

Gradient Descent Methods

Enrique Zuazua

Chair of Dynamics, Control and Numerics
Department of Data Science

FAU

Erlangen, Germany

<https://caa-avh.nat.fau.eu>

Classical steepest descent

$J : H \rightarrow \mathbf{R}$. Two main assumptions:

$$\langle \nabla J(u) - \nabla J(v), u - v \rangle \geq \alpha |u - v|^2, \quad |\nabla J(u) - \nabla J(v)|^2 \leq M |u - v|^2.$$

Then, for

$$u_{k+1} = u_k - \rho \nabla J(u_k),$$

we have

$$|u_k - u^*| \leq (1 - 2\rho\alpha + \rho^2 M)^{k/2} |u_1 - u^*|.$$

Convergence is guaranteed for $0 < \rho < 1$ small enough.

Compare with the continuous marching gradient system

$$u'(\tau) = -\nabla J(u(\tau)).$$

LaSalle's invariance principle

Consider the dynamical system

$$u'(t) = -\nabla J(u(t))$$

Multiply it by $\nabla J(u(t))$ and then

$$dJ(u(t))/dt = -|\nabla J(u(t))|^2.$$

Thus, for the gradient system, $J(u(t))$ constitutes a Lyapunov function whose value diminishes along trajectories.

Assume that J is bounded below. This is typically the case when searching the minimizers of J under the standard coercivity and continuity assumptions.

Then, necessarily, $J(u(t))$ has a limit l as $t \rightarrow \infty$.

Furthermore, when J is coercive, this necessarily means that the trajectory $\{u(t)\}_{t \geq 0}$ is bounded. In the finite-dimensional context this means that the trajectory is precompact. In the infinite-dimensional case this requires further analysis of the dynamical properties of the evolution system under consideration.

Let us then define the ω -limit set. Given the initial datum u_0 of the solution of the gradient system, $\omega(u_0)$ is the set of accumulation points of the trajectory as $t \rightarrow \infty$. Obviously $J(z) = l$ for all $z \in \omega(u_0)$. On the other hand, if we denote by $z(t)$ the trajectory of the same gradient system starting at z at time $t = 0$, by the semigroup property, as J is convex, we also deduce that $J(z(t)) = l$ for all $t \geq 0$. This implies, in particular, that z is a critical point of J : $J'(z) = 0$. In case J has a unique minimizer, as it happens when J is strictly convex, then z is this minimizer. Taking into account that the accumulation point is unique, we deduce that $\omega(u_0) = \{z\}$. This implies that the whole trajectory $u(t)$ converges to z .

As we mentioned above, in the infinite-dimensional case, the boundedness of trajectories does not necessarily imply that they are relatively compact. The compactness of trajectories is normally achieved by imposing further monotonicity properties.

Indeed, when J is convex, distances diminish along trajectories. Indeed, if u and v are two trajectories of the same system then $|u(t) - v(t)|$ diminishes as time evolves.

According to this it is sufficient to prove convergence towards equilibrium for a dense set of initial data. This dense set is chosen normally to ensure compactness through the compactness of the embedding into the phase space, and the boundedness of the trajectories in that subspace.

The Direct Method of the Calculus of Variations (DMCV)

Consider a continuous, convex and coercive functional $J : H \rightarrow \mathbb{R}$ in a Hilbert space H . Then, the functional achieves its minimum in at least one point:

$$\exists h \in H : J(h) = \min_{g \in H} J(g). \quad (1)$$

This can be easily proved in a systematic manner by means of the DMCV:

Step 1. Define the infimum

$$I = \inf_{g \in H} J(g)$$

that, by the coercivity of J , necessarily satisfies $I > -\infty$.

Step 2. Consider the minimizing sequence

$$(g_n)_{n \in \mathbb{N}} \subset H : J(g_n) \searrow I. \quad (2)$$

By the coercivity of the functional J we deduce that $(g_n)_{n \in \mathbb{N}}$ is bounded in H .

Step 3. H being a Hilbert space, there exists a weakly convergent subsequence $(g_n)_{n \in \mathbb{N}}$

$$g_n \rightharpoonup g \text{ en } H. \quad (3)$$

Step 4. J being continuous in H and convex it is lower semicontinuous with respect to the weak topology. Therefore,

$$J(g) \leq \varliminf_{n \rightarrow \infty} J(g_n). \quad (4)$$

We deduce that $J(g) \leq I$ which, by the definition of infimum, implies that $J(g) = I$, which shows that the minimum is achieved.

