Ingenieurmathematik

Skript zur Vorlesung im Master-Studiengang Geodäsie

Dr. Martin Lenz, Dr. Patrick Penzler

Version vom 8. Juni 2012

Inhaltsverzeichnis

Zur besseren Übersicht werden in diesem Skript folgende Farben verwendet:

Satz, Lemma, Folgerung	
Definition, Notation	
Schema	
Problemstellung	
Programmieraufgabe	

0 Einleitung

Thema: Numerische Lösungsverfahren für partielle Differentialgleichungen
Modellproblem:

$$\frac{\partial^2}{\partial x^2}u(x,y) + \frac{\partial^2}{\partial y^2}u(x,y) = f(x,y)$$

Diese sogenannte „Poisson-Gleichung“ tritt beispielsweise bei der Datenglättung oder der Modellierung eines Schwerefelds auf.

Notation 0.1. *Wir schreiben*

$$\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} =: \partial_{xx} + \partial_{yy} =: \Delta.$$

Diskretisierungsmethoden: Folgende Diskretisierungsmethoden werden in der Vorlesung behandelt:

- Finite Differenzen
- Finite Elemente
- Randelemente

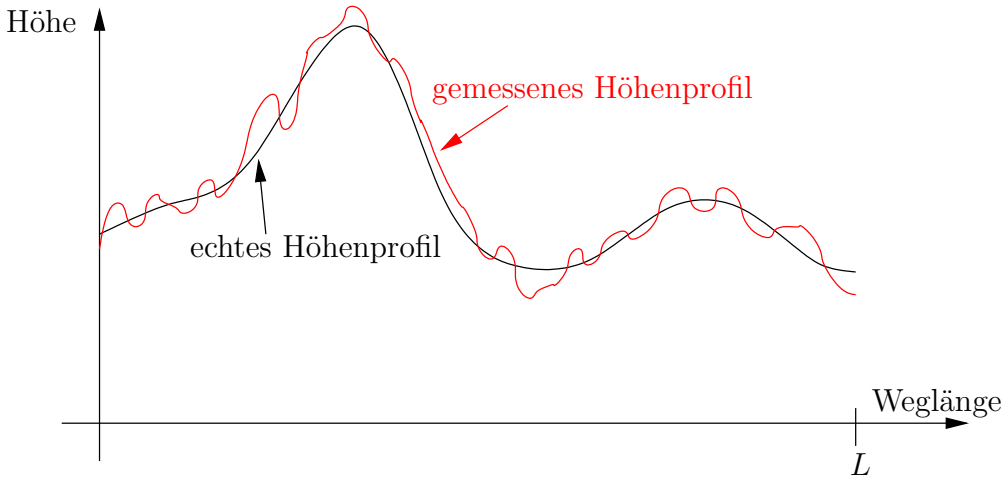
Algorithmische Umsetzung: `MATLAB`
Literatur:

- Dietrich Braess: *„Finite Elemente“* (Kapitel I & II), Springer.
- Wolfgang Dahmen und Arnold Reusken: *„Numerik für Ingenieure und Naturwissenschaftler“* (Kapitel 12), Springer.
- Lothar Gaul, Martin Kögl und Marcus Wagner: *„Boundary Element Methods for Engineers and Scientists“* (Kapitel 4), Springer.

1 Finite Differenzen

1.1 Finite Differenzen in 1D

Modellproblem: Datenglättung für Funktionen $f : \mathbb{R} \to \mathbb{R}$
Als Beispiel wollen wir gemessene Höhendaten entlang eines Weges betrachten. Die Höhenwerte mögen dabei starkes Rauschen (z.B. durch schlechten GPS-Empfang) enthalten.
Wie erhält man daraus ein „vernünftiges“ Höhenprofil?



Einfacher Ansatz: Interpretiere die Höhendaten als *kontinuierliche* Funktion

$$f : [0, L] \to \mathbb{R}.$$

Gesucht ist nun eine Funktion

$$u : [0, L] \to \mathbb{R},$$

die die wesentlichen Eigenschaften von f widerspiegelt, aber weniger Rauschen enthält.
Ansatz: Was wollen wir nicht?

- Die Funktion u soll nicht weit weg von f sein, d.h. $|u(x) - f(x)|$ soll klein sein.
- u soll nicht verrauscht sein, d.h. $|u'(x)|$ soll klein sein.

Dies führt uns auf das folgende Funktional, das wir minimieren wollen:

$$E[u] := \int_0^L (u(x) - f(x))^2 + \beta(u'(x))^2 \, dx, \quad \beta > 0 \text{ konstant.}$$

Erinnerung: Hat eine differenzierbare Funktion ein Minimum, so ist die Ableitung dort 0.
Wie sieht nun die Ableitung von E nach u aus? Wir betrachten hierzu die Richtungsableitung $\frac{d}{dt}E[u+tv]\Big|_{t=0}$ von E in Richtung einer differenzierbaren Funktion $v : [0, L] \to \mathbb{R}$ für die $v(0) = v(L) = 0$ gilt. Dann gilt:

$$
\begin{aligned}
0 &= \frac{d}{dt}E[u+tv]\Big|_{t=0} \\
&= \frac{d}{dt}\int_0^L (u(x)+tv(x)-f(x))^2 + \beta(u'(x)+tv'(x))^2\,dx\Big|_{t=0} \\
&= \int_0^L \frac{d}{dt}(u(x)+tv(x)-f(x))^2 + \beta\frac{d}{dt}(u'(x)+tv'(x))^2\,dx\Big|_{t=0} \\
&\overset{\text{Kettenregel}}{=} \int_0^L 2\,(u(x)+tv(x)-f(x))\,v(x) + 2\beta(u'(x)+tv'(x))v'(x)\,dx\Big|_{t=0} \\
&= \int_0^L 2\,(u(x)-f(x))\,v(x) + 2\beta u'(x)v'(x)\,dx \\
&\overset{\text{part. Int.}}{=} \int_0^L 2\,(u(x)-f(x))\,v(x)\,dx - \int_0^L 2\beta u''(x)v(x)\,dx + u'(x)v(x)\Big|_{x=0}^{x=L} \\
&= 2\int_0^L \Big(u(x)-f(x)-\beta u''(x)\Big)v(x)\,dx
\end{aligned}
$$

