# Computational PDEs: Viscosity Solutions

(Theorie und Numerik partieller Differentialgleichungen: Viskositätslösungen)

D. Gallistl

Lecture in Winter 2021/22 at the Universität Jena
Last update: 12th April 2023

Our winter semester has 15 weeks (weeks 42–50 and 1–6). Each section in this lecture notes contains material for one week, corresponding to one lecture and (optionally) a problem session.

## Contents

**Topic 1: Finite differences for Poisson's equation** **2**
§1 Basic notions *(week 42/2021)* . . . 2
§2 Finite difference discretization of the Laplacian *(week 43)* . . . 6
§3 Basic error analysis of the finite difference method *(week 44)* . . . 10
§A Supplement: Nine-point stencils and complements on FDM ***(week 49)*** . . . 14

**Topic 2: Viscosity solutions to degenerate elliptic PDEs** **18**
§4 Jets and the definition of viscosity solutions *(week 45)* . . . 18
§5 Viscosity solutions and semicontinuous envelopes *(week 46)* . . . 23
§6 Existence and uniqueness of viscosity solutions *(week 47)* . . . 27
§7 A sufficient criterion for the comparison principle *(week 48)* . . . 31
§8 Semiconvex functions: Jensen's lemma and sup-convolutions *(week 50)* . . . 35
§9 Proof of Ishii's lemma *(week 1/2022)* . . . 39

**Topic 3: Monotone finite differences** **43**
§10 Abstract convergence; a model problem *(week 2)* . . . 43
§11 Non-negative operators *(week 3)* . . . 48
§12 Construction of monotone finite differences *(week 4)* . . . 52
§13 Discrete Alexandrov estimate *(week 5)* . . . 58
§14 Finite difference ABP estimate and stability analysis *(week 6)* . . . 62

**Literature** **64**

# Topic 1: Finite differences for Poisson's equation

## §1 Basic notions *(week 42/2021)*

In this lecture we study a class of partial differential equations (PDEs) and their numerical approximation. We confine ourselves to linear equations of second order. Let us first define what we mean by this. Throughout these notes, the space of symmetric real $n \times n$ matrices is denoted by $\mathbb{S}^{n\times n}$.

**Definition 1.1.** Let $n \in \mathbb{N}$ and $\Omega \subseteq \mathbb{R}^n$ be an open subset. Let furthermore a map

$$F : \Omega \times \mathbb{R} \times \mathbb{R}^n \times \mathbb{S}^{n\times n} \to \mathbb{R} \tag{1}$$

be given. We call the equation

$$F(x, u(x), \nabla u(x), D^2u(x)) = 0 \quad \text{for all } x \in \Omega \tag{2}$$

a *partial differential equation of 2nd order.* Any function $u : \Omega \to \mathbb{R}$ satisfying the above relation is called a *solution.*

The foregoing definition is rather abstract. At the same time, it implicitly requires further properties (differentiability) of the solution, which are not stated explicitly. We will work with this basic definition and will proceed with examples. The equation is called *partial* differential equation because it involves partial derivatives of the solution (in contrast to *ordinary differential equations (ODEs)*, which only depend on one scalar variable. The notion of *2nd order* describes that the highest involved derivative of $u$ has order 2. At this point, the function $F$ can be arbitrarily nonlinear.

**Example 1.2** (Poisson's equation)**.** Recall the Laplacian

$$\Delta u(x) = \operatorname{div} \nabla u(x) = \sum_{j=1}^{n} \partial_{jj} u(x) = \operatorname{tr} D^2u(x),$$

where $\operatorname{tr} A$ denotes the trace of a matrix $A$. For a given function $f \in C(\Omega)$ (usually referred to as *right-hand side*) and $F(x, r, p, X) = \operatorname{tr} X - f(x)$ we obtain *Poisson's equation*

$$\Delta u(x) = f(x).$$

**Example 1.3** ($\infty$-Poisson equation)**.** For a given function $f \in C(\Omega)$ and $F$ given by $F(x, r, p, X) = \operatorname{tr}(pp^\top X) - f(x)$, we obtain the equation

$$\Delta_\infty u = f(x)$$

where the $\infty$-Laplacian is defined as

$$\Delta_\infty u := \operatorname{tr}(\nabla u (\nabla u)^\top D^2u)$$

It is called $\infty$-*Poisson equation.*

**Example 1.4** (heat equation)**.** Let $\Omega = \mathbb{R}\times\mathbb{R}$. For this equation, the first variable (referring to time) is usually denoted by $t$ and the second (spatial) variable is denoted by $x$ so that we write $(t,x) \in \Omega$. For $F$ given by $F((t,x),r,p,X) = \left(\begin{smallmatrix}1\\0\end{smallmatrix}\right)\cdot p - \operatorname{tr}(\left(\begin{smallmatrix}0&0\\0&1\end{smallmatrix}\right)X)$, we obtain the so-called *heat equation*

$$\partial_t u(t,x) - \partial^2_{xx}u(t,x) = 0.$$

What are basic differences between these two examples? Poisson's equation is *linear.* This means that, given solutions $u$ to the right-hand side $f$ and $v$ to the right-hand side $g$, the equation

$$\Delta w(x) = \alpha f(x) + \beta g(x)$$

will be satisfied by the linear combination $w := \alpha u + \beta v$, $(\alpha,\beta \in \mathbb{R})$. This is easy to verify. Similar considerations show that the heat equation is linear as well. It is also elementary to verify that the $\infty$-Poisson equation does not have this property. We expect in general that

$$\Delta_\infty(u(x)+v(x)) \neq f(x)+g(x),$$

for solutions $u$ and $v$ to right-hand sides $f$ and $g$, respectively. Convince yourself of this fact by setting up suitable examples.

For $X,Y \in \mathbb{S}^{n\times n}$, the spectral theorem states that $X,Y$ are diagonalizable with real eigenvalues. We write $X \le Y$ whenever all eigenvalues of $Y - X$ are nonnegative. Another important classification is based on the following notion.

**Definition 1.5** (degenerate ellipticity)**.** The PDE (2) is *degenerate elliptic* if, given the coefficient $F$ from (1), any $(x,r,p) \in \Omega\times\mathbb{R}\times\mathbb{R}^n$, and any $X,Y{\in}\mathbb{S}^{n\times n}$ with $X \le Y$ the relation $F(x,r,p,X) \le F(x,r,p,Y)$ holds.

It follows from basic calculations that the Poisson and the $\infty$-Poisson equations are degenerate elliptic while the heat equation fails to satisfy this criterion.

In order to get started with a fairly simple setting, we will consider Poisson's equation in the first lectures.

Generally, we pose the questions of *existence* of a solution to a PDE and its *uniqueness.* Clearly, solutions to Poisson's equation are not unique without any further constraints being imposed. For instance, any solution can be shifted by an arbitrary affine function and will still remain a solution. We will thus consider the *Dirichlet problem*, which imposes a zero boundary condition on the solution. This PDE is posed on a domain $\Omega \subseteq \mathbb{R}^n$ which is open, bounded, and connected.

**Definition 1.6** (Dirichlet problem for the Laplacian)**.** Let $\Omega \subseteq \mathbb{R}^n$ be open, bounded, and connected. A function $u \in C^2(\Omega)\cap C(\bar\Omega)$ is said to be a classical solution to the Dirichlet problem (for the Laplacian) with right-hand side $f \in C(\Omega)$ and boundary values $g \in C(\partial\Omega)$ if it satisfies

$$\Delta u = f \text{ in } \Omega \quad \text{und} \quad u = g \text{ on } \partial\Omega.$$

The question under which circumstances solutions to the Dirichlet problem exist is difficult to answer in general. At this stage, we confine ourselves to study uniqueness.

**Lemma 1.7** (maximum principle). *Let $\Omega \subseteq \mathbb{R}^n$ be open, bounded, and connected and let $u \in C^2(\Omega) \cap C(\bar{\Omega})$ satisfy $\Delta u \geq 0$ in $\Omega$. Then the maximum of $u$ is attained on the boundary, i.e.,*

$$\max_{\bar{\Omega}} u = \max_{\partial\Omega} u.$$

*Proof.* We note that $\bar{\Omega}$ is compact and thus the maximum of $u$ is attained in $\Omega \cup \partial\Omega$. Let us first assume the strict inequality $\Delta u > 0$ in $\Omega$. At any point $x_0 \in \Omega$ with $u(x_0) = \max_{\bar{\Omega}} u$, the Hessian is necessarily negative-semidefinite, written $D^2 u(x_0) \leq 0$, and so has only non-positive eigenvalues. In particular its trace (the sum of all eigenvalues) is non-positive, whence $\operatorname{tr}(D^2 u(x_0)) = \Delta u(x_0) \leq 0$. In view of the assumed inequality $\Delta u > 0$, such a point $x_0 \in \Omega$ cannot exist, which implies that the maximum is attained on $\partial\Omega$. In the general case of $\Delta u \geq 0$ in $\Omega$ we let $\varepsilon > 0$ and define $u_\varepsilon(x) = u(x) + \varepsilon |x|^2$ where $|\cdot|$ denotes the Euclidean norm. We then have for any $\varepsilon > 0$ that $\Delta u_\varepsilon > 0$ in $\Omega$ and the above argument shows that

$$\max_{\bar{\Omega}} u_\varepsilon = \max_{\partial\Omega} u_\varepsilon.$$

We observe for any $x \in \bar{\Omega}$ that

$$u(x) \leq u_\varepsilon(x) \leq \max_{\bar{\Omega}} u_\varepsilon = \max_{\partial\Omega} u_\varepsilon = \max_{x\in\partial\Omega} u(x) + \varepsilon |x|^2 \leq \max_{\partial\Omega} u + \varepsilon R^2$$

for $R := \max_{x\in\bar{\Omega}} |x|^2$. The assertion then follows from letting $\varepsilon \to 0$. $\square$

**Corollary 1.8** (uniqueness). *There is at most one classical solution to the Dirichlet problem from Definition 1.6.*

*Proof.* Let $u_1$, $u_2$ be two classical solutions. Then, $w := u_1 - u_2$ satisfies $w \in C^2(\Omega) \cap C(\bar{\Omega})$ and solves $\Delta w = 0$ in $\Omega$ with $w = 0$ on $\partial\Omega$. The maximum principle implies that $w$ attains its maximum on $\partial\Omega$ and thus $w \leq 0$ in $\Omega$. On the other hand, $\Delta w = 0$ also implies $\Delta(-w) \geq 0$. The maximum principle applied to $-w$ thus proves $-w \leq 0$. In consequence $w = 0$ in $\Omega$ and thus $u_1 = u_2$. $\square$

**Corollary 1.9** (comparison principle). *Let $u, v \in C^2(\Omega) \cap C(\bar{\Omega})$ be such that $u \leq v$ on $\partial\Omega$ and $\Delta u \geq \Delta v$ in $\Omega$. Then $u \leq v$ in $\Omega$.*

*Proof.* The difference $w := u - v$ satisfies $\Delta w \geq 0$ and by the maximum principle $w$ attains its maximum on $\partial\Omega$. But there we have $w \leq 0$. Therefore $w \leq 0$ in $\Omega$ or equivalently $u \leq v$ in $\Omega$. $\square$

## Synopsis of §1.

We have formulated an abstract PDE and classified some examples with respect to (non) linearity and degenerate ellipticity. We have formulated the Dirichlet problem of the Laplacian and proved the maximum principle. Its consequences are uniqueness of solutions and a comparison principle.

# Problems to §1

**Problem 1.** Let the following function be given

$$\Phi(x) = \begin{cases} -\frac{1}{2\pi}\log|x| & \text{if } n = 2 \\ \frac{1}{n(n-2)\alpha(n)}\frac{1}{|x|^{n-2}} & \text{if } n \geq 2. \end{cases}$$

Here, $\alpha(n) \neq 0$ is some real number. Show that $\Delta\Phi(x) = 0$ holds for all $x \in \mathbb{R}^n \setminus \{0\}$.

**Problem 2.** Prove that the Laplacian is represented in polar coordinates $(r, \varphi)$ as follows

$$\Delta = \frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \varphi^2}.$$

**Problem 3.** Let a linear 2nd order PDE be given by

$$\sum_{|\alpha|\leq 2} a_\alpha(x)\partial^\alpha u(x) = f(x).$$

Here, $a_\alpha$ and $f$ are given functions over $\Omega$. The above sum runs over all multi-indices $\alpha$ of length $\leq 2$, and $\partial^\alpha$ is the partial derivative with respect to $\alpha$. Show that the linear PDE is degenerate ellitpic if and only if the matrix $(a_\alpha)_{|\alpha|=2}$ of the indices belonging to multiindices of length 2 is positive semi-definite.

**Problem 4.** Write the following PDEs in the format (2) and decide which of them are linear or degenerate elliptic.

- Poisson's equation
- $\infty$-Poisson equation
- heat equation
- $\partial^2 u_{tt} - \partial^2 u_{xx} = f(t,x)$ for $(t,x) \in \mathbb{R} \times \mathbb{R}$
- $\partial_{x_1} u(x) - u(x)\,\mathrm{tr}(\left(\begin{smallmatrix} 2 & 1 \\ 0 & -10 \end{smallmatrix}\right) D^2 u(x)) = f(x)$ for $x = (x_1, x_2) \in \mathbb{R}^2$
- $\Delta u(x) - |u(x)|^3 = 0$ for $x \in \mathbb{R}^3$

**Problem 5.** Install a suitable Python environment on your computer. Use the `NumPy` library to perform the elementary matrix-vector multiplications

$$\begin{bmatrix} 2 & 4 \\ -3 & 1 \end{bmatrix}\begin{bmatrix} 2 \\ -3 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 1 \\ 3 & 4 & 1 & 2 \\ 4 & 1 & 2 & 3 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

## §2 Finite difference discretization of the Laplacian *(week 43)*

We want to design a numerical method to approximately solve the Dirichlet problem. For the sake of a clear presentation, we confine ourselves to the case of $\Omega$ being the two-dimensional square domain $\Omega = (0,1)^2$ and to homogeneous boundary conditions, i.e., $g = 0$ in Definition 1.6. Generalizations will be discussed later (problem sessions).

The idea of the so-called Finite Difference Method (FDM) is to replace partial derivatives by difference quotients.

**Definition 1.10** (first-order difference quotients)**.** Given a step size $h > 0$ and a sequence $(U_j)_{j=0,\dots,J}$ of elements of some vector space, we define

$$\partial^+ U_j := \frac{U_{j+1} - U_j}{h}, \quad (j = 0, \dots, J-1) \quad \textit{(forward difference quotient)}$$

and

$$\partial^- U_j := \frac{U_j - U_{j-1}}{h}, \quad (j = 1, \dots, J) \quad \textit{(backward difference quotient).}$$

**Definition 1.11** (second-order central difference quotient)**.** Given a step size $h > 0$ and a sequence $(U_j)_{j=0,\dots,J}$ of elements of some vector space, the quantity

$$\partial^+ \partial^- U_j = \frac{U_{j+1} - 2U_j + U_{j-1}}{h^2},$$

is called the *second-order central difference quotient.*

For a function $u$ over $[0,1]$ we let

$$\partial^+ u(x) = \frac{u(x+h) - u(x)}{h}$$

with analogous notation for $\partial^-$. The following approximation properties can be proven via Taylor expansion.

**Lemma 1.12.** *Given $u \in C^2([0,1])$, we have for $\partial_x^+$ and $\partial_x^-$ that*

$$|\partial_x^+ u(x) - \partial_x u(x)| \leq \frac{h}{2} \|\partial_{xx}^2 u\|_{C([0,1])} \quad \textit{for all } x \in [0, 1-h]$$
$$|\partial_x^- u(x) - \partial_x u(x)| \leq \frac{h}{2} \|\partial_{xx}^2 u\|_{C([0,1])} \quad \textit{for all } x \in [h, 1].$$

*Given $u \in C^4([0,1])$, we have for $\partial_x^+$ and $\partial_x^-$ that*

$$|\partial_x^+ \partial_x^- u(x) - \partial_x^2 u(x)| \leq \frac{h^2}{12} \|\partial_{xxxx}^4 u\|_{C([0,1])} \quad \textit{for all } x \in [h, 1-h].$$

*Proof.* Problem 6. □

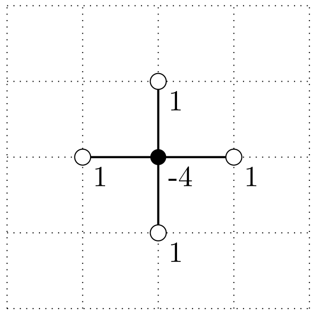


Figure 1: Scematic diagram of the 5-point stencil with weights.

Let $J \geq 0$ and $h = 1/J$. We set up a grid with $J+1$ points in every coordinate direction by letting

$$x_{j,k} = (jh, kh) \quad j, k = 0, \ldots, J.$$

We wish to approximate the solution $u$ by a grid function $U$ whose value at $x_{j,k}$ we denote by $U_{j,k}$. For interior points we define a discrete version of the Laplacian $\Delta = \partial^2_{x_1 x_1} + \partial^2_{x_2 x_2}$ through central differences

$$\Delta_h U_{j,k} = \partial^+_{x_1} \partial^-_{x_1} U_{j,k} + \partial^+_{x_2} \partial^-_{x_2} U_{j,k} \quad \text{for } j, k = 1, \ldots, J-1.$$

It is straightforward to compute the representation

$$\Delta_h U_{j,k} = \frac{1}{h^2}(U_{j+1,k} + U_{j,k+1} - 4U_{j,k} + U_{j-1,k} + U_{j,k-1}). \tag{3}$$

We see that the value $\Delta_h U_{j,k}$ depends on the point $x_{j,k}$ and its four neighbours in the grid. The stencil is called *five-point stencil*, see Figure 1.

**Definition 1.13.** Let $\Omega = (0,1)^2$ and $f \in C(\Omega)$. The discretized Poisson problem (with zero boundary conditions) seeks $(U_{j,k} : j, k = 0, \ldots, J)$ such that

$$\begin{cases} \Delta_h U_{j,k} = f(x_{j,k}) & \text{for } j, k = 1, \ldots, J-1 \\ U_{0,k} = U_{J,k} = U_{j,0} = U_{j,J} = 0 & \text{for } j, k = 0, \ldots, J. \end{cases}$$

We briefly comment on the implementation. In order to represent $U$ as a vector, we choose the *lexicographic enumeration* and identify $\{0, \ldots, J\}^2$ with $\{1, \ldots, L\}$ (where $L = (J+1)^2$) through the map

$$(j, k) \mapsto j + k(J+1) + 1 =: \ell.$$

Loosely speaking, we enumerate the grid by taking rows from left to right starting on the left bottom. We see from (3) that the discrete Laplacian takes the form

$$\Delta_h U_\ell = \frac{1}{h^2}(U_{\ell+1} + U_{\ell+(J+1)} - 4U_\ell + U_{\ell-1} + U_{\ell-(J+1)})$$

for any interior point $x_\ell$. We see that $U_{j,k}$ for $j$ or $k$ in $\{0, J\}$ are no unknowns because they are known through the boundary condition. We are therefore merely interested in computing $U_{j,k}$ for $j, k \in \{1, \ldots, J-1\}$. We consider the sub-list $(\mathring{U}_1, \ldots \mathring{U}_N)$ corresponding to the interior points and define the matrix

$$X := \begin{bmatrix} -4 & 1 & & \\ 1 & \ddots & \ddots & \\ & \ddots & \ddots & 1 \\ & & 1 & -4 \end{bmatrix}.$$

This results in the system

$$\begin{bmatrix} X & I & & \\ I & \ddots & \ddots & \\ & \ddots & \ddots & I \\ & & I & X \end{bmatrix} \begin{bmatrix} \mathring{U}_1 \\ \vdots \\ \vdots \\ \mathring{U}_N \end{bmatrix} = h^2 \begin{bmatrix} f_1 \\ \vdots \\ \vdots \\ f_N \end{bmatrix}.$$

Here $f_\ell = f(x_\ell)$ for every interior node. We note that this is a system of the type $Ax = b$ for a sparse matrix $A$. In an implementation, a sparse matrix format should be used.