On linear convergence of iterative methods for the variational inequality problem[★]

Paul Tseng^{*}

Department of Mathematics, University of Washington, Seattle, WA 98195, United States

Received 5 October 1993; revised 7 February 1994

Abstract

We present new linear convergence results for iterative methods for solving the variational inequality problem. The methods include the extragradient method, the proximal point method, a matrix splitting method and a certain feasible descent method. The proofs of the results are based on certain error bounds related to the algorithmic mappings. Moreover, we show that all these error bounds hold if a projection-type error bound holds.

Keywords: Variational inequality; Error bound; Linear convergence

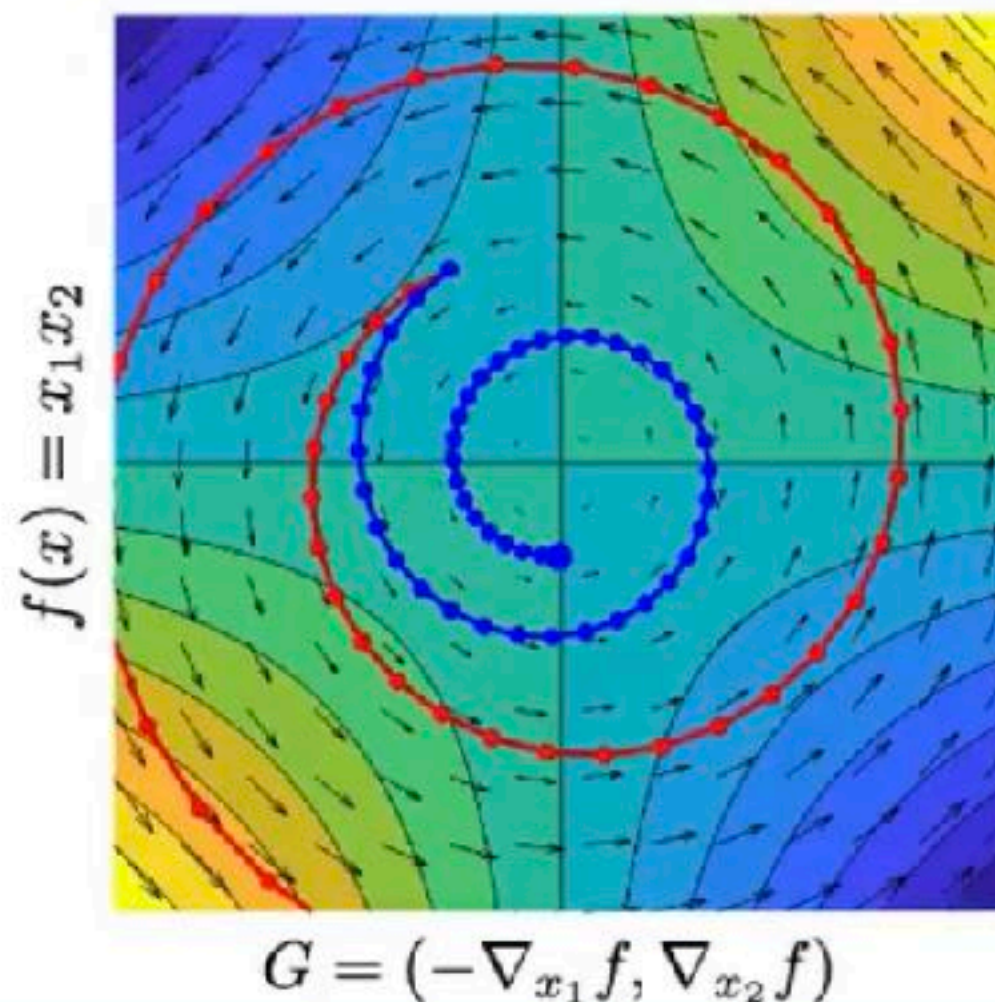


Paul Tseng

$$x_{t+1} \stackrel{\text{def.}}{=} x_t + \tau G(x_t)$$

$$x_{t+\frac{1}{2}} \stackrel{\text{def.}}{=} x_t + \tau G(x_t)$$
$$x_{t+1} \stackrel{\text{def.}}{=} x_t + \tau G(x_{t+\frac{1}{2}})$$

[Korpelevich, 1976]



**INTRODUCTION
À L'ANALYSE
NUMÉRIQUE MATRICIELLE
ET À L'OPTIMISATION**
Cours et exercices corrigés

Philippe G. Ciarlet

DUNOD