Die Richtungsableitung muss im Minimum für alle Richtungen 0 ergeben, d.h. es muss gelten:

$$2\int_0^L \Big(u(x)-f(x)-\beta u''(x)\Big)v(x)\,dx = 0$$

für alle differenzierbaren $v : [0, L] \to \mathbb{R}$ mit $v(0) = v(L) = 0$. Daraus ergibt sich

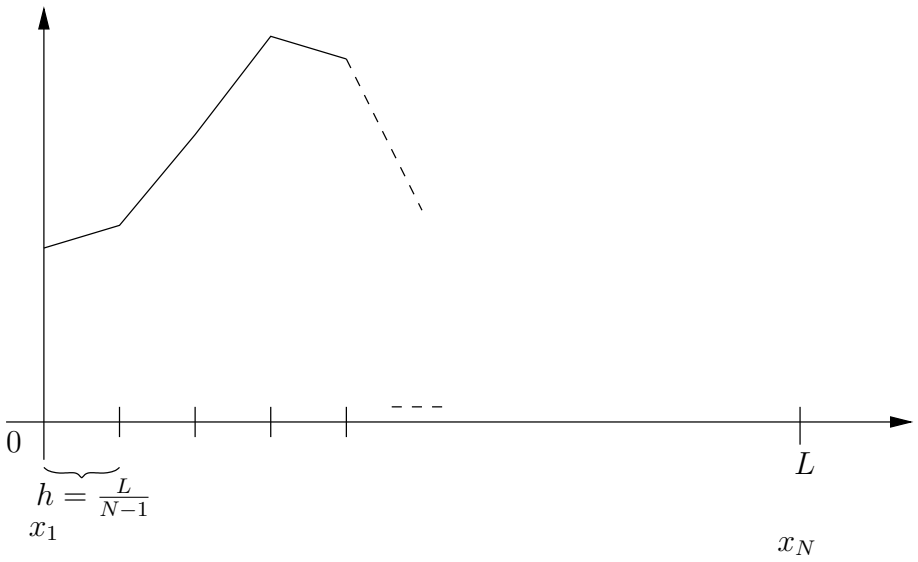
$$u(x) - f(x) - \beta u''(x) = 0 \quad \text{für alle } x \in (0, L).$$

Was passiert am Rand?
Einfacher Fall: Wir kennen die Höhen am Start- und Zielpunkt sehr genau. Dann können wir diese einfach vorschreiben: $u(0) = f(0)$, $u(L) = f(L)$. Insgesamt erhalten wir:

Problemstellung 1.1. *Gegeben sei eine stetige Funktion $f : [0, L] \to \mathbb{R}$ und $\beta > 0$. Wir suchen eine zweimal stetig differenzierbare Funktion $u : [0, L] \to \mathbb{R}$ mit*

$$
\begin{aligned}
-\beta u''(x) + u(x) &= f(x), \quad x \in (0, L) \\
u(0) &= f(0), \\
u(L) &= f(L).
\end{aligned}
\tag{P}
$$

Wie findet man u näherungsweise?
Wir approximieren u auf einem Gitter der Schrittweite $h = \frac{L}{N-1}$. Die Funktion $u : [0, L] \to \mathbb{R}$ wird dann durch Werte an den Knoten $x_i = (i-1) \cdot h$ approximiert. $U \in \mathbb{R}^N$ ist der Vektor mit den Knotenwerten, $U_i \approx u(x_i)$.
Zu einer gegebenen Funktion u kann man natürlich stets einen Vektor U mit $U_i = u(x_i)$ finden. Wenn jedoch u die (unbekannte) Lösung von (P) bezeichnet, so werden wir in der Regel nur eine Approximation der Knotenwerte von u berechnen können, d. h. $U_i \approx u(x_i)$.



Approximation der Ableitung? Hierbei dürfen nur Werte an den Knoten verwendet werden! Idee: Differenzenquotienten
Nach dem **Satz von Taylor** gelten für eine viermal stetig differenzierbare Funktion u die Identitäten

$$\begin{aligned} u(x+h) &= u(x) + u'(x)h + \frac{1}{2}u''(x)h^2 + \frac{1}{6}u'''(x)h^3 + O(h^4), \\ u(x-h) &= u(x) - u'(x)h + \frac{1}{2}u''(x)h^2 - \frac{1}{6}u'''(x)h^3 + O(h^4). \end{aligned}$$

Addition ergibt $u(x+h) + u(x-h) = 2u(x) + u''(x)h^2 + O(h^4)$, d.h. es ergibt sich der folgende Differenzenquotient für die zweite Ableitung

$$\begin{aligned} u''(x) &= \frac{1}{h^2}\Big(u(x-h) - 2u(x) + u(x+h)\Big) + O(h^2) \\ \Rightarrow u''(x_i) &= \frac{1}{h^2}\Big(u(x_{i-1}) - 2u(x_i) + u(x_{i+1})\Big) + O(h^2) \end{aligned}$$

Die Differentialgleichung (P) wird also wie folgt approximiert.