### Synopsis of §2.

We have defined various difference quotients and studied their approximation properties.

We have formulated the Dirichlet problem of the Laplacian and proved the maximum principle. Its consequences are uniqueness of solutions and a comparison principle. We defined a discrete version of the Laplacian by using central differences in $x$ and $y$ (5-point stencil). Finally, we have discussed how to represent the discrete system of equations as a (sparse) matrix-vector problem.

# Problems to §2

**Problem 6.** Prove Lemma 1.12.

**Problem 7.** Show that the discrete problem from Definition 1.13 and the stated matrix-vector system are equivalent.

**Problem 8.** Have a look at the `scipy.sparse` library, in particular `dia_matrix` and `linalg`. Use these tools to set up the (sparse) system matrix of the finite difference method.

**Problem 9.** Implement the finite difference method for the Poisson problem on the square domain for zero boundary conditions and the right-hand side $f(x) = e^{x_1} x_1(x_1(x_2^2 - x_2 + 2) + 3x_2^2 - 3x_2 - 2)$. You can use the command `spsolve` for a direct solver for sparse matrices. Use different mesh sizes $h = 2^{-2}, 2^{-3}, 2^{-4}, 2^{-5}$. Compare the computed solution with the exact solution (given by $u(x) = e^{x_1}(x_1 - x_1^2)(x_2 - x_2^2)$) at the grid-points by considering the error in the maximum-norm. Visualize the computed solutions using surface plot tools from Python (see also Problem 10).

**Problem 10.** Inform yourself about the possibilities of creating surface plots in Python and visualize the finite difference solution from the previous exercise.

*Hint:* A basic example taken from https://www.geeksforgeeks.org/3d-surface-plotting-in-python-using-matplotlib/

```
# Import libraries
from mpl_toolkits import mplot3d
import numpy as np
import matplotlib.pyplot as plt
# Creating dataset
x = np.outer(np.linspace(-3, 3, 32), np.ones(32))
y = x.copy().T # transpose
z = (np.sin(x **2) + np.cos(y **2) )
# Creating figure
fig = plt.figure(figsize =(14, 9))
ax = plt.axes(projection ='3d')
# Creating plot
ax.plot_surface(x, y, z)
# show plot
plt.show()
```

## §3 Basic error analysis of the finite difference method *(week 44)*

We want to quantify the error $u - U$ between the true solution $u$ to the Dirichlet problem and its finite difference approximation $U$. The fundamental tool is a discrete version of the maximum principle for $\Delta_h$.

**Lemma 1.14** (discrete maximum principle)**.** *Let $\Omega$ be the unit square. Let the mesh function $U$ satisfy $\Delta_h U_{j,k} \geq 0$ for all $j,k \in \{1,\dots,J-1\}$. Then, $U$ attains its maximum at a boundary point (i.e., at some $x_{j,k}$ with $j \in \{0,J\}$ or $k \in \{0,J\}$.*

*Proof.* Let $x_{j,k}$ with $j,k \in \{1,\dots,J-1\}$ be an interior point. From the definition of $\Delta_h$ we obtain

$$U_{j,k} = \frac{1}{4}(U_{j-1,k} + U_{j+1,k} + U_{j,k+1} + U_{j,k-1}) - \frac{h^2}{4}\Delta_h U_j.$$

From $\Delta_h U_{j,k} \geq 0$ we thus infer

$$U_{j,k} \leq \frac{1}{4}(U_{j-1,k} + U_{j+1,k} + U_{j,k+1} + U_{j,k-1}).$$

Assume $U_{j,k}$ is the maximum of $U$. Then it is not smaller than any of the four neighbouring values. Hence, equality holds in the foregoing estimate. In particular

$$U_{j,k} = U_{j-1,k} = U_{j+1,k} = U_{j,k+1} = U_{j,k-1}.$$

Iterating this argument up to the boundary shows that $U$ is constant and therefore the maximum is attained at the boundary. □

The foregoing lemma was formulated for the unit square. It is clear how to generalize it to other geometries.

We denote the set of boundary points of the grid by $\Gamma$. For mesh functions $V$ we use the following notation on maximum norms

$$|V|_{\infty,\bar\Omega} := \max_{\substack{j,k=0,\dots,J\\ \text{s.t. } x_{j,k}\in\Omega\cup\Gamma}} |V_{j,k}|$$

$$|V|_{\infty,\Omega} := \max_{\substack{j,k=0,\dots,J\\ \text{s.t. } x_{j,k}\in\Omega}} |V_{j,k}|$$

$$|V|_{\infty,\Gamma} := \max_{\substack{j,k=0,\dots,J\\ \text{s.t. } x_{j,k}\in\Gamma}} |V_{j,k}|$$

For the unit square we have $\Gamma \subseteq \partial\Omega$. Note, however, that for more complicated geometries the 'boundary points' of the grid need not lie on $\partial\Omega$.

The discrete maximum principle implies the following stability estimate.

**Lemma 1.15** (stability)**.** *Let $\Omega$ be the unit square. There exists a constant $C > 0$ with the following property. Given a mesh over $\Omega$ and a mesh function $U$, we have*

$$|U|_{\infty,\bar\Omega} \leq |U|_{\infty,\Gamma} + C|\Delta_h U|_{\infty,\Omega}.$$

*Proof.* We define the mesh function

$$W_{j,k} = \frac{1}{4}|x_{j,k}|^2 \quad \text{(squared Euclidean norm)}.$$

Then $\Delta_h W_{j,k} = 1$ for any pair $(j,k)$. Let $r := |\Delta_h U|_{\infty,\Omega}$ and define the mesh functions $V^{\pm} := \pm U + rW$. Then

$$\Delta_h V^{\pm} = \pm \Delta_h U + r \geq 0.$$

By the discrete maximum principle, $V^{\pm}$ attains its maximum on the boundary. This means

$$\pm U + rW \leq |\pm U + rW|_{\infty,\Gamma} \quad \text{over } \bar{\Omega}.$$

The triangle inequality on the right-hand side and $W \geq 0$ on the left hand side thus prove

$$|U|_{\infty,\bar{\Omega}} \leq |U|_{\infty,\Gamma} + r|W|_{\infty,\Gamma}.$$

This proves the assertion with $C = |W|_{\infty,\Gamma}$. □

**Remark 1.16.** The generalization of the stability estimate to domains different from the square is immediate.

**Corollary 1.17.** *The finite difference method has a unique solution $U$.*

*Proof.* We have already seen that the finite difference system is a quadratic finite-dimensional system of linear equations. Thus, uniqueness implies existence. Suppose there exist two solutions $U$, $V$ satisfying $\Delta_h U = F = \Delta_h V$ (where $F$ is the mesh function interpolating $f$ at the grid points) and $U|_\Gamma = 0 = V_\Gamma$. Then $\Delta_h(U - V) = 0$, and the stability estimate implies $|U - V|_{\infty,\bar{\Omega}} = 0$. Thus $U = V$. □

**Remark 1.18.** For a mesh function $F$ we denote by $\Delta_h^{-1} F$ the solution to the finite difference system with zero boundary conditions. The stability estimate can then be written as follows

$$|\Delta_h^{-1} F|_{\infty,\bar{\Omega}} \leq C|F|_{\infty,\Omega}.$$

We thus see that $\Delta_h^{-1}$ has a uniformly bounded continuity constant ($C$ is independent of the grid size $h$).

When operating on grids we identify $u$ with the mesh function having values $u(x_{j,k})$.

**Lemma 1.19** (consistency). *Assume $u \in C^4(\bar{\Omega})$. Then*

$$|\Delta_h u - \Delta u|_{\infty,\Omega} \leq \frac{1}{2} h^2 \sum_{j=1,2} \|\partial_{x_j}^4 u\|_{C(\bar{\Omega})}.$$

*Proof.* This is an immediate consequence of Lemma 1.12. □

**Theorem 1.20** (FDM convergence)**.** *Assume the solution $u$ to the Poisson problem $\Delta u = f$ over the unit square $\Omega$ with homogeneous boundary conditions satisfies $u \in C^4(\bar\Omega)$. Then the finite difference error satisfies*

$$|u - U|_{\infty,\bar\Omega} \leq Ch^2 \sum_{j=1,2} \|\partial_{x_j}^4 u\|_{C(\bar\Omega)}$$

*with a constant $C$ independent of the mesh size and $f$.*

*Proof.* Stability implies

$$|u - U|_{\infty,\bar\Omega} \leq C|\Delta_h(u - U)|_{\infty,\Omega} = C|\Delta_h u - \Delta u|_{\infty,\Omega}$$

because $\Delta_h U = F = \Delta u$ at the grid points. The right-hand side is then estimated with the consistency estimate, which concludes the proof. $\square$

**Remark 1.21.** The simple proof of convergence shows the general principle of convergence proofs for finite difference methods:

$$\text{stability } + \text{consistency} \implies \text{convergence.}$$

This can be formalized in a general framework (Lax–Richtmyer theorem), but we confine ourselves to formulating this rule of thumb. The above convergence proof contains the whole essence of the reasoning behind.

## Synopsis of §3.

We have formulated a discrete version of the maximum principle. The main consequence was a stability estimate. Together with consistency (which is a consequence of Taylor expansions), this led to the convergence proof.

# Problems to §3

**Problem 11** (convergence rates in Hölder norms)**.** Let $k \in \mathbb{N}_0$ and $0 < \alpha \leq 1$ and define the following norm

$$\|v\|_{C^{k,\alpha}(\bar{\Omega})} = \|v\|_{C^k(\bar{\Omega})} + \max_{|\beta|=k} \sup_{\substack{x,y\in\Omega \\ x\neq y}} \frac{|\partial^\beta v(x) - \partial^\beta v(y)|}{|x-y|^\alpha}.$$

A continuous function $v$ with finite norm $\|v\|_{C^{k,\alpha}(\bar{\Omega})}$ is said to be uniformly Hölder continuous of class $C^{k,\alpha}$. Prove that the finite difference method satisfies the following convergence estimate

$$|u - U|_{\infty,\bar{\Omega}} \leq Ch^\alpha \max_{j=1,2} \|\partial^2_{x_j} u\|_{C^{0,\alpha}(\bar{\Omega})}$$

provided $\|u\|_{C^{2,\alpha}(\bar{\Omega})} < \infty$.

*Hint:* Use first-order Taylor expansion with Lagrange form of the remainder.

**Problem 12.** Given an inhomogeneous Dirichlet boundary condition $u = g$ on $\partial\Omega$, we can extend the interpolated boundary condition to the interior by zero to a grid function $U_g$. We then solve the auxiliary FDM problem

$$\Delta_h U_0 = F - \Delta_h g$$

and see that

$$U := U_0 + U_g$$

solves $\Delta_h U = F$ and satisfies the boundary condition at the boundary grid points. Implement the FDM for the problem

$$\Delta u = 0 \text{ in } \Omega \quad \text{ and } u|_{\partial\Omega} = g$$

with $g(x,y) = x^3 - 3xy^2$. Plot the computed solution and perform an experimental convergence study (the exact solution is given by $u(x,y) = x^3 - 3xy^2$).

**Problem 13.** Let $\Omega = (-1,1)^2 \setminus ([0,1] \times [-1,0])$ be the Γ-shaped (or L-shaped) domain. Let $u$ be given by

$$u(x,y) = (1-x^2)(1-y^2)r^{2/3} \sin\left(\frac{2\varphi}{3}\right).$$

Here, we use polar coordinates $0 < r < 1$ and $0 < \varphi < 3\pi/2$; note that $x = r\cos\varphi$ and $y = r\sin\varphi$.

(a) Prove that $u$ satisfies $\Delta u = f$ for some $f \in C^0(\bar{\Omega})$ and $u|_{\partial\Omega} = 0$. Compute $f$.

(b) Prove that $u$ does not possess bounded derivatives and, thus, does not belong to $C^1(\bar{\Omega})$.

**Problem 14.** Find a way to extend the FDM to the L-shaped domain by eliminating points outside $\bar{\Omega}$ from the resulting system. Test the method for the setting from Problem 13 where the boundary condition is given by the (known) exact solution. Which convergence properties do you observe?

## §A Supplement: Nine-point stencils and complements on FDM *(week 49)*

The five-point stencil studied so far is somehow a minimal choice. One can think of improving accuracy by increasing the dependence on neighbouring grid points. In two dimensions, nine-point stencil take into account the diagonal neighbours as well. We note that the distance of a point $x_{j,k}$ to its diagonal neighbour is $\sqrt{2}h$. We then have the central differences

$$\frac{U_{j+1,k} - 2U_{j,k} + U_{j-1,k}}{2h}$$
$$\frac{U_{j,k+1} - 2U_{j,k} + U_{j,k-1}}{2h}$$
$$\frac{U_{j+1,k-1} - 2U_{j,k} + U_{j-1,k+1}}{2\sqrt{2}h}$$
$$\frac{U_{j+1,k+1} - 2U_{j,k} + U_{j-1,k-1}}{2\sqrt{2}h},$$

see also Figure 2.

We next discuss how to design a linear combination that consistently discretizes the Laplacian and has higher-order convergence properties.

Consider the function $u(x_{j,k} + te_m)$ where $m \in \{1, 2\}$ is the $m$th cartesian unit vector. Taylor expansion of fourth order results in

$$\begin{aligned} u(x_{j,k} + te_m) = u(x_{j,k}) + \partial_m u(x_{j,k})t + \frac{1}{2}\partial_m^{(2)} u(x_{j,k})t^2 + \frac{1}{6}\partial_m^{(3)} u(x_{j,k})t^3 \\ + \frac{1}{24}\partial_m^{(4)} u(x_{j,k})t^4 + \frac{1}{120}\partial_m^{(5)} u(x_{j,k})t^6 + O(t^5). \end{aligned}$$

If we evaluate this expression for $t = \pm h$ and add the results, the odd-order terms cancel and we obtain

$$u(x_{j,k} + he_m) + u(x_{j,k} - he_m) = 2u(x_{j,k}) + \partial_m^{(2)} u(x_{j,k})h^2 + \frac{1}{12}\partial_m^{(4)} u(x_{j,k})h^4 + O(h^6).$$

Adding this identity for $m = 1, 2$ results in the well known relation of the 5-point stencil

$$\begin{aligned} &u(x_{j+1,k}) + u(x_{j-1,k}) + u(x_{j,k+1}) + u(x_{j,k-1}) \\ &= 4u(x_{j,k}) + \Delta u(x_{j,k})h^2 + \frac{1}{12}(\partial_{xxxx} + \partial_{yyyy})u(x_{j,k})h^4 + O(h^6). \end{aligned} \tag{4}$$

We can apply similar arguments to the diagonal directions

$$d_1 = 2^{-1/2}(1, -1) \quad \text{and} \quad d_1 = 2^{-1/2}(1, 1)$$

and obtain with $t = \pm\sqrt{2}h$ and analogous computations

$$\begin{aligned} &u(x_{j+1,k-1}) + u(x_{j-1,k+1}) + u(x_{j+1,k+1}) + u(x_{j-1,k-1}) \\ &= 4u(x_{j,k}) + 2\Delta u(x_{j,k})h^2 + \frac{1}{6}(\partial_{xxxx} + \partial_{yyyy} + 6\partial_{xxyy})u(x_{j,k})h^4 + O(h^6). \end{aligned} \tag{5}$$

Here, various sums of mixed derivatives have cancelled out. We now add 4 times (4) to (5) and obtain

$$\begin{aligned}
4u(x_{j+1,k}) + 4u(x_{j-1,k}) + 4u(x_{j,k+1}) + 4u(x_{j,k-1})\\
+u(x_{j+1,k-1}) + u(x_{j-1,k+1}) + u(x_{j+1,k+1}) + u(x_{j-1,k-1})\\
= 20u(x_{j,k}) + 6\Delta u(x_{j,k})h^2 + \frac{1}{2}(\partial_{xxxx} + \partial_{yyyy} + 2\partial_{xxyy})u(x_{j,k})h^4 + O(h^6)
\end{aligned}$$

We use that

$$\Delta u(x_{j,k}) = f(x_{j,k})$$

and

$$(\partial_{xxxx} + \partial_{yyyy} + 2\partial_{xxyy})u(x_{j,k}) = (\partial_{xx} + \partial_{yy})\Delta u(x_{j,k}) = (\partial_{xx} + \partial_{yy})f(x_{j,k})$$

and derive the relations

$$S^{9p}_{(j,k)}u = 6h^2 f(x_{j,k}) + \frac{1}{2}h^4 \Delta f(x_{j,k}) + O(h^6).$$

for the 9-point stencil $S^{9p}_{(j,k)}$ symbolized as follows

$$\begin{bmatrix} 1 & 4 & 1\\ 4 & -20 & 4\\ 1 & 4 & 1 \end{bmatrix}.$$

The corresponding finite difference equations are then

$$\begin{aligned}
-20U_{j,k} + 4U_{j+1,k} + 4U_{j-1,k} + 4U_{j,k+1} + 4U_{j,k-1} + U_{j+1,k-1} + U_{j-1,k+1} + U_{j+1,k+1} + U_{j-1,k-1}\\
= 6h^2 f(x_{j,k}) + \frac{1}{2}h^4 \Delta f(x_{j,k}).
\end{aligned}$$

**Remark 1.22.** We expect $U$ to converge at a better order than the ordinary 5-point stencil provided the exact solution is sufficiently regular. We will not provide a detailed proof in this lecture but remark that it can in principle be worked out with the basic tools from the previous section.

**Remark 1.23.** The 9-point stencil can be viewed as a weighted average of two (rotated) 5-point stencils. From the above derivation it is clear that any convex combination of the stencils yields a first-order scheme. The special choice 4 : 1 and a modification of the right-hand side, however, result in an even higher-order scheme.

We know that convergence of any finite difference scheme follows from stability and consistency. We do not work out an error analysis of the nine-point stencil here; it will be part of the problem sessions.

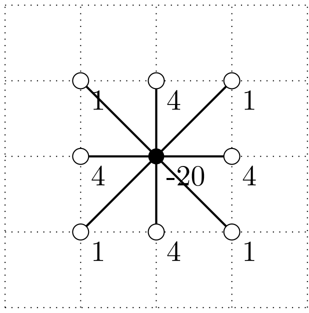


Figure 2: Scematic diagram of the 9-point stencil with weights.

**Curved geometries.** We end this section by commenting on practical aspects of the FDM (be it the 5- or 9-point stencil). For convenience, we formulated many results for the square where the domain could be exactly covered by a cartesian mesh. For more complicated situations with possibly curved geometries this is no longer possible. Assume for example that domain $\Omega \subseteq [0,1]^2$ can be embedded in the unit square (or any other box after appropriate scaling). Generally we cannot expect that the boundary $\partial\Omega$ has a meaningful intersection with the gridpoints. Instead, we define

$$\Omega_h := \{x_{j,k} : x_{j,k} \in \Omega \text{ and all neighbours belong to } \bar{\Omega}\}$$

and

$$\Gamma_h := \{x_{j,k} : x_{j,k} \in \Omega \text{ and a neighbour does not belong to } \bar{\Omega}\}.$$

By neighbour we mean a gridpoint belonging to the stencil at $x_{j,k}$. The FDM equations then read $\Delta_h U_{j,k} = F_{j,k}$ for all $x_{j,k} \in \Omega_h$. The results proven in the foregoing sections transfer to this situation.

**More general elliptic operators.** We can reduce a PDE of the form

$$\operatorname{tr}(AD^2u) = f$$

with a (constant) positive definite and symmetric matrix $A$ to an equation involving only the diagonal entries of $D^2$ by diagonalizing $A = R\Lambda R^T$ with an orthogonal matrix $R$ and a diagonal matrix $\Lambda$. Since the trace is independent of the chosen coordinate system we see that the above PDE is equivalent to

$$\operatorname{tr}(\Lambda R^T D^2 u R) = f.$$

It is easy to check that this PDE only depends on $\partial_{r_1,r_1}$ and $\partial_{r_2,r_2}$ where $r_1, r_2$ are the chosen eigenvectors of $A$. Thus, after rotating the coordinate system, a (weighted) 5-point stencil can be used.

