbf{u}_{\hat{k}})$ is smaller or equal to ε.

Numerical Results

Careful interpretation of this result yields that an iteration number satisfying that the associated error is smaller or equal to ε is at most polynomial in the ratio between largest and smallest directional element size of order equal to the domain dimension n. As a matter of fact, the subsequent simulation results show that this dependency is also true numerically.



(a) 1-D (b) 2-D (c) 3-D

Figure: Iterations of the steepest descent method with exact line search and ℓ^2 represented gradients, using linear Lagrange finite elements (CG1) on non-uniform refined meshes with respect to $h_{\max}/h_{\min}$ on a log-log scale. The problem is defined by $H = L^2([0,1])$, $\alpha = 1, \beta = -1$ and $\gamma = -1/2$.

Refinement	*CG1*	*CG2*	*CG3*	*CG4*	*CG5*
1 (96)	13	20	19	37	43
2 (192)	12	20	16	36	41
3 (384)	11	19	15	35	40
4 (768)	10	19	13	35	38
5 (1536)	8	19	13	35	37

Table: Iterations of the steepest descent method with exact line search and ℓ^2 represented gradients for Lagrange finite elements of different order (CG1-5) on uniformly refined meshes in 1D depending on the refinement level (number of elements).

When a mesh is refined uniformly over the entire domain, we speak of uniform refinement. In that case, the ratio $h_{\max}/h_{\min}$ remains unaltered. Consequently, we do not expect the iteration number to increase under uniform refinement according to $\hat{k}$. Indeed, the results (left) confirm this. In fact, the iterations tend to decrease slightly for increasing refinement levels as the relative impact of the numerical error due to the boundaries gets smaller as the total number of elements increases.

Conclusion

Efficiently dealing with non-uniformity in the mesh is essential for robust optimisation. Here, this was achieved by using the inner product representation for the gradient of J that corresponds naturally to the control space. In this situation, the optimisation is mesh independent and the solution is found after only one iteration! This compares to an iteration count that is polynomially increasing for increasing non-uniformity in the mesh when using an Euclidean inner product! This makes apparent how important it is to respect inner products in optimisation.

References

[Fri72] Fried, I.: Bounds on the extremal eigenvalues of the finite element stiffness and mass matrices and their spectral condition number. In: *Journal of Sound and Vibration* 22 (1972), Nr. 4, S. 407–418

[LY08] Luenberger, D G. ; Ye, Y.: *Linear and nonlinear programming*. Bd. 116. Springer Science & Business Media, 2008

[SHF16] Schwedes, T. ; Ham, D. A. ; Funke, S. W.: An iteration count estimate for a mesh-dependent steepest descent method based on finite elements and Riesz inner product representation, preprint. (2016)

t.schwedes14@imperial.ac.uk