Bemerkung 1.2. *Bezeichnen wir die erste Ableitung mit*

$$w(x) := \frac{u(x+h) - u(x)}{h},$$

so ist die Diskretisierung der zweiten Ableitung durch

$$\frac{w(x) - w(x-h)}{h}$$

gegeben. Wir haben also erst Vorwärtsdifferenzen *und dann* Rückwärtsdifferenzen *verwendet. Dies hängt damit zusammen, dass wir in höheren Dimensionen* $\Delta = \mathrm{div}\nabla$*, also die Divergenz des Gradienten, haben. Dabei werden Divergenz und Gradient unterschiedlich diskretisiert, siehe auch Bemerkung 1.18.*

Problemstellung 1.3. *Gegeben sei eine stetige Funktion* $f : [0, L] \to \mathbb{R}$ *und* $\beta > 0$*. Sei* $N \in \mathbb{N}$ *und* $h = L/(N-1)$*. Wir suchen* $U \in \mathbb{R}^N$ *mit*

$$\begin{aligned} -\frac{\beta}{h^2}(U_{i-1} - 2U_i + U_{i+1}) + U_i &= f(x_i), \quad i = 2, ..., N-1, \\ U_1 &= f(x_1), \\ U_N &= f(x_N). \end{aligned} \tag{P_h}$$

Dies ist ein lineares Gleichungssystem für $U \in \mathbb{R}^N$. In Matrixschreibweise ergibt sich die folgende Darstellung.

Schema 1.4. *(Matrixdarstellung von* (P_h)*)*

$$\left(-\frac{\beta}{h^2} \begin{pmatrix} 0 & & & & 0 \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \\ 0 & & & & 0 \end{pmatrix} + \begin{pmatrix} 1 & & & & 0 \\ & 1 & & & \\ & & \ddots & & \\ & & & 1 & \\ 0 & & & & 1 \end{pmatrix} \right) \begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_{N-1} \\ U_N \end{pmatrix} = \begin{pmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_{N-1}) \\ f(x_N) \end{pmatrix}.$$

Der zugehörige `MATLAB`-Code sieht folgendermaßen aus:

Schema 1.5. *(`MATLAB`-Code zu* (P_h)*)*

```
e = ones (N, 1);
% Matrix for second derivative
A = spdiags ([e -2*e e], [-1 0 1], N, N);
% First and last line
A (1, :) = zeros (1, N);
A (N, :) = zeros (1, N);
% Identity matrix
B = spdiags (e, 0, N, N);
% Complete matrix
L = - beta / (h*h) * A + B;
```

Programmieraufgabe 1. *Implementieren Sie das beschriebene Finite-Differenzen-Verfahren zur Datenglättung in* `MATLAB`*. Experimentieren Sie mit dem Parameter* β*.*

1.2 Konvergenztheorie für Differenzenverfahren

(P_h) führt auf ein lineares Gleichungssystem. Wir wenden uns nun den folgenden Fragen zu:

1. Ist dieses stets *eindeutig lösbar*?
2. Konvergiert die Lösung von (P_h) für $N \to \infty$ (bzw. $h \to 0$) gegen die Lösung von (P)? Wenn ja, wie schnell?

Dazu benötigen wir zunächst das folgende Lemma, das sowohl für die kontinuierliche Lösung von (P) als auch für die diskrete Lösung von (P_h) gilt. Um die Analogie herauszuarbeiten, beweisen wir beide Varianten.

Lemma 1.6 (Maximumprinzip)**.**

a) Sei u Lösung von (P). Dann ist

$$\begin{aligned} \max_{x\in[0,L]} u(x) &\leq \max_{x\in[0,L]} f(x), \\ \min_{x\in[0,L]} u(x) &\geq \min_{x\in[0,L]} f(x). \end{aligned}$$

b) Sei U Lösung von (P_h). Dann ist

$$\begin{aligned} \max_{i=1,\dots,N} U_i &\leq \max_{i=1,\dots,N} f(x_i), \\ \min_{i=1,\dots,N} U_i &\geq \min_{i=1,\dots,N} f(x_i). \end{aligned}$$

Da u und f stetig sind, können wir in a) von Maxima (und nicht nur von Suprema) reden.

Beweis.

a) Sei $\hat{x} \in [0, L]$ Maximalstelle von u. Falls $\hat{x} = 0$ oder $\hat{x} = L$, so ist

$$\max_{x\in[0,L]} u(x) = u(\hat{x}) = f(\hat{x}) \leq \max_{x\in[0,L]} f(x).$$

Falls $\hat{x} \in (0, L)$, so folgt ($\hat{x}$ ist Maximalstelle, u ist zweimal stetig differenzierbar)

$$\begin{aligned} u''(\hat{x}) \leq 0 &\Rightarrow -\beta u''(\hat{x}) \geq 0 \\ &\Rightarrow f(\hat{x}) - u(\hat{x}) \geq 0 \\ &\Rightarrow u(\hat{x}) \leq f(\hat{x}), \end{aligned}$$

wobei wie oben $f(\hat{x}) \leq \max_{x\in[0,L]} f(x)$. Für das Minimum betrachte $\tilde{u} = -u, \tilde{f} = -f$.

b) Sei $\hat{\imath} \in \{1, \ldots, N\}$ Maximalstelle von U. Falls $\hat{\imath} = 1$ oder $\hat{\imath} = N$, so ist

$$\max_{i=1,\ldots,N} U_i = U_{\hat{\imath}} = f(x_{\hat{\imath}}) \leq \max_{i=1,\ldots,N} f(x_i).$$