When lower-order terms are present, for instance as

$$\operatorname{tr}(AD^2u) + b \cdot \nabla u + cu = f$$

for a vector $b$ and a constant $c$, these can be included as well. The zero-order term is simply discretized by $cU$. The term involving the gradient can be discretized through first-order difference quotients.

### Synopsis.

We have derived the nine-point stencil. We have chosen the coefficients in such a way that appropriate terms in the Taylor expansion cancel, leading to higher asymptotic accuracy. Derivatives of $f$ enter the right-hand side. We concluded with some comments on curved domains and more complicated PDE operators.

## Problems.

**Problem 15.** Work out the details in the Taylor expansions for the derivation of the 9-point stencil.

**Problem 16.** Prove that the 9-point stencil satisfies a discrete maximum principle and work out an error estimate for the finite difference error $|u - U|_{\infty,\Omega}$ for the Laplacian on the unit square with homogeneous Dirichlet boundary conditions.

**Problem 17.** Implement the 9-point stencil finite difference method for the Poisson problem on the square domain for zero boundary conditions and the right-hand side from Problem 9 Use different mesh sizes and compare the computed solution with the exact solution at the grid-points by considering the error in the maximum-norm. Compare the convergence speed with that of the 5-point stencil.

**Problem 18.** Compare (experimentally) the performance (in terms of convergence rates) of the 5-point and the 9-point stencil for the example on the L-shaped domain (see Problems 13 and 14). Give a theoretical explanation of what you observe.

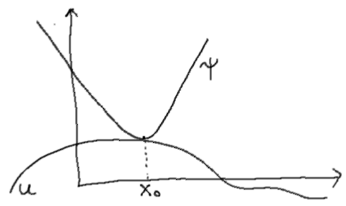


Figure 3: The graph of $u$ is touched from above by $\psi$ at $x_0$.

# Topic 2: Viscosity solutions to degenerate elliptic PDEs

## §4 Jets and the definition of viscosity solutions *(week 45)*

The basic concept of 'weak solution' we will work with in this lecture are *viscosity solutions.* Before we state precise definitions, we shall explain the underlying idea. Let us start with Poisson's equation $F(\cdot, D^2u) = 0$ for $F(\cdot, D^2u) = \Delta u - f$.

A function $u \in C^2(\Omega) \cap C(\bar{\Omega})$ is a classical solution to $F(\cdot, D^2u) = 0$ if and only if it is a subsolution, that is $F(\cdot, D^2u) = \Delta u - f \geq 0$, and a supersolution, that is $F(\cdot, D^2u) = \Delta u - f \leq 0$. We will now weaken the latter properties and thereby generalize them to merely continuous functions.

Suppose $u$ is a subsolution and suppose further that there is a function $\psi \in C^2(\mathbb{R}^n)$ such that $u$ and $\psi$ coincide in $x_0$,

$$u(x_0) - \psi(x_0) = 0,$$

and $u - \psi$ has a local maximum at $x_0$. The latter means that there is some $r > 0$ such that

$$u - \psi \leq 0 \quad \text{in } B_r(x_0) \subseteq \Omega.$$

In a visual imagination one can think of the graph of $u$ being touched from above by the graph of $\psi$ in the point $x_0$, see Figure 3.

Since $u - \psi$ has a maximum at $x_0$, the Hessian is negative semidefinite in $x_0$ and thus $D^2u(x_0) \leq D^2\psi(x_0)$. From degenerate ellipticity of $F$ we conclude

$$0 \leq F(x_0, D^2u(x_0)) \leq F(x_0, D^2\psi(x_0)).$$

We shall say that $u$ is a *viscosity subsolution*, if for every $x_0 \in \Omega$ and every $\psi \in C^2(\mathbb{R}^n)$ such that $u - \psi$ has a local maximum at $x_0$, the relation $F(x_0, D^2\psi(x_0)) \geq 0$ is valid. We thus generalize the notion of subsolution by replacing the differential inequality $\Delta u - f \geq 0$ at $x_0$ by the inequality $\Delta\psi - f \geq 0$ for every suitable test function $\psi$ touching $u$ from above in $x_0$. The principal achievement of this definition is that we do not require any differentiability of $u$ for being a viscosity subsolution. But it can be easily checked (and is outlined above) that every supersolution $u \in C^2(\Omega)$ is automatically also a viscosity supersolution. We say that $u$ is a *viscosity supersolution*, if for every $x_0 \in \Omega$ and every $\psi \in C^2(\mathbb{R}^n)$ such that $u - \psi$ has a local minimum at $x_0$, we have $F(x_0, D^2\psi(x_0)) \leq 0$. A continuous function is then called a *viscosity solution*, if it is simultaneously a sub- and supersolution. We stress the fact that we do not require more differentiability than mere continuity of $u$. For more general functions $F$, possibly depending on $u(x)$ and $\nabla u(x)$, the same reasoning applies because if $\psi$ touches $u$ at $x_0$ we have $\psi(x_0) = u(x_0)$ and, from the extremal property, $\nabla\psi(x_0) = \nabla u(x_0)$. We thus can compare

$$F(x_0, u(x_0), \nabla u(x_0), D^2u(x_0)) \geq \text{(resp. } \leq) F(x_0, \psi(x_0), \nabla\psi(x_0), D^2\psi(x_0))$$

in the above arguments.

We will now formalize the above ideas using slightly more general notions, which will turn out to be of help in many arguments.

We start by defining generalizing pointwise derivatives for semicontinuous functions.

**Definition 2.24** (semicontinuity)**.** Let $\Omega \subseteq \mathbb{R}^n$. A function $u : \Omega \to \mathbb{R}$ is *upper semicontinuous* on $\Omega$ if

$$\limsup_{y\to x} u(y) \leq u(x) \quad \text{for all } x \in \Omega.$$

We then write $u \in USC(\Omega)$. A function $u : \Omega \to \mathbb{R}$ is *lower semicontinuous* on $\Omega$ if

$$\liminf_{y\to x} u(y) \geq u(x) \quad \text{for all } x \in \Omega.$$

We then write $u \in LSC(\Omega)$.

**Definition 2.25** (second-order jets)**.** Let $u \in USC(\Omega)$ and $x \in \Omega$. The set

$$\mathcal{J}^{2,+}u(x) := \left\{ (p, X) \in \mathbb{R}^n \times \mathbb{S}^{n\times n} \,\middle|\, u(x+z) \leq u(x) + p\cdot z + \frac{1}{2}z^\top Xz + o(|z|^2) \text{ as } z \to 0 \right\}$$

is called the *second-order super-jet* of $u$ at $x$.

For $u \in LSC(\Omega)$ and $x \in \Omega$, the set

$$\mathcal{J}^{2,-}u(x) := \left\{ (p, X) \in \mathbb{R}^n \times \mathbb{S}^{n\times n} \,\middle|\, u(x+z) \geq u(x) + p\cdot z + \frac{1}{2}z^\top Xz + o(|z|^2) \text{ as } z \to 0 \right\}$$

is called the *second-order sub-jet* of $u$ at $x$. □

**Remark 2.26.** The second-order super-jet of $u$ at $x$ describes all paraboloids that can touch $u$ from above at $x$. The symmetric statement holds for the second-order sub-jet. The jets generalize (one-sided) derivatives. In the problems below we shall see that there are continuous functions not differentiable in a point $x_0$ but having nonempty intersection $\mathcal{J}^{2,+}u(x_0) \cap \mathcal{J}^{2,-}u(x_0)$. The latter is then interpreted as the generalized second-order derivative in $x_0$.

**Definition 2.27** (viscosity solutions)**.** Consider the degenerate elliptic PDE from Definition 1.1 with continuous coefficients $F \in C(\Omega \times \mathbb{R} \times \mathbb{R}^n \times \mathbb{S}^{n\times n})$. A function $u \in USC(\Omega)$ is called a *viscosity subsolution* if $(p, X) \in \mathcal{J}^{2,+}u(x)$ implies $F(x, u(x), p, X) \geq 0$ for every $x \in \Omega$. A function $u \in LSC(\Omega)$ is called a *viscosity supersolution* if $(p, X) \in \mathcal{J}^{2,-}u(x)$ implies $F(x, u(x), p, X) \leq 0$ for every $x \in \Omega$. A continuous function $u \in C(\Omega)$ is called *viscosity solution* if it is both a viscosity sub- and supersolution.

**Remark 2.28.** This definition in terms of jets matches the intuition from the heuristic derivation above. Also, the property of being a subsolution $F(x, u(x), \nabla u(x), D^2u(x)) \geq 0$ is weakened and only required for the one-sided generalized derivative, namely

$$\inf_{(p,X)\in \mathcal{J}^{2,+}u(x)} F(x, u(x), p, X) \geq 0.$$

Analogous statements apply to supersolutions. Note that the jets may well be empty, in which case there is nothing to be checked.

We need the following technical lemma.

**Lemma 2.29.** *Let $\sigma \in [0, \infty) \to \mathbb{R}$ be a nondecreasing function. Then there exists $\tau \in C^2(0, \infty)$ with*

$$\sigma(t) \leq \tau(t) \leq 8\sigma(4t) \quad \text{for all } t > 0.$$

*Proof.* It is well known that monotone functions are measurable. Let

$$\tau(t) := \frac{1}{2t^2} \int_0^{4t} \int_0^r \sigma(s)\, ds\, dr.$$

Since $\sigma$ is increasing, we increase the inner integral from the definition of $\tau$ if we replace $r$ by $4t$ and then estimate $\sigma(s)$ from above by $\sigma(4t)$. This proves the stated upper bound for $\tau(t)$. We furthermore note from the definition of $\tau$ and the monotonicity of $\sigma$ that

$$\tau(t) \geq \frac{1}{2t^2} \int_{2t}^{4t} \int_0^r \sigma(s)\, ds\, dr.$$

In order to estimate this from below, we shrink the integration range for the $s$-integral to $[t, 2t]$ and estimate $\sigma(s)$ from below by $\sigma(t)$. This shows $\tau(t) \geq \sigma(t)$. ☐

**Theorem 2.30** (touching by $C^2$ functions)**.** *Let $u \in USC(\Omega)$. For every $x \in \Omega$ we have that*

$$\mathcal{J}^{2,+}u(x) = \left\{ (\nabla\psi(x), D^2\psi(x)) \,\middle|\, \begin{array}{l} \psi \in C^2(\mathbb{R}^n) \text{ with } u(x) = \psi(x) \text{ and} \\ u - \psi \leq 0 \text{ in an open subset containing } x \end{array} \right\}.$$

*Proof of Theorem 2.30.* The inclusion $\supseteq$ follows from Problem 19. We now show the inclusion $\subseteq$. Let $(p, X) \in \mathcal{J}^{2,+}u(x)$. By definition of the jet we have

$$u(x + z) \leq u(x) + p \cdot z + \frac{1}{2} z^\top X z + \sigma(|z|)|z|^2 \quad \text{for } |z| \text{ small enough}$$

for some nondecreasing function $\sigma$ with $\sigma(0) = 0$. We use $\tau$ from Lemma 2.29 and deduce

$$u(x + z) \leq u(x) + p \cdot z + \frac{1}{2} z^\top X z + \tau(|z|)|z|^2 \quad \text{for } |z| \text{ small enough.}$$

We now take the right-hand side of this estimate, make the change of variables $z = y - x$, and define

$$\psi(y) := u(x) + p \cdot (y - x) + \frac{1}{2}(y - x)^\top X(y - x) + \tau(|y - x|)|y - x|^2.$$

By construction we have $\psi \in C^2(\mathbb{R}^n)$, $u(x) = \psi(x)$ and $u \leq \psi$ near $x$. From Taylor expansion of $\psi$ we see that necessarily $p = \nabla u$ and $X = D^2\psi$. $\square$

**Remark 2.31.** (a) Theorem 2.30 states that the second-order upper semi-jet consists of all pairs of gradient and Hessian in $x$ of a $C^2$ function $\psi$ touching the graph of $u$ from above at $x$. By setting $\bar{\psi}(z) := \psi(z) + |z - x|^4$ we even see the stronger statement that the touching function can be assumed to satisfy

$$u - \bar{\psi} < 0 \in B_r(x) \setminus \{x\}$$

for some $r > 0$, i.e., the function $\bar{\psi}$ touches $u$ only in $x$.

(b) We see that Definition 2.27 is equivalent to what was described at the beginning of this section.

## Synopsis of §4.

In a heuristic derivation we have illustrated the idea of viscosity solutions, namely, shifting derivatives to smooth test functions touching the graph from above/below. We have formalized the concept using semicontinuous functions and second-order jets.

# Problems to §4

**Problem 19.** Let $u \in C^2(\mathbb{R}^n)$. Prove that

$$\mathcal{J}^{2,+}u(x) \cap \mathcal{J}^{2,-}u(x) = \{(\nabla u(x), D^2u(x))\}$$

for every $x \in \mathbb{R}^n$.

**Problem 20.** Let $W : \mathbb{R} \to [0,1]$ be a bounded function that is nowhere differentiable and define the function $w \in C(\mathbb{R})$ by $w(x) = W(x)|x|^3$.

(a) Prove that $w$ is differentiable in $x = 0$.

(b) Prove that $w$ is nowhere differentiable in $\mathbb{R} \setminus \{0\}$ and that $w$ is not twice differentiable in $x = 0$.

(c) Prove that $w(z) = o(|z|^2)$ as $z \to 0$.

(d) Prove that $(0_{\mathbb{R}^n}, 0_{\mathbb{S}^{n\times n}}) = \mathcal{J}^{2,+}w(0) \cap \mathcal{J}^{2,-}w(0)$.

**Problem 21.** Let $u(x) = -|x|$. Compute $\mathcal{J}^{2,\pm}u(x)$ for every $x \in \mathbb{R}$. Draw a picture of representative elements of the upper and lower semijets of in the point $x = 0$.

**Problem 22.** Let $\Omega = \{x \in \mathbb{R}^2 : |x| < R\}$ for $R = 1/2$. Define the functions

$$f(x) = \begin{cases} 0 & \text{for } x = 0 \\ \frac{x_2^2 - x_1^2}{2|x|^2}\left(\frac{4}{-\log|x|} + \frac{1}{2(-\log|x|)^{3/2}}\right) & \text{for } x \neq 0 \end{cases}$$

and

$$u(x) = (x_1^2 - x_2^2)\sqrt{-\log|x|}.$$

(a) Prove that $f \in C(\bar{\Omega})$ and $u \in C(\bar{\Omega}) \cap C^2(\Omega \setminus \{0\})$.

(b) Show $\Delta u = f$ in $\Omega$.

(c) Show that $u$ is not a classical solution (consider second-order partial derivatives near 0).

(d) Show that $u$ is a viscosity solution to $\Delta u = f$.

**Problem 23.** Check the following real functions for upper/lower semicontinuity in $x = 0$.

$$u(x) = \begin{cases} x/|x| & x \neq 0 \\ 1 & \text{else} \end{cases}, \quad v(x) = \begin{cases} x^{-1} & x > 0 \\ 0 & \text{else} \end{cases}, \quad w(x) = \begin{cases} \cos(x^{-1}) & x \neq 0 \\ 0 & x = 0. \end{cases}$$

## §5 Viscosity solutions and semicontinuous envelopes *(week 46)*

We start with a definition and a remark related to the previous section. For future considerations, the concept of jet closures will turn out useful. We introduce it now to be able to discuss some properties in the exercises.

**Definition 2.32** (jet closures)**.** Let $u \in USC(\Omega)$ and $x \in \Omega$. The jet closure $\bar{\mathcal{J}}^{2,+}u(x)$ is defined as

$$\bar{\mathcal{J}}^{2,+}u(x) := \left\{ (p, X) \in \mathbb{R}^n \times \mathbb{S}^{n\times n} \;\middle|\; \begin{array}{l} (x,p,X) = \lim_{m\to\infty}(x_m, p_m, X_m) \\ \text{for some sequence } (p_m, X_m) \in \mathcal{J}^{2,+}(x_m) \end{array} \right\}.$$

The jet closure $\bar{\mathcal{J}}^{2,-}u(x)$ for $u \in LSC(\Omega)$ is defined in an analogous fashion.

**Remark 2.33.** Definition 2.27 remains the same if we replace the jets $\mathcal{J}^{2,+}u(x)$, $\mathcal{J}^{2,-}u(x)$ by their closures $\bar{\mathcal{J}}^{2,+}u(x)$, $\bar{\mathcal{J}}^{2,-}u(x)$, see Problem 28.

**Definition 2.34** (Semicontinuous envelopes)**.** Let $u : \Omega \to \mathbb{R}$ be a function. The function $u^* : \Omega \to \mathbb{R} \cup \{+\infty\}$ given by

$$u^*(x) := \lim_{r\to 0} \sup_{B_r(x)} u$$

is called the *upper semicontinuous envelope* of $u$. The function $u_* : \Omega \to \mathbb{R} \cup \{-\infty\}$ given by

$$u_*(x) := \lim_{r\to 0} \inf_{B_r(x)} u$$

is called the *lower semicontinuous envelope* of $u$.

**Remark 2.35.** It is easy to verify that indeed $u^* \in USC(\Omega)$ and $u_* \in LSC(\Omega)$, see Problem 25. The envelope $u^*$ [$u_*$] is the smallest [largest] function in $USC(\Omega)$ [$LSC(\Omega)$] that is larger [smaller] than $u$; see Problem 26.

**Lemma 2.36** (envelopes of suprema and jets)**.** *Let $\Omega \subseteq \mathbb{R}^n$ be open. Given a subset $\mathcal{U} \subseteq USC(\Omega)$, define its pointwise supremum $U : \Omega \to \mathbb{R}$ by*

$$U(x) := \sup_{u\in\mathcal{U}} u(x)$$

*and denote its upper semicontinuous envelope by $U^*$. For every $x \in \Omega$ and every $(p, X) \in \mathcal{J}^{2,+}U^*(x)$ there exist sequences $(x_m)_m \in \Omega^{\mathbb{N}}$ and $(u_m)_m \in \mathcal{U}^{\mathbb{N}}$ and*

$$(p_m, X_m) \in \mathcal{J}^{2,+}u_m(x_m)$$

*such that*

$$(x_m, u(x_m), p_m, X_m) \to (x, U^*(x), p, X) \quad \text{as } m \to \infty.$$

*Proof.* From the definition of the pointwise supremum and its upper semicontinuous envelope we infer that there exist sequences $\hat{x}_m \to x$ and $u_m$ with $u_m(\hat{x}_m) \to U^*(x)$.

By Theorem 2.30 and Remark 2.31 we see that $(p, X) = (\nabla\psi(x), D^2\psi(x))$ for some $\psi \in C^2(\mathbb{R}^n)$ with $U^*(x) = \psi(x)$ and $u - \psi < 0$ in $B_R(x) \setminus \{x\}$ for some sufficiently small $R$. We may assume $\overline{B_R(x)} \subseteq \Omega$.