Falls $\hat{\imath} \in \{2, \ldots, N-1\}$, so gilt ($\hat{\imath}$ ist Maximalstelle)

$$\begin{aligned} U_{\hat{\imath}-1} &\leq U_{\hat{\imath}}, \\ U_{\hat{\imath}+1} &\leq U_{\hat{\imath}}. \end{aligned}$$

Daraus folgt

$$\begin{aligned} U_{\hat{\imath}-1} - 2U_{\hat{\imath}} + U_{\hat{\imath}+1} \leq 0 \quad &\Rightarrow \quad -\frac{\beta}{n^2}(U_{\hat{\imath}-1} - 2U_{\hat{\imath}} + U_{\hat{\imath}+1}) \geq 0 \\ &\Rightarrow \quad f(x_{\hat{\imath}}) - u_{\hat{\imath}} \geq 0 \\ &\Rightarrow \quad u_{\hat{\imath}} \leq f(x_{\hat{\imath}}), \end{aligned}$$

wobei wie oben $f(x_{\hat{\imath}}) \leq \max_{i=1,\ldots,N} f(x_i)$.

Für das Minimum betrachte $\tilde{U} = -U, \tilde{f} = -f$.

□

Folgerung 1.7. *Das lineare Gleichungssystem zu (P_h) ist stets eindeutig lösbar.*

Beweis. Das lineare Gleichungssystem ist eindeutig lösbar genau dann, wenn seine Matrix regulär ist. Die Matrix ist regulär, wenn das homogene Gleichungssystem (mit rechter Seite Null) nur die Lösung $U = 0$ hat.
Betrachten wir also das homogene Gleichungssystem d. h. für alle $i = 1, \ldots, N$ ist $f(x_i) = 0$. Es gilt also

$$\max_{i=1,\ldots,N} f(x_i) = \min_{i=1,\ldots,N} f(x_i) = 0.$$

Nach dem Maximumprinzip folgt dann

$$0 \leq \min_{i=1,\ldots,N} U_i \leq \max_{i=1,\ldots,N} U_i \leq 0,$$

d.h. $U_i = 0$ für alle $i = 1, \ldots, N$. □

Kommen wir nun zur zweiten Frage. Dazu benötigen wir vorweg einige Definitionen, die uns viel Schreibarbeit sparen werden.

Definition 1.8 (Diskreter Differentialoperator)**.**
Sei $\Omega = (0, L)$ *und* $u : \Omega \to \mathbb{R}$*. Wir schreiben*

$$L^\beta u(x) := -\beta u''(x) + u(x).$$

Sei weiter $N \in \mathbb{N}$*,* $h = \frac{L}{N-1}$ *und*

$$\Omega_h := \{x_i = (i-1)h \,|\, i = 1, \ldots, N\}$$

bezeichne die Menge der Gitterpunkte. Dann heißt $L_h^\beta u : \Omega_h \to \mathbb{R}$*,*

$$(L_h^\beta u)(x_i) := -\frac{\beta}{h^2}(u(x_{i-1}) - 2u(x_i) + u(x_{i+1})) + u(x_i),$$

diskreter Differentialoperator *zu* L^β*. Analog schreiben wir für* $U \in \mathbb{R}^N$

$$(L_h^\beta U)_i := -\frac{\beta}{h^2}(U_{i-1} - 2U_i + U_{i+1}) + U_i.$$

Mit diesen Bezeichnungen können wir nun den Begriff der *Konsistenz* definieren. Dies bezeichnet die Approximationsqualität der Ableitungen durch den gewählten Diskretisierungsansatz.

Satz 1.9 (Konsistenz)**.** *Sei* u *auf* Ω *viermal stetig differenzierbar. Dann gilt*

$$\max_{i=2,\ldots,N-1} |L^\beta u(x_i) - (L_h^\beta u)(x_i)| \leq Ch^2$$

(mit einer Konstanten C *unabhängig von* h*). Ein solches Verfahren heißt* konsistent der Ordnung 2*.*

Beweis.

$$\begin{aligned} |-\beta u''(x_i) + u(x_i) - (L_h^\beta u)(x_i)| &= \beta|-u''(x_i) + \frac{1}{h^2}(u(x_{i-1}) - 2u(x_i) + u(x_{i+1})| \\ &= \beta O(h^2) \\ &\leq \beta C h^2. \end{aligned}$$

Das zweite Gleichheitszeichen folgt aus den Überlegungen anhand des Satzes von Taylor, die wir in Abschnitt 1.1 angestellt haben. □

Satz 1.10 (Stabilität, stetige Abhängigkeit von der rechten Seite)**.**
Sei $U \in \mathbb{R}^N$ *Lösung von* (P_h)*. Dann gilt*

$$\max_{i=1,\ldots,N} |U_i| \leq C \max_{i=1,\ldots,N} |f(x_i)|$$

(mit einer Konstanten C *unabhängig von* h*). Ein solches Verfahren heißt* stabil*.*

Beweis. Aus dem Maximumprinzip folgt

$$\begin{aligned} \max_{i=1,\dots,N} U_i &\leq \max_{i=1,\dots,N} f(x_i), \\ -\min_{i=1,\dots,N} U_i &\leq -\min_{i=1,\dots,N} f(x_i). \end{aligned}$$