Let, for every $m$, $x_m \in \overline{B_R(x)}$ denote a point where $u_m - \psi$ attains its maximum in $\overline{B_R(x)}$. The resulting bounded sequence $(x_m)_m$ then has a subsequence converging to some $x^* \in \overline{B_R(x)}$. Without loss of generality we may assume that $(x_m)$ is identical to this subsequence (because $u_m(\hat{x}_m) \to U^*(x)$ remains valid for subsequences). We have

$$(U^* - \psi)(x) = \limsup_{m\to\infty}(u_m - \psi)(\hat{x}_m) \leq \limsup_{m\to\infty}(u_m - \psi)(x_m) \leq (U^* - \psi)(x^*).$$

Since $x$ is the unique maximizer of $(U^* - \psi)$, we thus infer $x = x^*$. In particular, for $m$ sufficiently large, we have $x_m \in B_R(x)$ and $u_m - \psi$ has a maximum at $x_m$ at this interior point. Thus,

$$(p_m, X_m) := (\nabla\psi(x_m), D^2\psi(x_m)) \in \mathcal{J}^{2,+}u_m(x_m).$$

as can be seen from Theorem 2.30. Since the first and second-order derivatives of $\psi$ are continuous, we have $(p_m, X_m) \to (p, X)$ as $m \to \infty$. Using continuity of $\psi$ and the maximality of $(u_m - \psi)(x_m)$ we finally obtain

$$\begin{aligned} U^*(x) &= \limsup_{m\to\infty} u_m(\hat{x}_m) \\ &= \limsup_{m\to\infty}(u_m(\hat{x}_m) - \psi(\hat{x}_m) + \psi(x_m)) \leq \limsup_{m\to\infty} u_m(x_m) \leq U^*(x). \end{aligned}$$

□

**Theorem 2.37** (supremum of subsolutions)**.** *Suppose $\mathcal{U} \subseteq USC(\Omega)$ is a family of viscosity subsolutions to the PDE $F(\cdot, u, \nabla u, D^2u) = 0$ with continuous $F$ and define*

$$U(x) := \sup_{u\in\mathcal{U}} u(x)$$

*and denote its upper semicontinuous envelope by $U^*$. If $U^*$ is finite over $\Omega$, then $U^*$ is a viscosity subsolution.*

*Proof.* Let $x \in \Omega$ be arbitrary and $(p, X) \in \mathcal{J}^{2,+}U^*(x)$. By Lemma 2.36, there exist sequences $(x_m)_m \in \Omega^{\mathbb{N}}$ and $(u_m)_m \in \mathcal{U}^{\mathbb{N}}$ and

$$(p_m, X_m) \in \mathcal{J}^{2,+}u_m(x_m)$$

such that

$$(x_m, u(x_m), p_m, X_m) \to (x, U^*(x), p, X) \quad \text{as } m \to \infty.$$

Each $u_m$ is a viscosity subsolution, whence

$$F(x_m, u_m(x_m), p_m, X_m) \geq 0.$$

By continuity of $F$ we conclude for $m \to \infty$ that

$$F(x, U^*(x), p, X) \geq 0.$$

Thus, $U^*$ satisfies the criterion for viscosity subsolutions. □

**Synopsis of §5.**

We have defined semicontinuous envelopes. We have seen that jets of the USC envelope of the pointwise supremum over some family of USC functions can be approximated by jets from that family. The USC envelope over a family of subsolutions is again a subsolution (if finite).

## Problems to §5

**Problem 24.** Let $K \subseteq \mathbb{R}^n$ be compact. Prove that any $v \in USC(K)$ attains its maximum and any $w \in LSC(K)$ attains its minimum.

**Problem 25.** Given a function $u : \Omega \to \mathbb{R}$, prove that $u^* \in USC(\Omega)$.

**Problem 26.** Let $u : \Omega \to \mathbb{R}$. Prove that $u \leq u^*$. Prove further that any $v \in USC(\Omega)$ with $v \geq u$ satisfies $v \geq u^*$. *(The comparison of functions is meant in the pointwise sense.)*

**Problem 27.** Compute the envelopes $w^*$ and $w_*$ of

$$w(x) = \begin{cases} \cos(x^{-1}) & x \neq 0 \\ 0 & x = 0. \end{cases}$$

**Problem 28.** Prove that the notion of viscosity solution remains the same if we replace the jets in Definition 2.27 by their jet closures.

## §6 Existence and uniqueness of viscosity solutions *(week 47)*

Let $F$ continuous an degenerate elliptic. We study the boundary-value problem (the so-called *Dirichlet problem*) for the PDE from (1.1) with continuous $F$: Seek $u \in C(\bar{\Omega})$ such that

$$F(x, u(x), \nabla u(x), D^2 u(x)) = 0 \quad \text{for all } x \in \Omega \tag{6a}$$

$$u(x) = g(x) \text{ for all } x \in \partial\Omega. \tag{6b}$$

Here, $g \in C(\partial\Omega)$ prescribes the boundary values. A simple instance of this Dirichlet problem is the Dirichlet Laplacian from Definition 1.6.

**Definition 2.38.** A function $\underline{u} \in USC(\bar{\Omega})$ (resp. $\bar{u} \in LSC(\bar{\Omega})$) is a viscosity subsolution (resp. supersolution) to the Dirichlet problem (6), if it is a viscosity subsolution (resp. supersolution) to the PDE (6a) and satisfies $u \leq g$ (resp. $u \geq g$) on $\partial\Omega$. A continuous function is called viscosity solution to the Dirichlet problem, it it is simultaneously sub- and supersolution.

The fundamental concept for our theory of unique solvability in the viscosity sense is the following.

**Definition 2.39** (comparison principle)**.** We say that the Dirichlet problem (6) satisfies the *comparison principle*, if and subsolution $\underline{u} \in USC(\bar{\Omega})$ and any supersolution $\bar{u} \in LSC(\bar{\Omega})$ satisfy

$$\underline{u} \leq \bar{u} \quad \text{in } \Omega.$$

We will learn sufficient criteria for the comparison principle to hold in the subsequent lectures. An immediate implication is the uniqueness of viscosity solutions.

**Proposition 2.40** (uniqueness)**.** *Assume the Dirichlet problem* (6) *satisfies the comparison principle. Then there is at most one viscosity solution to* (6)*.*

*Proof.* Assume $u_1$, $u_2$ are viscosity solutions to the Dirichlet problem. Since $u_1$ is in particular a subsolution with $u_1 \leq g$ and $u_2$ a supersolution with $u_2 \geq g$, we obtain from the comparison principle

$$u_1 \leq u_2 \text{ in } \Omega \quad \text{and} \quad u_1 \leq g \leq u_2 \text{ on } \partial\Omega.$$

From interchanging the roles of $u_1$, $u_2$ we obtain $u_1 = u_2$ in $\bar{\Omega}$. $\square$

The existence of solutions relies on an explicit construction referred to as *Perron's method.*

**Theorem 2.41** (existence)**.** *Assume the Dirichlet problem* (6) *satisfies the comparison principle. Assume furthermore that there exist a subsolution $\underline{u}$ with $\underline{u}_* = g$ on $\partial\Omega$ and $\underline{u}_* \geq -\infty$ and a supersolution $\bar{u}$ with $\bar{u}^* < \infty$ in $\Omega$ and $\bar{u}^* = g$ on $\partial\Omega$. Then*

$$V(x) := \sup\{v(x) : v \textit{ is subsolution with } \underline{u} \leq v \leq \bar{u}\}$$

*is a viscosity solution to the Dirichlet problem.*

*Proof.* Obviously $\underline{u} \leq V \leq \bar{u}$. We consider upper semicontinuous envelopes and see that

$$\underline{u}_* \leq V_* \leq V \leq V^* \leq \bar{u}^* \quad \text{in } \Omega.$$

From our assumptions on the boundary values of the envelopes we deduce that

$$V_* = V = V^* = g \quad \text{on } \partial\Omega.$$

From the assumption $\bar{u}^* < \infty$ we deduce that $V^*$ is finite with $V^* \leq g$ on the boundary $\partial\Omega$. Thus, Theorem 2.37 shows that $V^*$ is a viscosity subsolution to the Dirichlet problem. From elementary considerations and the assumed comparison principle we further deduce

$$\underline{u} = \underline{u}^* \leq V^* \leq \bar{u}.$$

Since $V$, in its definition, is the pointwise supremum and $V^*$ is a subsolution, we see $V \geq V^*$. We thus deduce $V = V^* \in USC(\Omega)$ is a subsolution. Let us show that $V_*$ is a supersolution as well. Assume (for contradiction) that there is $\hat{x} \in \Omega$ where the supersolution property fails to hold. Then, by the *bump construction* in Lemma 2.42 below, we can locally modify $V$ on some $B_r(\hat{x})$ such that the modified function $U$ is still a viscosity subsolution, satisfies $V \leq U$, and $V < U$ in a nonempty subset of $B_r(\hat{x})$. $\hat{x}$. The comparison principle then implies $U \leq \bar{u}$. This contradicts the maximality of $V$. Thus, such a point $\hat{x}$ cannot exist, whence $V_*$ is a viscosity supersolution. From $V_* \leq V^*$, the boundary conditions, and the comparison principle, we deduce $V_* = V = V^*$. Thus, $V$ is a viscosity solution to the Dirichlet problem. □

**Lemma 2.42** (the bump construction). *Let $u \in USC(\Omega)$ be a viscosity subsolution to the PDE* (2) *with continuous $F$ and assume $u_* > -\infty$. Let $\hat{x} \in \Omega$ be a point such that there exists $(p, X) \in \mathcal{J}^{2,-}u_*(\hat{x})$ such that $F(\hat{x}, u_*(\hat{x}), p, X) > 0$ (which means that $u_*$ fails to be a viscosity supersolution at $\hat{x}$). Then, for some sufficiently small $r > 0$, there exists a viscosity subsolution $U \in USC(\Omega)$ with $U \geq u$ in $\bar{\Omega}$ and $U = u$ on $\bar{\Omega} \setminus B_r(\hat{x})$ such that*

$$U > u \text{ on a nonempty subset of } B_r(\hat{x}).$$

*Proof.* Without loss of generality we assume for convenience that $\hat{x} = 0$. By assumption there exists $(p, X) \in \mathcal{J}^{2,-}(0)$ with $F(0, u_*(0), p, X) > 0$. In particular, we have

$$u(z) \geq u_*(z) \geq u_*(0) + p \cdot z + \frac{1}{2} z^\top X z + o(|z|^2) \quad \text{for small } |z|. \tag{7}$$

We introduce parameters $\gamma, \delta, r > 0$ and let

$$w_{\gamma,\delta}(z) := \delta + u_*(0) + p \cdot z + \frac{1}{2} z^\top X z - \frac{\gamma}{2}|z|^2.$$

For small $r$ we evaluate near $0 < |z| \leq r$ and obtain

$$F(z, w_{\gamma,\delta}(z), \nabla w_{\gamma,\delta}, D^2 w_{\gamma,\delta}) = F(z, \delta + u_*(0) + O(r), p + O(r), X - \gamma I_{n\times n}).$$

We obtain from continuity of $F$ that

$$F(z, w_{\gamma,\delta}(z), \nabla w_{\gamma,\delta}, D^2 w_{\gamma,\delta}) \geq F(0, u_*(0), p, X) + o(1) \quad \text{as } (\gamma, \delta, r) \to 0.$$

Therefore, by the above assumption $F(0, u_*(0), p, X) > 0$,

$$F(z, w_{\gamma,\delta}(z), \nabla w_{\gamma,\delta}, D^2 w_{\gamma,\delta}) \geq 0 \quad \text{for } (\gamma, \delta, r) \text{ sufficiently small.}$$

Let such a sufficiently small $(\gamma, \delta, r)$ be given. Note that we have just computed that the continuous function $w_{\gamma,\delta}$ is a subsolution on the ball $B_r(0)$. Comparing the definition of $w_{\gamma,\delta}$ with (7), we see that for fixed $\gamma$ we can choose $\delta$ and $r$ so small that $u > w_{\gamma,\delta}$ in $B_r(0) \setminus B_{r/2}(0)$. We then define

$$U(x) := \begin{cases} \max\{u(x), w_{\gamma,\delta}(x)\} & \text{for } x \in B_r(0), \\ u(x) & \text{for } x \in \Omega \setminus B_r(0). \end{cases}$$

Theorem 2.37 implies that $U$ is a viscosity subsolution (because the USC functions $u$ and $w_{\gamma,\delta}$ are). This function satisfies

$$\limsup_{z \to 0} (U(z) - u(z)) = U(0) - u_*(0) \geq w_{\delta,\gamma}(0) - u_*(0) = \delta > 0.$$

Thus, the function has the claimed properties. □

### Synopsis of §6.

We have formulated the Dirichlet problem. The comparison principle is a criterion directly implying uniqueness. It also implies existence via Perron's method.

## Problems to §6

**Problem 29** (lack of comparison)**.** Consider the one-dimensional elliptic Dirichlet problem

$$u''(x) + 18x(u'(x))^4 = 0 \quad \text{in } (-1,1), \quad u(-1) = b, \quad u(1) = -b$$

for some $b > 1$. Let furthermore the functions

$$\underline{u}(x) = \begin{cases} x^{1/3} - 1 + b, & x \in [0,1], \\ x^{1/3} + 1 - b, & x \in [-1,0), \end{cases} \qquad \bar{u}(x) = \begin{cases} x^{1/3} - 1 + b, & x \in (0,1], \\ x^{1/3} + 1 - b, & x \in [-1,0], \end{cases}$$

be given. Prove that $\underline{u}$ is subsolution and $\bar{u}$ is supersolution to the Dirichlet problem but $\underline{u}(z) > \bar{u}(z)$ for some point $z$.

**Problem 30.** Let $\Omega \subseteq \mathbb{R}^n$ be open, $x \in \bar{\Omega}$, $u \in USC(\bar{\Omega})$, and $\psi \in C^2(\bar{\Omega})$. Prove that

$$\mathcal{J}^{2,+}(u+\psi)(x) = \left\{ (\nabla\psi(x), D^2\psi(x)) + (p,X) : (p,X) \in \mathcal{J}^{2,+}u(x) \right\}$$

and

$$\bar{\mathcal{J}}^{2,+}(u+\psi)(x) = \left\{ (\nabla\psi(x), D^2\psi(x)) + (p,X) : (p,X) \in \bar{\mathcal{J}}^{2,+}u(x) \right\}.$$

**Problem 31.** Let $u \in USC(\Omega)$ have a local maximum at $x \in \Omega$. Prove that any symmetric positive semidefinite matrix $0 \leq A \in \mathbb{S}^{n\times n}$ satisfies

$$(0, A) \in \mathcal{J}^{2,+}u(x).$$

**Problem 32.** Let $u, v \in USC(\Omega)$ and $\psi \in C^2(\Omega \times \Omega)$. Assume the function $u(x) + v(x) - \psi(x,y)$ has a maximum at $(x,y) \in \Omega \times \Omega$. Prove that there exist $p_1, p_2 \in \mathbb{R}^n$ and $B \in \mathbb{S}^{2n\times 2n}$ such that

$$((p_1,p_2), B) \in \mathcal{J}^{2,+}(u(x) + v(y)) \quad \text{and} \quad B \leq D^2\psi(x,y).$$

## §7 A sufficient criterion for the comparison principle *(week 48)*

Precise characterizations of the comparison principle are not known. We give a sufficient criterion under which the comparison principle holds. We consider the simplified setting where the PDE has no dependence of $u(x)$ and $\nabla u(x)$, i.e.,

$$F(x, r, p, X) = F(x, X)$$

and we assume that $F : \mathbb{R} \times \mathbb{S}^{n\times n} \to \mathbb{R}$ is continuous.

Throughout the next lectures, we assume that we are given fixed constants $0 < \lambda \leq \Lambda < \infty$. They quantify what we call *uniform ellipticity*. By $I$ we denote the $n \times n$ unit matrix.

**Definition 2.43** (Pucci's operators)**.** Given $X \in \mathbb{S}^{n\times n}$ we define

$$\mathcal{P}^+(X, \lambda, \Lambda) = \mathcal{P}^+(X) := \max\{\operatorname{tr}(AX) : A \in \mathbb{S}^{n\times n} \text{ with } \lambda I \leq A \leq \Lambda I\}$$
$$\mathcal{P}^-(X, \lambda, \Lambda) = \mathcal{P}^-(X) := \min\{\operatorname{tr}(AX) : A \in \mathbb{S}^{n\times n} \text{ with } \lambda I \leq A \leq \Lambda I\}$$

**Warning 2.44.** The choice of signs in the definition of the Pucci operators and the notion of (uniform/degenerate) ellipticity is not uniform in the literature. Some authors use reversed signs.

**Definition 2.45** (uniform ellipticity)**.** The map $F : \Omega \times \mathbb{S}^{n\times n} \to \mathbb{R}$ is *uniformly elliptic* if for all $x \in \Omega$ and all $X, Y \in \mathbb{S}^{n\times n}$ we have

$$\mathcal{P}^-(Y - X) \leq F(x, Y) - F(x, X) \leq \mathcal{P}^+(Y - X).$$

Note that the definition depends on $\lambda$ and $\Lambda$, which are referred to as the uniform ellipticity constants.

**Definition 2.46** (structure condition)**.** Let $F$ be given. We assume that there exists a continuous nonnegative function $\omega_F$ with $\omega_F(0) = 0$ such that the following is satisfied. If $X, Y \in \mathbb{S}^{n\times n}$ and $\mu > 1$ satisfy

$$-3\mu \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \leq \begin{bmatrix} X & 0 \\ 0 & -Y \end{bmatrix} \leq 3\mu \begin{bmatrix} I & -I \\ -I & I \end{bmatrix}$$

then

$$F(x, X) - F(y, Y) \leq \omega_F\Big(|x - y|(1 + \mu|x - y|)\Big) \quad \text{for all } x, y \in \Omega.$$

Recall the jet closures from Definition 2.32.

**Theorem 2.47** (Ishii's lemma)**.** *Let $w_1, w_2 \in USC(\bar{\Omega})$ and consider the sum $w(x, y) = w_1(x) + w_2(y)$. Let $(x, y) \in \Omega^2$ with an element*

$$((p_1, p_2), A) \in \mathcal{J}^{2,+}w(x, y).$$

*Then, for each $\varepsilon > 0$ there exist matrices $X, Y \in \mathbb{S}^{n\times n}$ such that*

$$(p_1, X) \in \bar{\mathcal{J}}^{2,+}w_1(x) \quad \textit{and} \quad (p_2, Y) \in \bar{\mathcal{J}}^{2,+}w_2(y)$$

*and*

$$-\left(\frac{1}{\varepsilon} + \|A\|\right) I \leq \begin{bmatrix} X & 0 \\ 0 & Y \end{bmatrix} \leq A + \varepsilon A^2.$$

*Proof.* The proof is postponed to later sections. □

**Theorem 2.48** (comparison principle)**.** *Consider the PDE*

$$F(\cdot, D^2u) = 0$$

*for a continuous and uniformly elliptic $F$ and let the structure assumption be satisfied. Let $u \in USC(\bar{\Omega})$ be a viscosity subsolution and $v \in LSC(\bar{\Omega})$ a viscosity supersolution. Then, $u \leq v$ on $\partial\Omega$ implies $u \leq v \in \bar{\Omega}$.*

*Proof.* Assume (for contradiction) that $\theta := \max_{\bar{\Omega}} u - v > 0$. We choose $\delta > 0$ such that

$$\delta \max_{x \in \bar{\Omega}} e^{x_1/\lambda} \leq \theta/2$$

and set $\tau := \max_{x \in \bar{\Omega}}(u(x) - v(x) + \delta e^{x_1/\lambda}) \geq \theta > 0$. For $\alpha > 0$ we define the map $\Phi_\alpha : \bar{\Omega} \times \bar{\Omega} \to \mathbb{R}$ as

$$\Phi_\alpha(x, y) = u(x) - v(y) - \frac{\alpha}{2}|x - y|^2 + \delta e^{x_1/\lambda}.$$

and let $(x_\alpha, y_\alpha)$ be a point where $\Phi_\alpha$ attains its maximum. Note that this maximum satisfies $\max \Phi_\alpha \geq \tau$ (choose $x = y$ and maximize). In particular we have

$$|x_\alpha - y_\alpha|^2 \leq \frac{2}{\alpha}(u(x_\alpha) - v(y_\alpha) + \delta e^{x_{\alpha,1}/\lambda} - \tau) \leq \frac{2}{\alpha}(\max u - \min v + \theta/2 - \tau). \tag{8}$$

Since $\bar{\Omega}$ is compact, there is a point $(\hat{x}, \hat{y}) \in \bar{\Omega}^2$ and a subsequence (not relabelled) such that $(x_\alpha, y_\alpha) \to (\hat{x}, \hat{y})$ as $\alpha \to \infty$. From (8) we thus see that $\hat{x} = \hat{y}$. There we have

$$u(\hat{x}) - v(\hat{x}) + \delta e^{\hat{x}_1/\lambda} = \tau.$$

This implies $\hat{x} \in \Omega$, because of $u \leq v$ on $\partial\Omega$ and the choice of $\delta$. We further note from the first inequality in (8) that

$$\lim_{\alpha \to \infty} \alpha |x_\alpha - y_\alpha|^2 = 0. \tag{9}$$

We have that eventually $\Phi_\alpha$ attains its maximum at interior points $(x_\alpha, y_\alpha)$. We thus have that for some ($\alpha$-dependent) $p_1, p_2$, $B$, and $A = D^2(\frac{\alpha}{2}|x_\alpha - y_\alpha|^2)$

$$((p_1, p_2), B) \in \mathcal{J}^{2,+}(w_1(x_\alpha) + w_2(y_\alpha))$$

and

$$B \leq A$$

with

$$w_1(x) = u(x) + \delta e^{x_1/\lambda} \quad \text{and} \quad w_2(y) = -v(y)$$

(see Problem 32). We now apply Ishii's lemma (with $\varepsilon = 1/\alpha$) to $w_1$, $w_2$ and see that there are matrices $X_\alpha, Y_\alpha$ with

$$(p_1, X_\alpha) \in \bar{\mathcal{J}}^{2,+}(u(x_\alpha) + \delta e^{x_{\alpha,1}/\lambda}) \quad \text{and} \quad (p_2, Y_\alpha) \in \bar{\mathcal{J}}^{2,-}(v(y_\alpha))$$

such that

$$-(\alpha + \|A\|)\, I \leq \begin{bmatrix} X_\alpha & 0 \\ 0 & -Y_\alpha \end{bmatrix} \leq A + \alpha^{-1} A^2$$

where we used $B \leq A$.