Also erhalten wir

$$\begin{aligned} \max_{i=1,\dots,N} |U_i| &= \max\{\max_{i=1,\dots,N} U_i, -\min_{i=1,\dots,N} U_i\} \\ &\leq \max\{\max_{i=1,\dots,N} f(x_i), -\min_{i=1,\dots,N} f(x_i)\} \\ &= \max_{i=1,\dots,N} |f(x_i)| \qquad (\text{hier } C = 1). \end{aligned}$$

□

Aus diesen beiden Eigenschaften können wir nun die Konvergenz des Verfahrens herleiten. Ein wichtiger **Merksatz** ist

Konsistenz + Stabilität ⇒ Konvergenz

Satz 1.11 (Konvergenz)**.** *Sei u Lösung von (P) und viermal stetig differenzierbar, sei U Lösung von (P_h). Dann gilt*

$$\max_{i=1,\dots,N} |u(x_i) - U_i| \leq Ch^2.$$

Ein solches Verfahren heißt konvergent zweiter Ordnung*.*

Beweis. Zunächst zeigen wir, dass der Fehler $e : \Omega_h \to \mathbb{R}$, $e(x_i) := U_i - u(x_i)$ eine Variante des diskreten Problems (P_h) mit einer speziellen rechten Seite erfüllt:

$$\begin{aligned} (L_h^\beta e)(x_i) &= (L_h^\beta U)_i - (L_h^\beta u)(x_i) \\ &= f(x_i) - (L_h^\beta u)(x_i) \\ &= -\beta u''(x_i) + u(x_i) - (L_h^\beta u)(x_i) \\ &= (L^\beta u)(x_i) - (L_h^\beta u)(x_i) =: r(x_i) \quad \text{für } i = 2, \dots N-1. \end{aligned}$$

Setzt man zusätzlich $r(x_1) := r(x_N) := 0$, so löst e tatsächlich das Problem (P_h) mit r anstelle von f. Da unser Verfahren *stabil* ist, wissen wir also

$$\max_{i=1,\dots,N} |e(x_i)| \leq C \max_{i=1,\dots,N} |r(x_i)|$$

und aus der *Konsistenz* erhalten wir für $i = 2, \dots N-1$

$$|r(x_i)| \leq Ch^2.$$

An den Randpunkten gilt dies sowieso, da r dort gleich Null ist. Zusammen folgt somit

$$\max_{i=1,\dots,N} |e(x_i)| \leq Ch^2.$$

□

Dabei verwenden wir die Bezeichnung C für alle von h unabhängigen Konstanten und verzichten auf eine Unterscheidung.

1.3 Ein zweidimensionales Beispiel

Als Beispiel betrachten wir das Entrauschen von Luftbildern. Dabei interpretieren wir ein quadratisches Graustufen-Bild als kontinuierliche Funktion

$$f : \overline{\Omega} \to [0,1] \quad \text{mit} \quad \Omega = (0,1)^2, \bar{\Omega} = [0,1]^2.$$

Dabei interpretieren wir beispielsweise den Wert 0 als schwarz und den Wert 1 als weiß.
Wir setzen wieder an

$$E[u] := \int_\Omega \beta \left|\nabla u(x,y)\right|^2 + \left(u(x,y) - f(x,y)\right)^2 \, \mathrm{d}x \, \mathrm{d}y$$

und wollen wie im eindimensionalen

$$\frac{d}{dt} E[u + tv]\Big|_{t=0} = 0$$

für alle differenzierbaren v mit $v = 0$ auf $\partial\Omega$ berechnen.
Bemerkung: Im Folgenden verwenden wir stets $|\cdot|$ für die euklidische Norm im $\mathbb{R}^n$. Die Schreibweise $\|\cdot\|$ ist für Normen auf Funktionenräumen (vgl. Definition 1.17) reserviert.
Zunächst erinnern wir uns an einige Definitionen und Resultate aus der Analysis. Dazu sei im Folgenden $u : \mathbb{R}^2 \to \mathbb{R}$ stetig differenzierbare Funktion und $F : \mathbb{R}^2 \to \mathbb{R}^2$ stetig differenzierbares Vektorfeld.

Notation 1.12 (Gradient, Divergenz, Laplace)**.** *Wir bezeichnen den* Gradienten *von u mit*

$$\begin{pmatrix} \partial_x u \\ \partial_y u \end{pmatrix} =: \nabla u := \operatorname{grad} u.$$

Wir unterscheiden insbesondere nicht zwischen ∇u und $\operatorname{grad} u$.
Die Divergenz *von F ist gegeben durch*

$$\partial_x F_1 + \partial_y F_2 =: \operatorname{div} F.$$

Schließlich schreiben wir

$$\Delta u := \operatorname{div} \nabla u = \partial_{xx} u + \partial_{yy} u.$$

Satz 1.13 (Satz von Gauß)**.** *Sei Ω eine offene und beschränkte Teilmenge von $\mathbb{R}^2$ mit stückweise glattem Rand. Dann gilt*

$$\int_\Omega \operatorname{div} F = \int_{\partial\Omega} F \cdot \nu,$$

wobei ν äußere Normale an Ω ist.

Lemma 1.14 (Produktregel)**.** *Es ist* $\operatorname{div}(uF) = \nabla u \cdot F + u(\operatorname{div} F)$.