From $A = \alpha \begin{bmatrix} I & -I \\ -I & I \end{bmatrix}$ and elementary calculations we see that this implies

$$-3\alpha I \leq \begin{bmatrix} X_\alpha & 0 \\ 0 & -Y_\alpha \end{bmatrix} \leq 3\alpha \begin{bmatrix} I & -I \\ -I & I \end{bmatrix}.$$

Thus, we are in the setting of the structure condition. From Problem 30 we see that

$$(p_1 - \delta\lambda^{-1} e^{x_{\alpha,1}/\lambda} \boldsymbol{e}_1, X_\alpha - \delta\lambda^{-2} e^{x_{\alpha,1}} \boldsymbol{e}_1 \otimes \boldsymbol{e}_1) \in \bar{\mathcal{J}}^{2,+} u(x_\alpha).$$

We now use the sub- and supersolution properties of $u$ and $v$ and see that

$$0 \leq F(x_\alpha, X_\alpha - \delta\lambda^{-2} e^{x_{\alpha,1}/\lambda} \boldsymbol{e}_1 \otimes \boldsymbol{e}_1) - F(y_\alpha, Y_\alpha).$$

The uniform ellipticity implies

$$F(x_\alpha, X_\alpha) - F(x_\alpha, X_\alpha - \delta\lambda^{-2} e^{x_{\alpha,1}/\lambda} \boldsymbol{e}_1 \otimes \boldsymbol{e}_1) \geq \mathcal{P}^-(\delta\lambda^{-2} e^{x_1/\lambda} \boldsymbol{e}_1 \otimes \boldsymbol{e}_1) = \delta\lambda^{-2} e^{x_1/\lambda}.$$

Thus,

$$0 \leq F(x_\alpha, X_\alpha) - F(y_\alpha, Y_\alpha) - \delta\lambda^{-2} e^{x_{\alpha,1}/\lambda}.$$

We now use the structure condition, which leads to

$$0 \leq F(x_\alpha, X_\alpha) - F(y_\alpha, Y_\alpha) \leq \omega_F\Big(|x_\alpha - y_\alpha|(1 + \alpha|x_\alpha - y_\alpha|)\Big) - \delta\lambda^{-2} e^{x_{\alpha,1}/\lambda}.$$

In the limit $\alpha \to \infty$, the $\omega_F$ term vanishes by (9). This implies $\lambda \leq 0$, which contradicts the uniform ellipticity. Thus, the initial assumption $\theta > 0$ cannot hold. This completes the proof.

□

## Synopsis of §7.

We have restricted our attention to uniformly elliptic PDEs without dependence of $u(x)$ and $\nabla u(x)$. For those we have shown the comparison principle provided the structure condition (Definition 2.46) holds. The main tool in the proof is Ishii's lemma, which allows to select suitable elements $X_\alpha$, $Y_\alpha$ from the jet closures of $u$ at $x_\alpha$, $y_\alpha$.

# Problems to §7

**Problem 33.** Prove that uniform ellipticity implies degenerate ellipticity.

**Problem 34** (properties of Pucci's operators I)**.** Let $0 < \lambda \leq \Lambda$ and $M \in \mathbb{S}^{n\times n}$. Let the eigenvalues of $M$ be denoted by $\alpha_1, \dots, \alpha_n$. Prove that

$$\mathcal{P}^-(M,\lambda,\Lambda) = \lambda \sum_{\alpha_j>0} \alpha_j + \Lambda \sum_{\alpha_j<0} \alpha_j \quad \text{and} \quad \mathcal{P}^+(M,\lambda,\Lambda) = \Lambda \sum_{\alpha_j>0} \alpha_j + \lambda \sum_{\alpha_j<0} \alpha_j.$$

**Problem 35** (properties of Pucci's operators II)**.** Let $M, N \in \mathbb{S}^{n\times n}$. Prove

1. $\mathcal{P}^-(M) \leq \mathcal{P}^+(M)$
2. $\mathcal{P}^-(M,\lambda',\Lambda') \leq \mathcal{P}^-(M,\lambda,\Lambda)$ and $\mathcal{P}^+(M,\lambda',\Lambda') \geq \mathcal{P}^-(M,\lambda',\Lambda')$ if $\lambda' \leq \lambda \leq \Lambda \leq \Lambda'$
3. $\mathcal{P}^-(M) = -\mathcal{P}^+(-M)$
4. $\mathcal{P}^\pm(\alpha M) = \alpha \mathcal{P}^\pm(M)$ if $\alpha \geq 0$
5. $\mathcal{P}^+(M) + \mathcal{P}^-(N) \leq \mathcal{P}^+(M+N) \leq \mathcal{P}^+(M) + \mathcal{P}^+(N)$
6. $\mathcal{P}^-(M) + \mathcal{P}^-(N) \leq \mathcal{P}^-(M+N) \leq \mathcal{P}^-(M) + \mathcal{P}^+(N)$
7. $\lambda\|N\| \leq \mathcal{P}^-(N,\lambda,\Lambda) \leq \mathcal{P}^+(N,\lambda,\Lambda) \leq n\Lambda\|N\|$ if $N \geq 0$
8. $\mathcal{P}^-$ and $\mathcal{P}^+$ are uniformly elliptic with ellipticity constants $\lambda$, $n\Lambda$.

**Problem 36.** For matrices $A, B \in \mathbb{R}^{n\times n}$ we define the Frobenius inner product $A : B = \sum_{j,k=1}^n A_{jk}B_{jk}$. Prove that $A : B = \operatorname{tr}(AB)$ and $x^\top Ay = A : x \otimes y$ for $x, y \in \mathbb{R}^n$. *(Recall $x \otimes y = xy^\top$.)*

**Problem 37** (Hamilton–Jacobi–Bellman operator)**.** Let $\mathcal{A}$ be an index set and $0 < \lambda \leq \Lambda$ be given. Let, for any $\alpha \in \mathcal{A}$, $A_\alpha : \Omega \to \mathbb{S}^{n\times n}$ be measurable and bounded $0 < \lambda I \leq A_\alpha \leq \Lambda I$ uniformly in $\Omega$; and let $f_\alpha \in L^\infty(\Omega)$. Prove that the operator

$$F(x, D^2u(x)) = \inf_{\alpha\in\mathcal{A}} \big(\operatorname{tr}(A_\alpha D^2u(x)) - f_\alpha(x)\big)$$

is uniformly elliptic.

**Problem 38.** Prove Ishii's lemma under the assumption $w_1, w_2 \in C^2(\bar{\Omega})$.

## §8 Semiconvex functions: Jensen's lemma and sup-convolutions *(week 50)*

We work towards the proof of Ishii's lemma. To this end we study semiconvex functions and sup-convolutions.

**Definition 2.49.** Let $\Omega \subseteq \mathbb{R}^n$. A function $f : \Omega \to \mathbb{R}$ is called *semiconvex*, if there exists $\varepsilon \in (0, \infty]$ such that

$$x \mapsto f(x) + \frac{|x|^2}{2\varepsilon}$$

is convex. The quantity

$$\inf\{\frac{1}{\varepsilon} : x \mapsto f(x) + \frac{|x|^2}{2\varepsilon} \text{ is convex}\}$$

is called *semiconvexity constant.*

**Proposition 2.50** (continuity)**.** *Semiconvex functions over open domains are continuous.*

*Proof.* It suffices that convex functions are continuous. This is shown in Problem 40. □

**Theorem 2.51** (Jensen's lemma)**.** *Let $f : \Omega \to \mathbb{R}$ be semiconvex with constant $\mu > 0$ (so that $f + |\cdot|^2/(2\mu)$ is convex). Suppose $f$ has a strict local maximum at $x \in \Omega$. Given $p \in \mathbb{R}^n$, we denote by $f_p$ the function $f_p(z) = f(z) + p \cdot (z - x)$. We define the set*

$$K_{\delta,\rho} := \{y \in B_\rho(x) : \text{there is some } p \in B_\delta(0) \text{ s.t. } f_p \text{ has a local maximum at } y\}$$

*for parameters $\delta, \rho > 0$. For sufficiently small $\rho > 0$ there exists $\delta = \delta(\rho) > 0$ such that there is the following lower bound on the Lebesgue measure*

$$\mathcal{L}^n(K_{\delta,\rho}) \geq \alpha(n)(\mu\delta)^n$$

*where $\alpha(n)$ is the volume of the unit ball of $\mathbb{R}^n$.*

*Proof.* Let $\rho > 0$ be so small that $f(x) > f(z)$ for all $z \in \bar{B}_\rho(x)$. Then there exists $\gamma(\rho) > 0$ such that

$$f(x) - \max_{\bar{B}_\rho(x)\setminus B_{\rho/2}(x)} f \geq \gamma(\rho).$$

Let $p \in \mathbb{R}^n$ with $|p| = \delta$. For any $z \in \bar{B}_\rho(x) \setminus B_{\rho/2}(x)$ we can estimate

$$\max_{\bar{B}_\rho(x)} f_p - f_p(z) \geq f(x) - \max_{\bar{B}_\rho(x)\setminus B_{\rho/2}(x)} f_p \geq f(x) - \max_{\bar{B}_\rho(x)\setminus B_{\rho/2}(x)} f - \delta\rho \geq \gamma(\rho) - \delta\rho.$$

If $\delta$ is small enough such that $\delta\rho < \gamma(\rho)$, this implies that

$$\max_{\bar{B}_\rho(x)} f_p > \max_{\bar{B}_\rho(x)\setminus B_{\rho/2}(x)} f_p.$$

Hence, $f_p$ has a local maximum at some interior point $y \in B_\rho(x)$, and thus $y \in K_{\delta,\rho}$. Assume first that $f \in C^2(\Omega)$. Then, at the local maximum $y$, we have $\nabla f_p(y) = 0$ and therefore

$p = -\nabla f(y)$. In summary, we have shown that for any $p$ with $|p| < \delta$, there is $y \in K_{\delta,\rho}$ such that $p = -\nabla f(y)$, which means

$$B_\delta(0) \subseteq \nabla f(K_{\delta,\rho}). \tag{10}$$

The $\mu$-semiconvexity of $f$ and the maximality furthermore imply

$$-\mu^{-1} I \leq D^2 f(y) \leq 0.$$

Taking the product of all eigenvalues leads to

$$|\det D^2 f| \leq \mu^{-n} \quad \text{on } K_{\delta,\rho}.$$

From the inclusion (10), the change-of-variables formula (with inequality because $\nabla f$ may be not one-to-one), and the last displayed estimate we obtain

$$\mathcal{L}^n(B_\delta(0)) \leq \mathcal{L}^n(\nabla f(K_{\delta,\rho})) \leq \int_{K_{\delta,\rho}} |\det D^2 f| d\mathcal{L}^n \leq \mathcal{L}^n(K_{\delta,\rho})\mu^{-n}$$

This proves the assertion for the case of $f \in C^2(\Omega)$.

For a general $f \in C(\Omega)$, we consider the regularization $f_\varepsilon = f * \eta_\varepsilon$ by convolution with a standard mollifier $\eta_\varepsilon$. For such $f_\varepsilon$, the first part of the proof shows the asserted bound for the corresponding set $K^\varepsilon_{\delta,\rho}$. We then have from locally uniform convergence $f_\varepsilon \to f$ for $\varepsilon \to 0$ over $B_\rho$ that

$$\bigcap_{j=1}^{\infty} \bigcup_{k=j}^{\infty} K^{1/k}_{\delta,\rho} \subseteq K_{\delta,\rho}$$

(prove this as an exercise, see Problem 41). Moreover, from the proven lower bound, elementary properties of the measure of the limsup of sets, and the stated inclusion we infer

$$\mathcal{L}^n(B_\delta(0)) \leq \lim_{j\to\infty} \mathcal{L}^n\left(\bigcup_{k=j}^{\infty} K^{1/k}_{\delta,\rho}\right) = \mathcal{L}^n\left(\bigcap_{j=1}^{\infty} \bigcup_{k=j}^{\infty} K^{1/k}_{\delta,\rho}\right) \leq K_{\delta,\rho},$$

which proves the assertion in the general case. □

**Definition 2.52** (sup-convolution)**.** Let $\Omega \subseteq \mathbb{R}^n$ be bounded and let $u : \bar{\Omega} \to \mathbb{R}$. For given $\varepsilon > 0$ we define the function $u^\varepsilon : \Omega \to \mathbb{R}$ at any $x \in \Omega$ by

$$u^\varepsilon(x) := \sup_{y\in\Omega} \left\{ u(y) - \frac{|x-y|^2}{2\varepsilon} \right\}.$$

The function $u^\varepsilon$ is called the *sup-convolution* of $u$.

**Proposition 2.53** (sup-convolution is semiconvex)**.** *Let $\Omega \subseteq \mathbb{R}^n$ be bounded. The sup-convolution $u^\varepsilon$ of any given function $u : \bar{\Omega} \to \mathbb{R}$ is semiconvex.*

*Proof.* We rewrite the definition of $u^\varepsilon$

$$u^\varepsilon(x) = \sup_{y\in\Omega}\left\{u(y) - \frac{|x-y|^2 - |x|^2}{2\varepsilon}\right\} - \frac{|x|^2}{2\varepsilon}.$$

and rearrange as follows

$$u^\varepsilon(x) + \frac{|x|^2}{2\varepsilon} = \sup_{y\in\Omega}\left\{(u(y) - \frac{|y|^2}{2\varepsilon}) + \varepsilon^{-1} y\cdot x\right\}.$$

This means that the function $u^\varepsilon(x) + \frac{|x|^2}{2\varepsilon}$ is the supremum of a family of affine functions. Thus, it is convex and so $u^\varepsilon$ is semiconvex. □

**Remark 2.54.** From Proposition 2.50 we deduce that sup-convolutions are continuous.

**Proposition 2.55** (magic property)**.** *Let $\Omega \subseteq \mathbb{R}^n$ be bounded and $u \in USC(\bar\Omega)$ with $(p, X) \in \mathcal{J}^{2,+}u^\varepsilon(x)$. Then*

$$(p, X) \in \mathcal{J}^{2,+}u(x+\varepsilon p) \quad \textit{and} \quad u^\varepsilon(x) + \frac{\varepsilon}{2}|p|^2 = u(x+\varepsilon p).$$

*Proof.* Let $(p, X) \in \mathcal{J}^{2,+}u^\varepsilon(x)$. From Theorem 2.30 we know that there is $\psi \in C^2(\mathbb{R}^n)$ such that

$$u^\varepsilon - \psi \le (u^\varepsilon - \psi)(x)$$

with $\nabla\psi(x) = p$ and $D^2\psi(x) = X$. We choose $y$ such that $u^\varepsilon(x) = u(y) - 1/(2\varepsilon)|y-x|^2$. Then, for all $z, \xi$,

$$\begin{aligned} u(\xi) - \frac{1}{2\varepsilon}|z-\xi|^2 &\le u^\varepsilon(z) \\ &\le u^\varepsilon(x) + p\cdot(z-x) + \frac{1}{2}(z-x)^\top X(z-x) + o(|z-x|^2) \\ &= u(y) - \frac{1}{2\varepsilon}|y-x|^2 + p\cdot(z-x) + \frac{1}{2}(z-x)^\top X(z-x) + o(|z-x|^2). \end{aligned}$$

Choosing $z = \xi - y + x$ we see that $(p, X) \in \mathcal{J}^{2,+}u(y)$. Choosing $\xi = y$ and $z = x - \beta(\varepsilon^{-1}(x-y) + p)$ above we obtain

$$-\frac{1}{2\varepsilon}|x-y-\beta(\varepsilon^{-1}(x-y)+p)|^2 \le -\frac{1}{2\varepsilon}|y-x|^2 - p\cdot(\beta(\varepsilon^{-1}(x-y)+p)) + O(\beta^2).$$

After rearranging terms, we arrive at

$$\beta(\varepsilon^{-1}(x-y)+p)\cdot(\varepsilon^{-1}(x-y)+p) \le O(\beta^2)$$

This proves $y = x + \varepsilon p$. □

## Synopsis of §8.

We have proven Jensen's lemma on semiconvex functions. We have defined the regularization by sup-convolution and shown that it satisfies the *magic property*.

## Problems to §8

**Problem 39.** Let $\mathcal{A}$ be a family of affine functions over $\mathbb{R}^n$. Prove that $\sup \mathcal{A}$ is a convex.

**Problem 40.** Let $\Omega \subseteq \mathbb{R}^n$ an open domain and $u : \Omega \to \mathbb{R}$ be convex. Prove that $u$ is locally Lipschitz continuous.

*Instruction (if needed):* To show Lipschitz continuity near $x_0 \in \Omega$, let $B_{2r}(x_0) \subseteq \Omega$ be an open ball with $x, y \in B_r(x_0)$ and define $z := x + \alpha(x - y)$ with $\alpha = r/(2|x - y|)$. Show $x = (1+\alpha)^{-1}z + \alpha(1+\alpha)^{-1}y$ and use this result to first estimate $f(x) - f(y) \leq (\alpha+1)^{-1}(f(z) - f(y))$ and then establish the Lipschitz bound. Prove the estimate for $|f(x) - f(y)|$ by interchanging the roles of $x$, $y$.

**Problem 41.** Prove the inclusion

$$\bigcap_{j=1}^{\infty} \bigcup_{k=j}^{\infty} K_{\delta,\rho}^{1/k} \subseteq K_{\delta,\rho}$$

claimed in the proof of Jensen's lemma.

## §9 Proof of Ishii's lemma *(week 1/2022)*

We quote a result on (semi)convex functions without proving it (the proof is nontrivial and is better worked worked out in separate seminar).

**Theorem 2.56** (Alexandrov). *Let $f : \Omega \to \mathbb{R}$ be semiconvex. Then $f$ possess second-order derivatives almost everythere in $\Omega$ in the sense that for a.e. $x \in \Omega$ there is $(p, X) \in \mathbb{R}^n \times \mathbb{S}^{n\times n}$ such that*

$$f(x+z) = f(x) + p \cdot z + \frac{1}{2} z^\top X z + o(|z|^2).$$

We start with a technical lemma.