Beweis.

$$\begin{aligned}\operatorname{div}(uF) &= \sum_{i=1}^{2} \partial_{x_i}(uF)_i = \sum_{i=1}^{2} \partial_{x_i}(uF_i) = \sum_{i=1}^{2} (\partial_{x_i} u + F_i + u\partial_{x_i} F_i) \\ &= \sum_{i=1}^{2} F_i \partial_{x_i} u + u \sum_{i=1}^{2} \partial_{x_i} F_i = F \cdot \nabla u + u(\operatorname{div} F).\end{aligned}$$

□

Satz 1.15 (Partielle Integration)**.** *Es gilt*

$$\int_\Omega \nabla u \cdot F = -\int_\Omega u(\operatorname{div} F) + \int_{\partial\Omega} uF \cdot \nu.$$

Beweis. Nach Satz von Gauß und der Rechenregel ist

$$\int_\Omega \nabla u \cdot F + \int_\Omega u(\operatorname{div} F) = \int_\Omega \operatorname{div}(uF) = \int_{\partial\Omega} uF \cdot \nu.$$

□

Folgerung 1.16. *Ersetzen wir* F *durch* ∇v*, so erhalten wir*

$$\int_\Omega \nabla u \cdot \nabla v = -\int_\Omega u(\Delta v) + \int_{\partial\Omega} u \nabla v \cdot \nu.$$

Definition 1.17 ($L^2(\Omega)$)**.** *Wir definieren durch*

$$\langle u, w \rangle_{L^2(\Omega)} := \int_\Omega u \cdot w$$

das L^2-Skalarprodukt*. Man rechnet leicht nach, dass* $\langle \cdot, \cdot \rangle$ *symmetrisch, bilinear und positiv definit ist und damit ein Skalarprodukt. Mit diesem Skalarprodukt erhält man eine Norm, die* L^2-Norm

$$\|u\|^2_{L^2(\Omega)} := \langle u, u \rangle_{L^2(\Omega)} = \int_\Omega u^2 \, \mathrm{d}x.$$

Schließlich bezeichnen wir die Menge aller quadratintegrierbaren *Funktionen mit* $L^2(\Omega)$*,*

$$L^2(\Omega) = \left\{ u \;\middle|\; \int_\Omega u^2 \, \mathrm{d}x < \infty \right\} = \left\{ u \;\middle|\; \|u\|_{L^2(\Omega)} < \infty \right\}.$$

$L^2(\Omega)$ *ist ein Vektorraum.*

Bemerkung 1.18. *Mit Hilfe des L^2-Skalarprodukts lässt sich die partielle Integration schreiben als*

$$\langle \nabla u, F\rangle_{L^2(\Omega)} = \langle u, -\mathrm{div} F\rangle_{L^2(\Omega)}.$$

Vergleichen wir dies mit dem Standardskalarprodukt auf $\mathbb{R}^n$ für eine Matrix A und Vektoren u und F:

$$\langle Au, F\rangle_{\mathbb{R}^n} = \langle u, A^t F\rangle_{\mathbb{R}^n},$$

wobei A^t die transponierte Matrix ist. Es liegt nun Nahe zu sagen, dass der zu ∇ transponierte Operator $-$div *ist. Dies wird uns später dabei helfen ∇ und* div *richtig zu diskretisieren, nämlich so dass der obige Zusammenhang gegeben ist.*

Nun können wir die "Richtungsableitungen" von E berechnen

$$\begin{aligned}
\frac{d}{dt} E[u+tv]\Big|_{t=0} &= \frac{d}{dt}\int_\Omega \beta\,|\nabla u + t\nabla v|^2 + (u+tv-f)^2\,\Big|_{t=0} \\
&= \int_\Omega 2\beta(\nabla u + t\nabla v)\cdot\nabla v + 2(u+tv-f)v\Big|_{t=0} \\
&= 2\beta\int_\Omega \nabla u\cdot\nabla v + 2\int_\Omega (u-f)v \\
&= -2\beta\int_\Omega \Delta u v + 2\beta\int_{\partial\Omega}(\nabla u\cdot\nu)\underbrace{v}_{=0} - 2\int_\Omega (f-u)v \\
&= 2\int_\Omega(-\beta\Delta u + u - f)v
\end{aligned}$$

Damit dieses Integral für alle v mit $v = 0$ auf $\partial\Omega$ verschwindet, muss gelten

$$-\beta\Delta u + u = f \qquad\qquad \text{in } \Omega\,.$$

Wir sagen $-\beta\Delta u + u - f$ ist die *erste Variation* von E.
Analog zum 1D-Fall nehmen wir hier ebenfalls an, dass auf dem Rand die Werte von u und f übereinstimmen sollen:

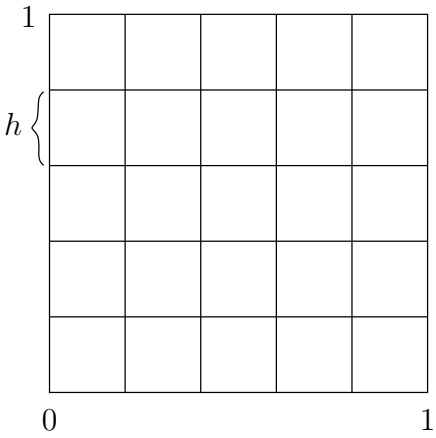
Problemstellung 1.19. *Es seien $\Omega = (0,1)^2$, $\bar\Omega = [0,1]^2$ und $\partial\Omega$ = Rand von Ω. Ferner sei $f : \bar\Omega \to \mathbb{R}$ stetig. Gesucht ist eine zweimal stetig differenzierbare Funktion $u : \bar\Omega \to \mathbb{R}$ mit*

$$\begin{aligned}
-\beta\Delta u(x,y) + u(x,y) &= f(x,y) \quad \text{für } (x,y)\in\Omega\,, \\
u(x,y) &= f(x,y) \quad \text{für } (x,y)\in\partial\Omega\,.
\end{aligned} \tag{P^2}$$

Bemerkung 1.20. *Aus dem Maximumprinzip folgt*

$$f : \bar\Omega \to [0,1] \quad \Rightarrow \quad u : \bar\Omega \to [0,1].$$

Nun approximieren wir die kontinuierliche Funktion u wiederum auf einem Gitter mit Gitterweite $h = \frac{1}{N-1}$, dabei soll jeder Gitterpunkt einem Pixel des ursprünglichen Bildes entsprechen.