**Lemma 2.57.** *Let $\Omega \subseteq \mathbb{R}^n$ with $0 \in \Omega$. Let $f : \Omega \to \mathbb{R}^n$ be $\mu$-semiconvex. If there exists $B \in \mathbb{S}^{n\times n}$ such that*

$$f(z) \le f(0) + \frac{1}{2} z^\top B z \quad \textit{for } |z| \textit{ small},$$

*then there exists $X \in \mathbb{S}^{n\times n}$ such that*

$$(0, X) \in \bar{\mathcal{J}}^{2,+} f(0) \cap \bar{\mathcal{J}}^{2,-} f(0)$$

*and*

$$-\mu^{-1} I \le X \le B.$$

*Proof.* For any $\mu > 0$, the function

$$g_\mu(z) := f(z) - \frac{1}{2} z^\top B z - \frac{\mu}{2} |z|^2$$

has a strict maximum at $z = 0$. By Jensen's lemma, the set of $|y_\mu| < \mu$ such that there is $|p_\mu| < \delta$ (for any $\delta_\mu < \delta_0(\mu)$) such that

$$g_\mu(z) + p_\mu \cdot z \text{ is maximal at } y_\mu$$

has positive measure. Since, by Alexandrov's theorem, $g_\mu$ is twice differentiable a.e. in $\Omega$, there exists such a $y_\mu$ where $g_\mu$ is twice differentiable. We thus have $\nabla g_\mu(y_\mu) = -p_\mu$ and so

$$\nabla f(y_\mu) = \nabla g_\mu(y_\mu) + B y_\mu + \mu y_\mu \to 0 \quad \text{as } \mu \to 0.$$

Furthermore the $\mu$-semiconvexity and $D^2 g_\mu(y_\mu) \le 0$ imply

$$-\mu^{-1} I \le D^2 f(y_\mu) \le B + \mu I.$$

The differentiability at $y_\mu$ clearly implies

$$(\nabla f(y_\mu), D^2 f(y_\mu)) \in J^{2,+} f(y_\mu) \cap J^{2,-} f(y_\mu).$$

The proof is thus concluded by taking the limit $\mu \to 0$. □

***Proof of Ishii's lemma (Theorem 2.47).*** Without loss of generality we assume that $0 \in \Omega$ and $x = 0 = y$ as well as $w_1(x) = 0 = w_2(y)$ and $p_1 = 0 = p_2$ (otherwise consider $w(z_1, z_2) - w(0,0) - p_1 \cdot z_1 - p_2 \cdot z_2$). From $((0,0), A) \in \mathcal{J}^{2,+}w(0,0)$ we see that

$$w(z) \leq \frac{1}{2} z^\top A z + o(|z|^2) \leq \frac{1}{2} z^\top (A + \sigma I) z \text{ for small } \sigma \text{ and } |z| < c(\sigma).$$

This means $(A + \sigma I) \in \mathcal{J}^{2,+}w(0,0)$ for small $\sigma$. Once the assertion is shown for $A_\sigma := A + \sigma I$, it will follow for $A$ for $\sigma \to 0$. Without loss of generality we therefore assume $A = A_\sigma$ and $w(z) \leq \frac{1}{2} z^\top A z$ for small $|z|$. Since the jets $\mathcal{J}^{2,+}w_1(0)$, $\mathcal{J}^{2,+}w_2(0)$, $\bar{\mathcal{J}}^{2,+}w(0,0)$ only depend on local information near 0, we may modify $w_1$, $w_2$ outside some open ball around 0 such that

$$w(y) \leq \frac{1}{2} y^\top A y \quad \text{for all } y \in \bar{\Omega}. \tag{11}$$

We choose

$$\lambda := \left(\varepsilon^{-1} + \|A\|\right)^{-1}$$

and recall the definition of the sup-convolutions $w_1^\lambda$, $w_2^\lambda$, $w^\lambda$ of $w_1$, $w_2$, $w$, namely

$$w_j^\lambda(z) := \sup_{y \in \Omega} \left\{ w_j(y) - \frac{|z-y|^2}{2\lambda} \right\}, \qquad z \in \Omega$$

and

$$w^\lambda(z) := \sup_{y \in \Omega^2} \left\{ w(y) - \frac{|z-y|^2}{2\lambda} \right\}, \qquad z \in \Omega^2.$$

A direct computation with the Euclidean norm shows that $w_1^\lambda(z_1) + w_2^\lambda(z_2) = w^\lambda(z_1, z_2)$. It is shown in Problem 42 that

$$-(\varepsilon^{-1} + \|A\|)|z-y|^2 \leq z^\top (A + \varepsilon A^2) z - y^\top A y.$$

The definitions of $w^\lambda$ and $\lambda$ thus show

$$w^\lambda(z) = \sup_{y \in \Omega^2} \left\{ w(y) - \frac{1}{2}(\varepsilon^{-1} + \|A\|)|z-y|^2 \right\} \leq \sup_{y \in \Omega^2} \left\{ w(y) - \frac{1}{2} y^\top A y \right\} + \frac{1}{2} z^\top (X + \varepsilon A^2) z.$$

By the assumption on $A$ from (11) the sup-term on the right-hand side is nonpositive, whence

$$w^\lambda(z) \leq \frac{1}{2} z^\top (A + \varepsilon A^2) z.$$

Since the sup-convolutions $w_1^\lambda$, $w_2^\lambda$ are $\lambda$-semiconvex, we can apply Lemma 2.57 with $B := A + \varepsilon A^2$ with $z_2 = 0$ or $z_1 = 0$. This yields the existence of

$$(0, X) \in \bar{\mathcal{J}}^{2,+}w_1^\lambda(0) \cap \bar{\mathcal{J}}^{2,-}w_1^\lambda(0) \quad \text{and} \quad (0, Y) \in \bar{\mathcal{J}}^{2,+}w_2^\lambda(0) \cap \bar{\mathcal{J}}^{2,-}w_2^\lambda(0)$$

with (recall the definition of $\lambda$)

$$-(\varepsilon^{-1} + \|A\|)I = -\lambda^{-1} I \leq \begin{bmatrix} X & 0 \\ 0 & Y \end{bmatrix} \leq B = A + \varepsilon A^2.$$

Then, the *magic property* from Proposition 2.55 implies

$$(0, X) \in \bar{\mathcal{J}}^{2,+}w_1(0) \cap \bar{\mathcal{J}}^{2,-}w_1(0) \quad \text{and} \quad (0, Y) \in \bar{\mathcal{J}}^{2,+}w_2(0) \cap \bar{\mathcal{J}}^{2,-}w_2(0),$$

which concludes the proof. $\square$

**Synopsis of §9.**

We have completed the proof of Ishii's lemma.

## Problems to §9

**Problem 42.** Let $X \in \mathbb{S}^{n\times n}$, $a, b \in \mathbb{R}^n$, and $\varepsilon > 0$. As usual, $\|X\| = \max|\sigma(X)|$ is the spectral norm of $X$ (natural matrix norm w.r.t. the Euclidean scalar product).

(i) Prove

$$y^\top X y = z^\top X z + (y-z)^\top X(y-z) + 2(y-z)^\top X z.$$

(ii) Prove that any $a, b \in \mathbb{R}$ satisfy $2ab \le \varepsilon^{-1}a^2 + \varepsilon b^2$. *(Hint: binomial formula.)*

(iii) Prove

$$y^\top X y \le z^\top X z + \|X\|\,|z-y|^2 + \varepsilon^{-1}|z-y|^2 + \varepsilon|Xz|^2.$$

(iv) Prove

$$y^\top X y \le z^\top (X + \varepsilon X^2) z + (\varepsilon^{-1} + \|X\|)|z-y|^2.$$

# Topic 3: Monotone finite differences

## §10 Abstract convergence; a model problem *(week 2)*

We write the boundary-value problem from Definition 2.38 in a compact format. We define

$$\mathsf{F}(x,r,p,X) = \begin{cases} F(x,r,p,X) & \text{if } x \in \Omega \\ g(x) - r & \text{if } x \in \partial\Omega. \end{cases}$$

and rewrite the boundary-value problem as

$$\mathsf{F}[u] := \mathsf{F}(x, u(x), \nabla u(x), D^2 u(x)) = 0 \quad \text{in the viscosity sense} \quad \text{for all } x \in \bar\Omega. \tag{12}$$

As usual, we assume $F$ to be continuous, but note that $\mathsf{F}$ is discontinuous at the boundary.

We introduce the following compact notation: Let a mesh size $h > 0$ be given and let

$$\mathbb{Z}^n_h := \{he : e \in \mathbb{Z}^n\}$$

denote the scaled $n$-dimensional integer grid. We then define

$$\bar\Omega_h := \mathbb{Z}^n_h \cap \bar\Omega$$

as the finite difference grid over the domain, which we know from prior sections in the two-dimensional case $n = 2$. The space of grid functions, that is the space of all functions $\bar\Omega_h \to \mathbb{R}$, is denoted by $X_h$. The discrete domain $\bar\Omega_h \subseteq \bar\Omega$ then approximates $\bar\Omega$ in the sense that for any $z \in \bar\Omega$ there is a sequence $(z_h)_h$ such that $z_h \to z$ as $h \to 0$.

We seek an approximation $u_h \in X_h$ to $u$ satisfying

$$\mathsf{F}_h[u_h](z) = 0 \quad \text{for all } z \in \bar\Omega_h. \tag{13}$$

Here $\mathsf{F}_h : X_h \to \mathbb{R}$ is a map which should suitably approximate $\mathsf{F}$. The precise structure is not relevant to the abstract arguments discussed here. However, we assume for simplicity that the grid exactly matches with the domain's boundary in the sense that the set $\Gamma$ of boundary grid points satisfies $\Gamma = \bar\Omega_h \cap \partial\Omega$. Then we can evaluate the boundary data $g$ over $\Gamma$. We will always assume that

$$\mathsf{F}_h[u_h](z) = g(z) - u(z) \quad \text{for all } z \in \Gamma,$$

that is, we interpolate the boundary data.

We now generalize the notions of consistency and stability which we already have encountered in the Laplacian case.

**Definition 3.58** (consistency)**.** The discrete problem (13) is *consistent* with (12) if there exists an operator $I_h : C(\bar\Omega) \to X_h$ such that $I_h$ converges uniformly to the identity as $h \searrow 0$, and for any sequence $(z_h)_h$ with $z_h \in \bar\Omega_h$ and $z_h \to z_0 \in \bar\Omega$ and $\phi \in C^2(\bar\Omega)$

$$\lim_{h \searrow 0} \mathsf{F}_h[I_h\phi](z_h) = \mathsf{F}[\phi](z_0).$$

**Remark 3.59.** In FDM, the operator $I_h$ is usually the interpolation in the grid points.

**Definition 3.60** (stability)**.** Problem (13) is said to be *stable* if for any $h > 0$ there exists a solution $u_h \in X_h$ to (13) and the following bound holds

$$|u_h - w_h|_{\infty,\bar{\Omega}_h} \le C|\mathsf{F}_h[w_h]|_{\infty,\bar{\Omega}_h} \quad \text{for any } w_h \in X_h$$

with a constant $C > 0$ independent on $h$ or $w_h$.

For $C^2$ solutions to the Laplacian we formulated the rule that stability and consistency are sufficient for convergence of the FDM. In the case of viscosity solutions, we need *monotonicity* as a third ingredient.

**Definition 3.61** (monotonicity)**.** The discrete operator $\mathsf{F}_h$ is said to be *monotone* if the following property is satisfied for any $u_h, v_h \in X_h$: if $u_h - v_h$ has a global non-negative maximum at some $z \in \bar{\Omega}_h$, then

$$\mathsf{F}_h[u_h](z) \le \mathsf{F}_h[v_h](z).$$

**Remark 3.62.** With our simplifying assumption above that $F_h[u_h](z) = g(z) - u_h(z)$ for all grid points $z \in \Gamma$ on the boundary, the monotonicity needs only be verified on interior grid points. Indeed, if $z \in \Gamma$ and $u_h - v_h$ has a global nonnegative maximum at $z$, then $u_h(z) - v_h(z) \ge 0$. Thus

$$F_h[v_h](z) - F_h[u_h](z) = g(z) - v_h(z) - g(z) + u_h(z) = u_h(z) - v_h(z) \ge 0.$$

From the sequence of discrete solutions (which exist by the stability assumption) we pass to the limits

$$\bar{u}(x) := \limsup_{\substack{y \to x \\ h \searrow 0}} u_h(y) \quad \text{and} \quad \underline{u}(x) := \liminf_{\substack{y \to x \\ h \searrow 0}} u_h(y) \quad \text{for any } x \in \bar{\Omega}.$$

From the stability (with $w_h = 0$ in Definition 3.60) we infer that these functions are indeed finite-valued. By construction we have $\bar{u} \in USC(\bar{\Omega})$ and $\underline{u} \in LSC(\bar{\Omega})$.

Next, we prove the principal result on monotone finite differences.

**Theorem 3.63** (Barles–Souganidis 1991)**.** *Let $\mathsf{F}$ satisfy the comparison principle and let $\mathsf{F}_h$ be consistent, stable, and monotone. Assume furthermore that $\bar{u} \le g$ and $\underline{u} \ge g$ on $\partial\Omega$. Then, $u_h$ converges locally uniformly to the (unique) viscosity solution to* (12).

*Proof.* We shall prove that the functions $\bar{u}$ and $\underline{u}$ are sub- and supersolution, respectively.

Fix an interior point $z_0 \in \Omega$ and let $\phi \in C^2(\Omega)$ be such that $\bar{u} - \phi$ has a strict local maximum at $z_0$. Without loss of generality (by modifying $\phi$ outside some ball around $z_0$) we may assume that $\bar{u} - \phi$ has a strict global maximum at $z_0$. Then there exist a sequence $(h_k)_k$ of mesh sizes and a sequence of grid points $z_{h_k}$ such that

$$h_k \to 0, \quad z_{h_k} \to z_0 \quad \text{as } k \to \infty$$

and the grid function

$$u_{h_k} - I_{h_k}\phi \text{ has a strict global maximum at } z_{h_k}$$

(details are worked out as Problem 44). The monotonicity thus implies

$$0 = \mathsf{F}_{h_k}[u_{h_k}](z_{h_k}) \leq \mathsf{F}_{h_k}[I_{h_k}\phi](z_{h_k}).$$

We pass to the limit $k \to \infty$ and use the consistency to infer

$$0 \leq \lim_{k\to\infty} \mathsf{F}_{h_k}[I_{h_k}\phi](z_{h_k}) = \mathsf{F}[\phi](z_0).$$

Since $z_0 \in \Omega$ is an interior point and $F$ (the original PDE operator not including the boundary conditions) is assumed to be continuous, we have that $\bar{u}$ is subsolution to (12) at $z_0$. An analogous argument shows that $\underline{u}$ is a supersolution. The assumption $\bar{u} \leq g$ and $\underline{u} \geq g$ on $\partial\Omega$ and the comparison principle finally show that $\bar{u} = \underline{u} = u$ is the viscosity solution to (12). □

**Warning 3.64.** For simplicity, we have put $\bar{u} \leq g$ and $\underline{u} \geq g$ as an assumption in the formulation of our theorem. Boundary conditions are an issue as they are generally also posed in a viscosity sense and need not be satisfied pointwise. We do not discuss this point in the lecture and only deal with boundary conditions in the pointwise sense. But the reader should be aware that the above assumption must be verified on a case-by-case basis depending on $F$.

In the forthcoming sections we will mainly focus on a simplified setting of a linear model problem $F(\cdot, D^2u) = \mathcal{L}u - f = 0$ where $\mathcal{L}u := A : D^2u$ for a continuous, bounded, and uniformly symmetric positive definite matrix function $A : \Omega \to \mathbb{S}^{n\times n}$, a right-hand side $f \in C(\bar{\Omega})$, and homogeneous Dirichlet boundary conditions.

We now show well-posedness in the following model situation. We assume that there exists a Lipschitz constant $M_1$ and functions $\sigma_j : \Omega \to \mathbb{R}$ such that

$$A_{jk}(x) = \sigma_j\sigma_k \quad \text{and} \quad |\sigma_j(x) - \sigma_j(y)| \leq M_1|x-y| \quad \text{for all } x, y \in \Omega, \quad j, k = 1, \dots, n.$$

We further assume that there is a continuous real function $\omega$ with $\omega(0) = 0$ such that

$$|f(x) - f(y)| \leq \omega(|x-y|) \quad \text{for all } x, y \in \Omega.$$

Let $X, Y \in \mathbb{S}^{n\times n}$ and $\mu > 1$ satisfy

$$-3\mu \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} \leq \begin{bmatrix} X & 0 \\ 0 & -Y \end{bmatrix} \leq 3\mu \begin{bmatrix} I & -I \\ -I & I \end{bmatrix}.$$

We define vectors $\xi = (\sigma_1(x), \dots, \sigma_n(x))$ and $\eta = (\sigma_1(y), \dots, \sigma_n(y))$ and compute

$$\begin{bmatrix} \xi \\ \eta \end{bmatrix}^\top \begin{bmatrix} X & 0 \\ 0 & -Y \end{bmatrix} \begin{bmatrix} \xi \\ \eta \end{bmatrix} \leq 3\mu \begin{bmatrix} \xi \\ \eta \end{bmatrix}^\top \begin{bmatrix} I & -I \\ -I & I \end{bmatrix} \begin{bmatrix} \xi \\ \eta \end{bmatrix} = 3\mu|\xi - \eta|^2 \leq 3n\mu M_1^2|x-y|^2.$$

Thus, using $A_{jk} = \sigma_j \sigma_k$, we obtain

$$\begin{aligned} F(x,X) - F(y,Y) &\leq |f(y) - f(x)| + (A(x) : X - A(y) : Y) \\ &= |f(y) - f(x)| + \xi^\top X \xi - \eta^\top Y \eta \leq \omega(|x-y|) + 3n\mu M_1^2 |x-y|^2. \end{aligned}$$

We have thus shown that the structure condition holds and the problem therefore satisfies the comparison principle.

**Remark 3.65.** The assumptions on $A$ can be weakened. It suffices that $A$ can be uniformly approximated by matrices of the above structure.

### Synopsis of §10.

We have shown that stability, consistency, and monotonicity imply convergence to viscosity solutions. We further formulated a linear model problem and verified that it it satisfies the structure assumptions and thus it is well posed.

# Problems to §10

**Problem 43.** Prove that the 5-point stencil for the Laplacian is consistent, stable, and monotone.

**Problem 44.** Prove the following detail of the abstract convergence theorem. Let $z_0 \in \Omega$ and $\phi \in C^2(\Omega)$ be such that $\bar{u} - \phi$ has a strict global maximum at $z_0$. Then there exist a sequence $(h_k)_k$ of mesh sizes and a sequence of grid points $z_{h_k}$ such that

$$h_k \to 0, \quad z_{h_k} \to z_0 \quad \text{as } k \to \infty$$

and the grid function

$$u_{h_k} - I_{h_k}\phi \text{ has a strict global maximum at } z_{h_k}.$$

**Problem 45.** Write a routine that provides, for given $M \geq 0$, all generalized neighbours

$$\{z + hy : y \in \mathbb{Z}^2, |y|_\infty \leq M\}.$$

of a given vertex $z$. Visualize the results.

## §11 Non-negative operators *(week 3)*

In view of the sufficient criteria formulated in Theorem 3.63, we will construct monotone finite difference methods for our model problem. In general, simple methods like the five-point stencil do not enjoy all those properties simultaneously. Thus, we have to invest a little more work. It will turn out that our finite differences have to consider more than only the first-order neighbours of a grid point.

As before, we think of $\Omega \subseteq \mathbb{R}^n$ being a simple domain like a square/box/etc. Our focus is not on the approximation of complicated boundaries but rather on the design of finite difference stencils. Any finite subset $S \Subset \mathbb{Z}^n$ is called a *stencil*. The stencils we shall work with are always assumed to be of the format

$$S = \{y \in \mathbb{Z}^n \setminus \{0\} : |y|_\infty \leq m\}$$

with some integer $m$, which is referred to as the *stencil size* of $S$. The notation $|\cdot|_\infty$ indicates the usual maximum norm of a vector. The stencil prescribes the dependency of the discrete operator at some grid point $z$ on the neighbours. The format is thus given by

$$F_h[v_h](z) = F_h(z, v_h(z), Tv_h(z)), \quad v_h \in X_h, \tag{14}$$

where $Tv_h(z) := \{v_h(z+hy) : y \in S\}$ is the set of function values in a neighbourhood around $z$ prescribed by the stencil $S$. Here, we use the notation $F_h[v_h](z)$ instead of $\mathsf{F}_h[v_h](z)$ from the foregoing section to highlight that we are concerned with discretization of the PDE operator (at interior grid points) and not of the boundary condition.