Wir definieren die Knoten $(x_i, y_j) = ((i-1)h, (j-1)h)$ mit den Knotenwerten $u(x_i, y_j)$. Ein Bild bestehe aus $N \times N$ Pixeln und U_{ij} entspreche dem Grauwert des Pixels (i, j). Wie approximieren wir nun $\Delta u(x, y)$? Im Kontinuierlichen ist

$$\Delta u(x, y) = \partial_{xx} u(x, y) + \partial_{yy} u(x, y),$$

also die Summe aus den zweiten Richtungsableitungen. Zur Approximation benutzen wir die bekannte eindimensionale Approximation der Richtungsableitungen:

$$\begin{aligned} \partial_{xx} u(x_i, y_j) &\approx \frac{1}{h^2} \left(u(x_{i-1}, y_j) - 2u(x_i, y_j) + u(x_{i+1}, y_j) \right), \\ \partial_{yy} u(x_i, y_j) &\approx \frac{1}{h^2} \left(u(x_i, y_{j-1}) - 2u(x_i, y_j) + u(x_i, y_{j+1}) \right). \end{aligned}$$

Zusammen ergibt sich für den diskreten Laplace-Operator

$$\Delta u(x_i, y_j) \approx \frac{1}{h^2} \left(-4u(x_i, y_j) + u(x_{i-1}, y_j) + u(x_{i+1}, y_j) + u(x_i, y_{j-1}) + u(x_i, y_{j+1}) \right).$$

Man spricht bei dieser Diskretisierung auch von dem *5-Punkte-Stern*

$$\frac{1}{h^2} \begin{bmatrix} & 1 & \\ 1 & -4 & 1 \\ & 1 & \end{bmatrix}.$$

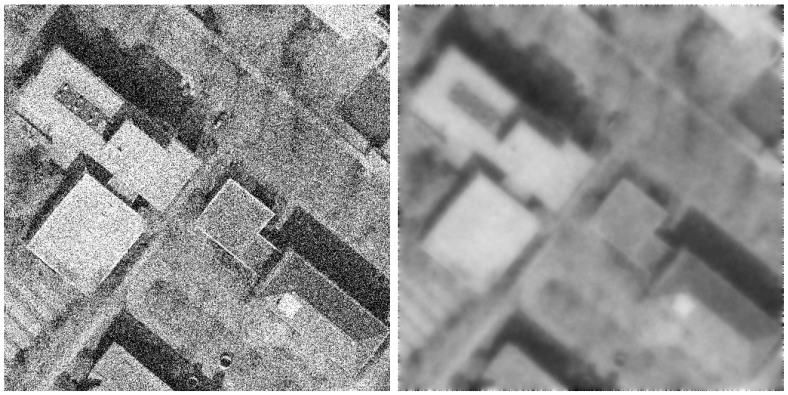
Wie im Eindimensionalen erhält man Konsistenz der Ordnung 2.
Diese Diskretisierung ergibt wie in 1D ein lineares Gleichungssystem. Um die unbekannte $U \in \mathbb{R}^{N^2}$ als Vektor zu schreiben, muss man die N^2 Pixel durchnummerieren. Es gibt zwei weit verbreitete Varianten

- column-major order: hierbei werden nacheinander die Spalten durchnummeriert, also $u(x_i, y_j) \approx U_k$ mit $k = (N(j-1) + i)$, oder
- row-major order, wobei die Zeilen durchnummeriert werden, also $u(x_i, y_j) \approx U_k$ mit $k = (N(i-1) + j)$.

Die erste Variante wird z.B. in MATLAB und Fortran benutzt, die zweite z.B. in C.

Programmieraufgabe 2. *Implementieren Sie das beschriebene Finite-Differenzen Verfahren zur Bildglättung in* `MATLAB`.

Als Ergebnis für $\beta = 10^{-4}$ erhalten wir



Beobachtung: Das Bild ist kaum noch verrauscht, aber "verschwommen". Woran liegt das?

Exkurs: Wärmeleitungsgleichung

Wir betrachten das Anfangswertproblem

$$\begin{aligned} \partial_t u - \Delta u &= 0 && t > 0, x \in \mathbb{R}^2, \\ u &= f && t = 0, x \in \mathbb{R}^2. \end{aligned}$$

Dann ist u gegeben durch

$$u(x,t) = \frac{1}{4\pi t} \int_{\mathbb{R}^2} e^{-\frac{|x-y|^2}{4t}} f(y)\, \mathrm{d}y.$$