We begin with formulating a handy criterion for monotonicity.

**Definition 3.66** (non-negative operator)**.** The operator $F_h$ is of *non-negative type* if, given $z \in \mathbb{R}^n$, $r \in \mathbb{R}$, and $p \in \mathbb{R}^{\operatorname{card} S}$, there holds

$$F_h(z, r+t, p+\tau) \leq F_h(z, r, p) \leq F_h(z, r, p+\tau) \tag{15}$$

for all translations $0 \leq t \in \mathbb{R}$ and $\tau \in \mathbb{R}^{\operatorname{card} S}$ with $|\tau|_\infty \leq t$.

**Lemma 3.67.** *A finite difference operator $F_h$ of the format* (14) *is of non-negative type if and only if it is monotone.*

*Proof.* Let $F_h$ be of non-negative type and let $u_h, v_h \in X_h$ be such that $u_h - v_h$ has a global non-negative maximum at a grid point $z \in \bar{\Omega}_h$. We choose

$$r := v_h(z), \quad t := u_h(z) - v_h(z)$$

as well as $p, \tau \in \mathbb{R}^{\operatorname{card} S}$ given by

$$p_j := v_h(z + hy_j), \quad \tau_j := \max\{0, u_h(z+hy_j) - v_h(z+hy_j)\} \quad \text{for all } 1 \leq j \leq \operatorname{card} S.$$

We observe that $t \geq 0$ and $0 \leq \tau_j \leq t$ because $u_h - v_h$ has a global maximum at $z$. We furthermore note

$$p_j + \tau_j = v_h(z+hy_j) + \max\{0, u_h(z+hy_j) - v_h(z+hy_j)\} \geq u_h(z+hy_j)$$

so that the upper bound in (15) yields

$$F_h[u_h](z) = F_h(z, r+t, u_h(z+hy_j)_{j=1}^{\operatorname{card} S}) \leq F_h(z, r+t, p+\tau).$$

The lower bound in (15) thus implies

$$F_h[u_h](z) \leq F_h(z, r+t, p+\tau) \leq F_h(z, r, p) = F_h[v_h](z).$$

Hence, the operator is monotone.

Let us conversely assume that the operator $F_h$ is monotone, let $z \in \bar{\Omega}_h$ be grid point and let $r, t \in \mathbb{R}$ and $p, \tau \in \mathbb{R}^{\operatorname{card} S}$ conforming to $0 \leq \tau_j \leq t$ as in Definition 3.66 be given. We construct grid functions $u_h, v_h$ by

$$v_h(z) := r, \quad v_h(z+hy_j) := p_j, \quad u_h(z) := r+t, \quad u_h(z+hy_j) := p_j + \tau_j.$$

We then have

$$u_h(z) - v_h(z) = t \geq \tau_j = u_h(z+hy_j) - v_h(z+hy_j)$$

and thus $u_h - v_h$ has a non-negative maximum at $z$. We choose an appropriate extension of the grid functions outside the stencil horizon such that $u_h - v_h$ has a global maximum at $z$. The monotonicity therefore implies

$$F_h(z, r+t, p+\tau) = F_h[u_h](z) \leq F_h[v_h](z) = F_h(z, r, p),$$

which proves the lower bound in (15) required for non-negativity. We define another grid function $w_h$ by

$$w_h(z) = r, \quad w_h(z+hy_j) = p_j + \tau_j.$$

Then, $v_h - w_h$ has a global maximum at $z$ and we obtain from the monotonicity

$$F_h(z, r, p) = F_h[v_h](z) \leq F_h[w_h](z) = F_h(z, r, p+\tau).$$

This proves the upper bound in (15). We conclude that $F_h$ is of non-negative type. □

Our discrete operators will be related to finite differences with respect to the stencil $S$. Similar as in prior sections, we define the first-order difference operators

$$\delta^+_{y,h} u(z) := \frac{1}{h}(u(z+hy) - u(z)), \quad \delta^-_{y,h} u(z) := \frac{1}{h}(u(z) - u(z-hy)),$$

and

$$\delta_{y,h} u(z) := \frac{1}{2}(\delta^+_{y,h} u(z) + \delta^-_{y,h} u(z)) = \frac{1}{2h}(u(z+hy) - u(z-hy))$$

as well as the second-order difference operator

$$\delta^2_{y,h} u(z) := \frac{1}{h^2}(u(z+hy) - 2u(z) + u(z-hy))$$

for any $y \in S$. With proofs similar to those in the case of the five-point stencil one can prove that these operators approximate the respective differential operators, see Problem 46. We denote

$$\delta_h u_h(z) := \{\delta_{y,h} u_h(z) : y \in S\} \quad \text{and} \quad \delta_h^2 u_h(z) := \{\delta_{y,h}^2 u_h(z) : y \in S\}.$$

With these finite differences, we want to construct operators of the format

$$F_h[u_h](z) = \mathcal{F}_h(z, u_h(z), \delta_h u_h(z), \delta_h^2 u_h(z)) \tag{16}$$

with a function

$$\mathcal{F}_h : \Omega_h \times \mathbb{R} \times \mathbb{R}^{\operatorname{card} S} \times \mathbb{R}^{\operatorname{card} S} \to \mathbb{R}.$$

We denote the points in the domain of $\mathcal{F}_h$ by $(z, r, q, s)$ and assume $\mathcal{F}_h$ to be symmetric with respect to $\pm q_{\pm j}$ and $\pm s_{\pm j}$.

From Lemma 3.67 and Definition 3.66 we see that such a scheme is monotone provided it satisfies the sufficient criterion

$$\frac{h}{2}\left|\frac{\partial \mathcal{F}_h}{\partial q_j}\right| \leq \frac{\partial \mathcal{F}_h}{\partial s_j} \text{ for all } j = 1, \dots, \operatorname{card} S, \quad \text{and} \quad \frac{\partial \mathcal{F}_h}{\partial r} \leq 0. \tag{17}$$

This will be worked out as Problem 47

In the continuous case we distinguished between degenerate ellipticity and uniform ellipticity. An analogous criterion formulating stronger conditions on $\mathcal{F}_h$ is as follows.

**Definition 3.68** (positive operator)**.** An operator of the format (16) is of *positive type* if (17) is satisfied and there exists a positive number $\lambda_{0,h} > 0$ and an orthogonal set of vectors $\{y_j\}_{j=1}^n \subseteq S$ such that

$$\lambda_{0,h} + \frac{h}{2}\left|\frac{\partial \mathcal{F}_h}{\partial q_j}\right| \leq \frac{\partial \mathcal{F}_h}{\partial s_j}.$$

Of course, positivity implies non-negativity. Note that the number $\lambda_{0,h}$ may depend on $h$.

## Synopsis of §11.

We have formulated non-negativity, a criterion equivalent to monotonicity. We furthermore defined finite differences with respect to a stencil $S$ and confinded ourselves to discrete operators based on these quantities and formulated a criterion (positivity) slightly stronger than monotonicity.

## Problems to §11

**Problem 46.** Let $u : \mathbb{R}^n \to \mathbb{R}$ be a sufficiently smooth. Prove that the finite difference operators defined above satisfy

$$\left|\frac{\partial u(z)}{\partial y} - \delta^{\pm}_{y,h} u(z)\right| = O(h|y|^2), \quad \left|\frac{\partial u(z)}{\partial y} - \delta_{y,h} u(z)\right| = O(h^2|y|^3)$$

and

$$\left|\frac{\partial^2 u(z)}{\partial^2 y} - \delta^2_{y,h} u(z)\right| = O(h^2|y|^4).$$

Here we use the notation $\partial u/\partial y = \nabla u \cdot y$ and $\partial^2 u/\partial y^2 = y^\top D^2 u y$.

**Problem 47.** Prove that the operator $F_h$ is of non-negative type provided that

$$\frac{\partial F_h}{\partial p_j} \geq 0 \text{ for all } j = 1, \dots, \operatorname{card} S \quad \text{and} \quad \frac{\partial F_h}{\partial r} + \sum_{j=1}^{\operatorname{card} S} \frac{\partial F_h}{\partial p_j} \leq 0.$$

Prove that an operator of the format (16) is monotone provided (17) is satisfied.

**Problem 48.** Recall the 5-point stencil discretization of Poisson's equation. Write the scheme in the formats (14) and (16) and prove that it is of positive type.

## §12 Construction of monotone finite differences *(week 4)*

In the remaining parts of this lecture we will restrict ourselves to the linear model problem $F[u] = \mathcal{L}u - f = 0$ where $\mathcal{L}u := A : D^2u$ for a continuous, bounded, and uniformly symmetric positive definite matrix function $A : \Omega \to \mathbb{S}^{n\times n}$ and $f \in C(\bar{\Omega})$. We confine ourselves to homogeneous Dirichlet boundary conditions. Note that wide stencils may exceed the domain boundary for grid points close to $\partial\Omega_h$. We will not deal with this problem from a theoretical point of view but will rather implement a practical solution in the excercises.

We will assume the discretized operator to be of the format

$$F_h[u_h](z) = \mathcal{L}_h u_h(z) - f(z) := \sum_{y\in S} a_y(z)\delta^2_{y,h}u_h(z) - f(z) \tag{18}$$

where the $a_y(z)$ are scalar coefficients. The criteria from the foregoing section reveal that $F_h$ is of non-negative type if all $a_y(z) \geq 0$ and of positive type if $a_{y_j}(z) \geq \lambda_{0,h} > 0$ for some orthogonal basis $(y_j)_j$.

**Remark 3.69.** One may wonder why we are studying wide-stencil schemes instead of discretizing the problem with a fixed-stencil method, say the 5-point stencil. Even in the linear model case, it is possible to prove that for any fixed stencil width there is a linear operator $\mathcal{L}$ such that any linear, consistent finite difference scheme of the above format (18) is not of positive type. We will not go into the details of this statement but should keep it in mind as a motivation for the design of wide-stencil methods.

We begin with formulating a simple positivity criterion based on a particular structure of the matrix $A$.

**Lemma 3.70.** *Suppose that the coefficient matrix $A$ of the operator $\mathcal{L}u = A : D^2u$ has the form*

$$A(x) = \sum_{\substack{y\in\mathbb{Z}^n \\ |y|_\infty \leq M}} a_y(x) y \otimes y$$

*for some positive integer $M$ and coefficients $a_y(x) \geq 0$ for every $x \in \Omega$. Assume further that there exists an orthogonal basis $(y_j)_j \subseteq \mathbb{Z}^n$ with $|y_j|_\infty \leq M$ such that $a_{y_j} \geq c$ for some $c > 0$. Then, the finite difference operator*

$$\mathcal{L}_h u_h(z) = \sum_{\substack{y\in\mathbb{Z}^n \\ |y|_\infty \leq M}} a_y(x)\delta^2_{y,h}u_h(z)$$

*is of positive type with $\lambda_{0,h} = c$ and a consistent approximation of $\mathcal{L}$.*

*Proof.* Positivity can be verified by differentiating $\mathcal{L}_h u_h$ with respect to the second-order differences. The claimed consistency follows from the approximation properties of the difference quotients. □

The foregoing lemma assumes a very special structure. However, we will now show that certain diagonal dominant matrices $A$ belong to that class.

Assume we are given an orthogonal basis $(y_j)_j \subseteq \mathbb{Z}^n$ of $\mathbb{R}^n$. Then, $A$ can be expanded as follows

$$A(x) = \sum_{j,k=1}^{n} a_{jk}(x) y_j \otimes y_k \quad \text{with} \quad a_{jk} = \frac{1}{|y_j|\,|y_k|} y_j^\top A y_k. \tag{19}$$

**Lemma 3.71** (diagonally dominant case)**.** *Suppose $A$ is given in the format* (19) *for a given orthogonal basis $(y_j)_j \subseteq \mathbb{Z}^n$ of $\mathbb{R}^n$. Additionally, suppose that there is $c > 0$ such that*

$$\sum_{\substack{k=1\\ j\neq k}}^{n} |a_{jk}(x)| \leq a_{jj}(x) - c \quad \textit{for all } j.$$

*Then, the operator can be written in the format of Lemma 3.70 and, thus, there exists a consistent positive finite difference method $\mathcal{L}_h$ with $\lambda_{0,h} = c$.*

*Proof.* We use elementary algebraic manipulations and infer

$$\begin{aligned} A = &\sum_{j=1}^{n} (a_{jj} - \sum_{\substack{k=1\\ k\neq j}}^{n} |a_{jk}|) y_j \otimes y_j \\ &+ \frac{1}{4} \sum_{\substack{j,k=1\\ k\neq j}}^{n} (|a_{jk}| + a_{jk})(y_j + y_k) \otimes (y_j + y_k) + \frac{1}{4} \sum_{\substack{j,k=1\\ k\neq j}}^{n} (|a_{jk}| - a_{jk})(y_j - y_k) \otimes (y_j - y_k). \end{aligned}$$

This shows that the conditions of Lemma 3.70 are satisfied with $M \leq 2 \max_j |y_j|_\infty$. $\square$

We now show that the construction carries over to general uniformly positive definite coefficients. In order to keep the technicalities to a minimum, we confine ourselves to the case $n = 2$ of two space dimensions.

**Theorem 3.72.** *Let $n = 2$. Suppose the coefficient $A$ is uniformly positive definite with the bounds*

$$\lambda I \leq A(x) \leq \Lambda I \quad \textit{for all } x \in \Omega$$

*for constants $0 < \lambda \leq \Lambda < \infty$. Then there exists a consistent and positive finite difference operator $\mathcal{L}_h$. The stencil size can be chosen proportional to $\Lambda/\lambda$.*

*Proof.* Let $\{\lambda_1, \lambda_2\} \subseteq [\lambda, \Lambda]$ denote the eigenvalues of $A$ with a corresponding orthonormal pair of eigenvectors $\varphi_1, \varphi_2$ so that

$$A = \sum_{j=1}^{2} \lambda_j \varphi_j \otimes \varphi_j.$$

We approximate the directions of $\varphi_j$ with the directions provided by the grid. In Figure 4 it is illustrated that such a vector $y_j$ may have a large norm. It will be shown in Problem 50

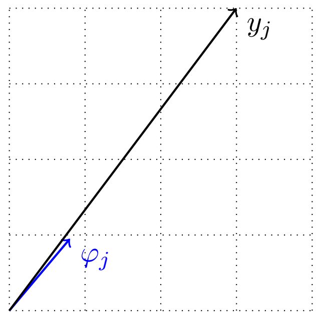


Figure 4: To achieve aligned directions of $\varphi_j$ and a vector $y_j$ matching with the grid, the stencil has to be sufficiently large.

that there is some $C > 0$ such that for any $s > 0$ (to be chosen later) there exist orthogonal $y_1, y_2 \in \mathbb{Z}^2$ with

$$\left|\varphi_j - \frac{y_j}{|y_j|}\right|_\infty \leq \frac{1}{s} \quad \text{and} \quad \frac{Cs}{2} \leq |y_j|_\infty \leq Cs. \tag{20}$$

We abbreviate $\tilde{y}_j := y_j/|y_j|$. and expand $A$ as follows

$$A = \sum_{j=1}^{2} \lambda_j \tilde{y}_j \otimes \tilde{y}_j + \sum_{j=1}^{2} \lambda_j (\varphi_j \otimes \varphi_j - \tilde{y}_j \otimes \tilde{y}_j).$$

Since the vectors $y_j \otimes y_k$ span $\mathbb{S}^{2\times 2}$, there exists a the matrix $B \in \mathbb{R}^{2\times 2}$ such that the second part can be written as

$$\sum_{j=1}^{2} \lambda_j (\varphi_j \otimes \varphi_j - \tilde{y}_j \otimes \tilde{y}_j) = \sum_{j=1}^{2} \sum_{k=1}^{2} B_{jk} \tilde{y}_j \otimes \tilde{y}_k.$$

Therefore, $A$ is of the form (19) with $a_{jk} = (\delta_{jk}\lambda_j + B_{jk})/(|y_j|\,|y_k|)$. Essentially we have shown that in the $(y_1, y_2)$ coordinate system the matrix $A$ is the sum of the diagonal matrix $\left(\begin{smallmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{smallmatrix}\right)$ and $B$. We want to establish diagonal dominance of $A$, so we need to bound the entries of $B$. To this end, let $x_1, x_2$ be a pair of normalized vectors in $\mathbb{R}^2$. We use elementary manipulations to deduce

$$\begin{aligned}
\left| x_1^\top \sum_{j=1}^{2} \lambda_j \left(\varphi_j \otimes \varphi_j - \tilde{y}_j \otimes \tilde{y}_j\right) x_2 \right| &= \left| x_1^\top \sum_j \lambda_j (\varphi_j \otimes (\varphi_j - \tilde{y}_j) + (\varphi_j - \tilde{y}_j) \otimes \tilde{y}_j) x_2 \right| \\
&= \left| \sum_j \lambda_j \Big( x_1 \cdot \varphi_j (\varphi_j - \tilde{y}_j) \cdot x_2 + x_1 \cdot (\varphi_j - \tilde{y}_j) \tilde{y}_j \cdot x_2 \Big) \right| \dots
\end{aligned}$$

We then use the spectral bound $\lambda_j \leq \Lambda$ and the bounds $|x_1 \cdot \varphi_j| \leq 1$ and $|\tilde{y}_j \cdot x_2| \leq 1$ which we obtain from the normalization of these vectors and Cauchy's inequality and bound the right-hand side of the foregoing expression by

$$\cdots \leq \Lambda \sum_j (|(\varphi_j - \tilde{y}_j)|_2 + |\varphi_j - \tilde{y}_j)|_2) = 4|\varphi_j - \tilde{y}_j|_2$$

The equivalence of vector norms in $\mathbb{R}^2$ reads as $|\cdot|_2 \leq \sqrt{2}|\cdot|_\infty$ (see Problem 49) and we therefore altogether obtain with the above bound (20) on $\varphi_j - \tilde{y}_j$ that

$$\left| x_1^\top \left( \sum_{j=1}^2 \lambda_j (\varphi_j \otimes \varphi_j - \tilde{y}_j \otimes \tilde{y}_j) \right) x_2 \right| \leq 4\sqrt{2}\Lambda |\varphi_j - \tilde{y}_j|_\infty \leq \frac{4\sqrt{2}\Lambda}{s}.$$

We therefore have established

$$|B_{jk}| \leq 4\sqrt{2}\Lambda/s.$$

The coefficients $a_{jk}$ therefore satisfy the following bounds

$$a_{jj} = \frac{\lambda_j + B_{jj}}{|y_j|^2} \geq \frac{\lambda - 4\sqrt{2}\Lambda/s}{|y_j|^2}$$

$$a_{jk} = \frac{B_{jk}}{|y_j|\,|y_k|} \leq \frac{4\sqrt{2}\Lambda}{s|y_j|\,|y_k|} \quad \text{if } j \neq k.$$

We now adjust the parameter $s$ such that the diagonal dominant structure from Lemma 3.71 is satisfied. From the above estimates we see for

$$\sum_{\substack{k=1\\ j\neq k}}^2 |a_{jk}| \leq a_{jj} - c \quad \text{for all } j$$

to hold, it is sufficient that

$$\sum_{\substack{k=1\\ j\neq k}}^2 \frac{4\sqrt{2}\Lambda/s}{|y_j|\,|y_k|} \leq \frac{\lambda}{|y_j|^2} - c$$

holds for all $j$ and some $c > 0$, or, equivalently,

$$\frac{\Lambda}{\lambda} 4\sqrt{2} \sum_{k=1}^2 \frac{|y_j|}{|y_k|} \leq s - \frac{|y_j|^2 cs}{\lambda}$$

In order to achieve this estimate, we choose $s := 64\Lambda/\lambda$ and $c := \lambda/(4(Cs)^2)$ . Noting that the ratio $|y_j|/|y_k|$ is bounded from above by $2\sqrt{2}$ (use Problem 49 and (20)), we then have

$$\frac{\Lambda}{\lambda} 4\sqrt{2} \sum_{k=1}^2 \frac{|y_j|}{|y_k|} \leq \frac{\Lambda}{\lambda} 4\sqrt{2} \cdot 2 \cdot 2\sqrt{2} = 32\Lambda/\lambda = s - s/2.$$

We further have from the definition of $c$ and (20)

$$-s/2 \leq -\frac{cs}{2} \frac{4(Cs)^2}{\lambda} \leq -\frac{cs}{2} \frac{4|y_j|^2}{\lambda} = -2cs|y_j|^2/\lambda$$

so that in summary the desired bound is achieved. We have thus shown that $A$ has the diagonally dominant structure from Lemma 3.71. From (20) we see that the stencil size is not larger than $Cs$. $\square$

**Synopsis of §12.**

Starting from the diagonal dominant case we have shown that it there exist positive and consistent FDM discretizations of the model problem. The proofs are constructive and provide details for actual numerical methods.

## Problems to §12

**Problem 49.** Show that $|x|_2 \leq \sqrt{n}|x|_\infty$ for any $x \in \mathbb{R}^n$.

**Problem 50.** Let $(\varphi_1, \varphi_2)$ be an orthonormal basis of $\mathbb{R}^2$. Show that there exists a universal constant $C > 0$ such that, given any $s > 0$, there exist orthogonal vectors $y_1, y_2 \in \mathbb{Z}^2$ such that

$$\left|\varphi_j - \frac{y_j}{|y_j|}\right|_\infty \leq \frac{1}{s} \quad \text{and} \quad \frac{Cs}{2} \leq |y_j|_\infty \leq Cs.$$

**Problem 51.** Consider the unit square $\Omega = (0,1)^2$. Derive and implement a FDM of stencil size $m = 2$ for the equation

$$A : D^2 u = f \text{ in } \Omega \quad \text{and} \quad u = 0 \text{ on } \partial\Omega$$

with

$$A = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \quad \text{for } \phi = \pi/6.$$

under homogeneous Dirichlet boundary conditions. Close to the boundary, use modified stencils as in the subsequent figure.

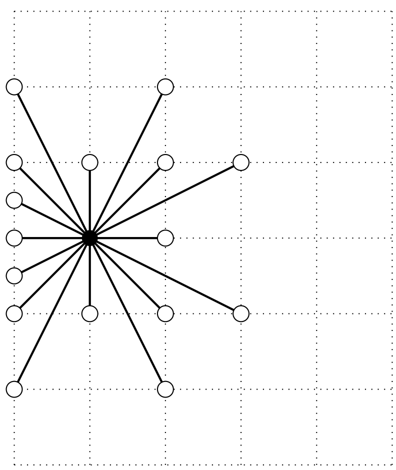

Figure 5: Modified stencil near the boundary.

## §13 Discrete Alexandrov estimate *(week 5)*

We will define a discrete notion of convexity and prove a fundamental estimate bounding the negative part of a grid function. In what follows, we use the disjoint splitting

$$\bar{\Omega}_h = \Omega_h^B \cup \Omega_h^I.$$

Here, $\Omega_h^I$ are the points $z \in \Omega_h$ such that $F_h[v_h](z)$ only depends on $v_h[y]$ for points $y \in \bar{\Omega}_h$. Loosely speaking, $\Omega_I$ are the interior gridpoints that are at least a stencil-width distant from the boundary. We then set $\Omega_h^B := \bar{\Omega}_h \setminus \Omega_h^I$.

**Definition 3.73.** The function $v_h \in X_h$ is a *convex nodal function* if at any interior $z \in \Omega_h^I$ there exists a supporting hyperplane, that is a vector $p \in \mathbb{R}^n$ such that

$$v_h(y) \geq v_h(z) + p \cdot (y - z) \quad \text{for all } y \in \bar{\Omega}_h.$$

We collect all supporting hyperplanes at $z$ in a set called the discrete subdifferential.

**Definition 3.74.** Given $v_h \in X_h$ and $z \in \bar{\Omega}_h$, we define by

$$\partial_h v_h(z) := \{p \in \mathbb{R}^n : \forall x \in \bar{\Omega}_h \ v_h(x) \geq v_h(z) + p \cdot (x - z)\}$$

the *discrete subdifferential.*

Assume we are given $R > 0$ such that $\bar{\Omega} \subseteq B_R(0)$. We use the following convention on the negative part $v_h^- := \max\{0, -v_h\}$ of $v_h$. Given any nodal function $v_h$ with $v_h \geq 0$ on $\Omega_h^B$, we extend the negative part or $v_h^-$ by 0 to the grid points

$$B_{R,h} := B_R \cap \{he : e \in \mathbb{Z}^n\}$$

in $B_R(\Omega) \setminus \Omega_h$.

**Definition 3.75.** Given $v_h \in X_h$, we define the *discrete convex envelope* of $-v_h^-$ as

$$\Gamma_h(v_h)(x) := \sup\{L(x) : L \text{ affine and } L(z) \leq -v_h^-(z) \text{ for all } z \in B_{R,h}\}$$

We note $\Gamma_h(v_h)(z) \leq v_h(z)$ in all nodes $z$.

**Definition 3.76.** Let $v_h \in X_h$ with $v_h \geq 0$ on $\Omega_h^B$. The *lower nodal contact set* of $v_h$ is defined as

$$\mathcal{C}_h^-(v_h) := \{z \in \Omega_h^I : \Gamma_h(v_h)(z) = v_h(z)\}.$$

**Lemma 3.77** (finite difference Alexandrov estimate)**.** *Let $v_h \in X_h$ satisfy $v_h \geq 0$ on $\Omega_h^B$. Then there exists a constant $C > 0$, only depending on $n$, such that*

$$\sup_{\bar{\Omega}_h} v_h^- \leq CR \left( \sum_{z \in \mathcal{C}_h^-(v_h)} \mathcal{L}^n(\partial_h \Gamma_h(v_h)(z)) \right)^{1/n}$$

*where $\mathcal{L}^n$ denotes the $n$-dimensional Lebesgue measure.*

*Proof.* Let us first reduce the statement to a statement on the discrete convex envelope $\Gamma_h(v_h)$. If $z_* \in B_{R,h}$ is a grid point where the supremum is attained, i.e., $\sup_{B_{R,h}} v_h^- = v_h^-(z_*)$, there exists a horizontal plane touching $v_h$ at $z_*$ from below. In other words, there is a constant function $L \equiv v_h^-$ with $v_h^- \geq L$ in all grid points $z$. By definition, the discrete convex envelope therefore satisfies at any $z \in B_{R,h}$ that

$$\Gamma_h(v_h)(z) \geq L(z) = L(z_*) = v_h(z_*),$$

which implies the converse estimate $\sup_{B_{R,h}} \Gamma_h(v_h)^- \leq v_h^-(z_*)$ for the negative part. Trivially we have from the definition of $\Gamma_h$ that $\Gamma_h(v_h) \leq v_h$ and in particular $\sup v_h^- \leq \sup \Gamma_h(v_h)^-$. Thus

$$\sup_{\bar{\Omega}} v_h^- = \sup_{B_{R,h}} v_h^- = \sup_{B_{R,h}} \Gamma_h(v_h)^- = \sup_{B_R} \Gamma_h(v_h)^-$$

(recall that $\Gamma_h(v_h)$ is a function defined not only at the grid points). It therefore suffices to prove

$$\sup_{B_R} \Gamma_h(v_h)^- \leq CR \left( \sum_{z \in \mathcal{C}_h^-(v_h)} \mathcal{L}^n(\partial_h \Gamma_h(v_h)(z)) \right)^{1/n}. \tag{21}$$

Let as above $z_* \in B_{R,h}$ be a grid point with $\sup_{B_{R,h}} v_h^- = v_h^-(z_*)$. Since $v_h \geq 0$ on $\Omega_h^B$, we can choose $z_* \in \Omega_h^I$ to be an interior point. We define $M := \sup_{B_R} \Gamma_h(v_h)^-$ and define a cone $K(x)$ with vertex $z_*$ (defined above) by the relations

$$K(z_*) = -M \quad \text{and} \quad K|_{\partial_h B_R} = 0.$$

If $p \in B_{M/(2R)}(0)$, we have that $L(x) := -M + p \cdot (x - z_*) \leq K(x)$ for any $x \in B_R$. In other words, $L$ is a supporting plane of $K$ at $z_*$ and thus $p \in \partial_h K(z_*)$. Therefore $B_{M/(2R)} \subseteq \partial_h K(z_*)$ whence

$$\frac{\alpha_n}{2^n}(M/R)^n = \mathcal{L}^n(B_{M/(2R)}) \leq \mathcal{L}^n(\partial_h K(z_*)).$$

In the remaining part of the proof we will show the relation

$$\partial_h K(z_*) \subseteq \bigcup_{z \in \mathcal{C}_h^-} \partial_h \Gamma_h(v_h)(z). \tag{22}$$

Once this has been shown, we can deduce with the previous estimate the inequality

$$\frac{\alpha_n}{2^n}(M/R)^n \leq \mathcal{L}^n(\partial_h K(z_*)) \leq \mathcal{L}^n \left( \bigcup_{z \in \mathcal{C}_h^-} \partial_h \Gamma_h(v_h)(z) \right) \leq \textstyle\sum_{z \in \mathcal{C}_h^-} \mathcal{L}^n(\partial \Gamma_h(v_h)(z))$$

which implies (21) and therefore proves the assertion.

To prove (22), we must show that any supporting plane $L$ of $K$ at $z_*$ can be shifted (by adding a constant) to a supporting plane $\tilde{L}$ of $\Gamma_h(v_h)$ at some $y \in \mathcal{C}_h^-(v_h)$. Since, by assumption, $v_h \geq 0$ on $\Omega_h^B$ and $K(z_*) = L(z_*) = v_h(z_*)$, the nodal function $v_h - L$ satisfies

$$v_h(z_*) - L(z_*) = K(z_*) - L(z_*) = 0 \quad \text{and} \quad v_h - L \geq K - L \geq 0 \text{ on } \Omega_h^B$$

and, hence, $v_h - L$ attains a non-positive minimum at some $x \in \Omega_h^I$ (recall that $z_* \in \Omega_h^I$). Thus, $\tilde{L}(z) := L(z) + v_h(x) - L(x)$ satisfies $\tilde{L} \leq v_h$ on $B_{R,h}$ and $\tilde{L}(x) = v_h(x)$. Comparing with the definition of $\Gamma_h(v_h)$ we obtain $\tilde{L} \leq \Gamma_h(v_h) \leq v_h$, which shows that $\tilde{L}$ is a supporting hyperplane of $\Gamma_h(v_h)$ at $x$ and that $x$ is a contact node, $x \in C_h^-(v_h)$. This proves (22). $\square$

## Synopsis of §13.

We have formulated basic notions of discrete/nodal convexity and proved the discrete Alexandrov estimate for grid functions.

# Problems to §13

**Problem 52.** Prove that the nodal interpolant of a convex function is a convex nodal function.

**Problem 53.** Let $v_h \in X_h$ be a convex nodal function with $v_h \leq 0$. Prove that $v_h(z) = \Gamma_h(v_h)(z)$ for all $z \in \Omega_h^I$.

**Problem 54.** Let $v_h \in X_h$. Prove that $\Gamma_h(v_h) = 0$ on $\partial B_R$.

**Problem 55.** Let $w_h$ and $v_h$ with $w_h \leq v_h$ be convex nodal functions with $w_h(z_*) = v_h(z_*)$ at some $z_* \in \bar{\Omega}_h$. Prove $\partial_h w_h(z_*) \subseteq \partial_h v_h(z_*)$.

**Problem 56.** Let $w_h$ and $v_h$ be convex nodal functions. Prove that $\partial w_h(z) + \partial v_h(z) \subseteq \partial(v_h + w_h)(z)$ for all $z \in \Omega_h^I$. Here we use the notation $A + B = \{a + b : a \in A, b \in B\}$.

## §14 Finite difference ABP estimate and stability analysis *(week 6)*

Recall the discrete ellipticity constant $\lambda_{0,h}$.

The following theorem is known as discrete Alexandrov–Bakelman–Pucci (ABP) estimate.

**Theorem 3.78** (discrete ABP estimate)**.** *Let $\mathcal{L}_h$ be a finite difference operator of the format* (18) *and of positive type. Let $g_h \in X_h$ be given and let $u_h \in X_h$ be a grid function satisfying*

$$\mathcal{L}_h u_h \leq f \text{ in } \Omega_h^I \quad \text{and} \quad u_h = g_h \text{ on } \Omega_h^B.$$

*Then*

$$\sup_{\bar{\Omega}_h} u_h^- \leq \sup_{\Omega_h^B} g_h^- + C\frac{MR}{\lambda_{0,h}}\left(\sum_{z\in\mathcal{C}_h^-(u_h)} h^n (f^+(z))^n\right)^{1/n}$$

*for the stencil size $M$ and a constant $C = C(n)$ only depending on the space dimension $n$.*

*Proof.* Without loss of generality, we may assume that $u_h \geq 0$ on $\Omega_h^B$, since otherwise we can consider $u_h + \max_{\Omega_h^B} g_h^-$. We will furthermore restrict our attention to the interesting case of $\max_{\bar{\Omega}_h} u_h^- > 0$, since otherwise nothing needs to be shown.

Let $z \in \mathcal{C}_h^-(u_h)$ be a point in the contact set and let $y \in S$ be a vector from the stencil. Since $\Gamma_h(u_h)$ is convex, the second-order difference with respect to $y$ satisfies $\delta^2_{y,h}\Gamma_h(u_h) \geq 0$. From the positivity of $\mathcal{L}_h$ (and hence nonnegative coefficients $a_y(z)$), the contact property $\Gamma_h(u_h)(z) = u_h(z)$, and $\Gamma_h(u_h) \leq u_h$, we thus obtain

$$\begin{aligned} 0 &\leq a_y(z)\delta^2_{y,h}\Gamma_h(u_h) \\ &= a_y(z)\frac{\Gamma_h(u_h)(z+hy) - 2\Gamma_h(u_h)(z) + \Gamma_h(u_h)(y-hy)}{h^2} \leq a_y(z)\delta^2_{y,h}u_h(z). \end{aligned}$$

Taking the sum over the stencil and using $\mathcal{L}_h u_h \leq f$ leads to

$$0 \leq a_y(z)\delta^2_{y,h}\Gamma_h(u_h) \leq \sum_{y'\in S} a_{y'}(z)\delta^2_{y',h}\Gamma_h(u_h) \leq f(z) \leq f^+(z).$$

Let now $\{y_j\}_{j=1}^n \subseteq S$ be the orthogonal set from Definition 3.68 (positivity). We expand $\delta^2_{y,h}\Gamma_h(u_h)$ in the above inequality and use $a_{y_j} \geq \lambda_{0,h}$ (see Definition 3.68) and obtain after elementary manipulations

$$\lambda_{0,h}\frac{\delta^+_{y_j,h}\Gamma_h(u_h)(z) - \delta^-_{y_j,h}\Gamma_h(u_h)(z)}{h} = \lambda_{0,h}\delta^2_{y_j,h}\Gamma_h(u_h)(z) \leq a_{y_j}(z)\delta^2_{y_j,h}\Gamma_h(u_h)(z) \leq f^+(z)$$

so that

$$\delta^+_{y_j,h}\Gamma_h(u_h)(z) \leq \delta^-_{y_j,h}\Gamma_h(u_h)(z) + \frac{h}{\lambda_{0,h}}f^+(z).$$

Let $p \in \partial_h \Gamma_h u_h(z)$ be any element in the discrete subdifferential. From the definition we have $\Gamma_h(u_h)(z \pm hy_j) \geq \Gamma_h(u_h)(z) \pm hp \cdot y_j$. This implies

$$\delta^-_{y_j,h}\Gamma_h(u_h)(z) \leq p \cdot y_j \leq \delta^+_{y_j,h}\Gamma_h(u_h)(z).$$

The combination with the foregoing estimate shows

$$\delta^-_{y_j,h}\Gamma_h(u_h)(z) \leq p \cdot y_j \leq \delta^+_{y_j,h}\Gamma_h(u_h)(z) \leq \delta^-_{y_j,h}\Gamma_h(u_h)(z) + \frac{h}{\lambda_{0,h}} f^+(z)$$

and thus, with $k := \delta^-_{y_j,h}\Gamma_h(u_h)(z)$,

$$k \leq p \cdot y_j \leq k + \frac{h}{\lambda_{0,h}} f^+(z) \quad \text{for any } p \in \partial_h \Gamma_h(u_h)(z) \text{ and any of the vectors } y_j.$$

Since the $\tilde{y}_j := y_j/|y_j|$ form an orthonormal basis of $\mathbb{R}^n$, we obtain with the bound $|y_j| \leq \sqrt{n}M$ for the stencil size $M$ that $p \cdot \tilde{y}_j \leq \sqrt{n}Mhf^+(z)/\lambda_{0,h}$. This means that any $p \in \partial_h\Gamma_h(u_h)(z)$ is contained in a box of side length $\sqrt{n}Mhf^+(z)/\lambda_{0,h}$, which yields the following bound on the Lebesgue measure

$$\mathcal{L}^n(\partial_h \Gamma_h(u_h)(z)) \leq \sqrt{n}^n M^n \frac{h^n f^+(z)^n}{\lambda^n_{0,h}}.$$

Summation over all contact points yields

$$\sum_{z \in \mathcal{C}^-_h(u_h)} \mathcal{L}^n(\partial_h \Gamma_h(u_h)(z)) \leq \sqrt{n}^n M^n \sum_{z \in \mathcal{C}^-_h(u_h)} \frac{h^n f^+(z)^n}{\lambda^n_{0,h}}.$$

We combine this estimate with Lemma 3.77 and obtain the assertion of the theorem. □

As a consequence of Theorem 3.78, any $u_h$ satisfying

$$\mathcal{L}_h u_h \leq f \text{ in } \Omega^I_h \quad \text{and} \quad u_h = g_h \text{ on } \Omega^B_h. \tag{23}$$

satisfies the stability estimate

$$\max_{\bar{\Omega}_h} |u_h| \leq \max_{\Omega^B_h} |g_h| + C \frac{R}{\lambda_{0,h}} \left( \sum_{z \in \mathcal{C}^-_h(u_h)} h^n |f(z)|^n \right)^{1/n}.$$

Since for finite-dimensional linear problems uniqueness of solutions implies existence, we deduce that there exists a unique solution to the finite difference system (23).

## Synopsis of §14.

We have proven the discrete ABP estimate and deduced stability of the FDM.

# Literature

## Principal references

Most of the material in this lecture is taken from these three references:


- Nikos Katzourakis. *An introduction to viscosity solutions for fully nonlinear PDE with applications to calculus of variations in $L^\infty$*. SpringerBriefs in Mathematics. Springer, Cham, 2015. free preprint copy from `https://arxiv.org/pdf/1411.2567`

- Shigeaki Koike. *A beginner's guide to the theory of viscosity solutions*, volume 13 of *MSJ Memoirs*. Mathematical Society of Japan, Tokyo, 2004. `http://www.math.tohoku.ac.jp/~koike/evis2013version.pdf`

- Michael Neilan, Abner J. Salgado, and Wujun Zhang. Numerical analysis of strongly nonlinear PDEs. *Acta Numer.*, 26:137–303, 2017


## Further reading

The following references contain further information on elliptic PDEs, viscosity solutions, or finite differences:


- Sören Bartels. *Numerical approximation of partial differential equations.*, volume 64 of *Texts Appl. Math.* Springer, Cham, 2016. Online im Netz der Uni Jena verfügbar

- Xiaobing Feng, Roland Glowinski, and Michael Neilan. Recent developments in numerical methods for fully nonlinear second order partial differential equations. *SIAM Rev.*, 55(2):205–267, 2013

- Stig Larsson and Vidar Thomée. *Partial differential equations with numerical methods*, volume 45. Berlin: Springer, 2003

- Michael G. Crandall, Hitoshi Ishii, and Pierre-Louis Lions. User's guide to viscosity solutions of second order partial differential equations. *Bull. Amer. Math. Soc. (N.S.)*, 27(1):1–67, 1992

- David Gilbarg and Neil S. Trudinger. *Elliptic partial differential equations of second order.* Classics in Mathematics. Springer-Verlag, Berlin, 2001. Reprint of the 1998 edition