Die Lösung der Wärmeleitungsgleichung nach einer Zeit t kann man also Gauß-Filterung der Anfangsdaten mit der Filterweite $\sigma = \sqrt{2t}$ ansehen. Was hat dies nun mit unserer Gleichung zu tun? Wenn wir die Zeiableitung ∂_t mittels Rückwärtsdifferenzen diskretisieren, erhalten wir

$$\begin{aligned} & \frac{u(x,\tau) - u(x,0)}{\tau} - \Delta u(x,\tau) = 0 \\ \iff & u(x,\tau) - \tau \Delta u(x,\tau) = u(x,0) \\ \iff & u(x,\tau) - \tau \Delta u(x,\tau) = f(x) \end{aligned}$$

und mit $\tau := \beta$

$$\iff -\beta \Delta u(x,\beta) + u(x,\beta) = f.$$

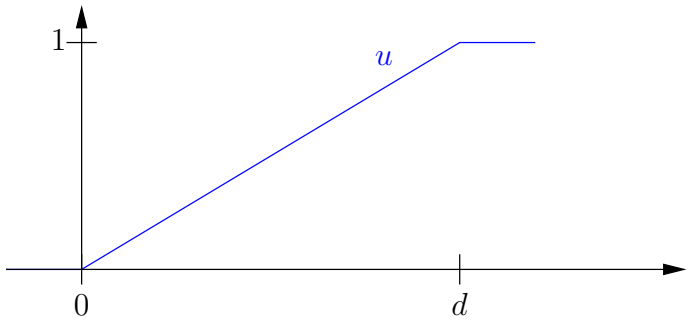
Wir können also die Differentialgleichung

$$-\beta \Delta u + u = f$$

interpretieren als den ersten Zeitschritt einer implizit diskretisierten Wärmeleitungsgleichung. Die Lösung dieser Gleichung approximiert also die Gauß-Filterung der Eingabedaten mit der Filterweite $\sqrt{2\beta}$.
Im obigen Beispiel entspricht dies einer Filterweite von ca. 7 Pixeln.

1.4 Totale Variation

Wir haben gesehen, dass die gewählte Energie dazu führt, dass Bilder stark geglättet werden. Um zu sehen, wie wir die Energie abändern müssen, um dies zu verhindern, schauen wir uns die Energie eines Übergangs von 0 nach 1 mit der Breite d (in 1D) an:



Die Funktion u ist also gegeben durch

$$u(x) = \begin{cases} 0 & x \leq 0 \\ \frac{1}{d}x & 0 \leq x \leq d \\ 1 & x \geq d \end{cases}.$$

Wir interessieren uns zunächst für den ersten Beitrag der Energie, also für $\int u'(x)^2 \, \mathrm{d}x$. Da die Ableitung außerhalb der Intervalls $(0, d)$ verschwindet, betrachten wir nur dieses Intervall:

$$\int_0^d u'(x)^2 \, \mathrm{d}x = \int_0^d \left(\frac{1}{d}\right)^2 \, \mathrm{d}x = \frac{1}{(d)^2} \int_0^d 1 = \frac{1}{(d)^2}(d) = \frac{1}{d}.$$

Dieser Wert wird kleiner, wenn d groß ist. Wir sehen also, dass dieser Energiebeitrag, den wir ja zur Verringerung des Rauschens angesetzt hatten, lange Übergänge bevorzugt und damit die Unschärfe erzeugt.
Würden wir statt dessen die Enegie $\int u'(x) \, \mathrm{d}x$ (also ohne das Quadrat) betrachten, so erhielten wir

$$\int_0^d u'(x) \, \mathrm{d}x = \int_0^d \frac{1}{d} \, \mathrm{d}x = \frac{1}{d} \int_0^d 1 = \frac{1}{d}(d) = 1.$$

Die Energie des Übergangs ist also unabhängig von der Länge des Übergangs! Es sollten also auch steile Übergänge und somit scharfe Kanten möglich sein. Da die Energie nicht von der Richtung des Übergangs abhängen darf und nicht negativ werden soll, müssen wir noch den Betrag von u' nehmen. Täten wir das nicht, könnte man durch Einfügen von zusätzlichen Übergängen in Gegenrichtung die Energie verringern.

1.4.1 TV-L^2

Wir benutzen also die neue Energie

$$E[u] = \int_\Omega \beta \, |\nabla u| + (u - f)^2. \tag{1.1}$$

Den Term $\int_\Omega |\nabla u|$ bezeichnet man mit *Totale Variation* von u, kurz TV. Der zweite Term, $\int_\Omega (u-f)^2$, ist die L^2-Norm von $u-f$. Daher bezeichnet man die Energie mit $TV\text{-}L^2$.
Da die Betragsfunktion nicht differenzierbar, können wir unsere bisherige Vorgehensweise, nämlich die erste Variation der Energie zu berechnen, nicht mehr anwenden. Um dieses Problem zu umgehen, verwenden wir folgende Darstellung

Satz 1.21. *Es gilt*

$$\int_\Omega |\nabla u(x)| \, \mathrm{d}x = \max_p \left\{ \int_\Omega u(x)(\mathrm{div} p(x)) \, \mathrm{d}x \right\},$$

wobei über alle stetig differenzierbaren Vektorfelder $p : \Omega \to \mathbb{R}^2$ maximiert wird, die punktweise maximal Länge eins haben, $|p(x)| \leq 1 \; \forall x \in \Omega$, und die in einer Umgebung des Randes $\partial\Omega$ verschwinden.

Beweis. Zunächst sehen wir mittels partieller Integration, dass

$$\int_\Omega u(\mathrm{div} p) = -\int_\Omega \nabla u \cdot p,$$

da p auf dem Rand verschwindet. Wir betrachten die Gleichung nun punktweise. Falls $\nabla u(x) = 0$, dann ist offenbar für alle p die Gleichung $|\nabla u| = \nabla u \cdot p$ erfüllt. Nehmen wir nun also an, dass $\nabla u(x) \neq 0$. Es ist $\nabla u \cdot p \leq |\nabla u| \, |p|$ (∇u und p sind in jedem Punkt Vektoren) und es gilt Gleichheit, falls ∇u und p parallel sind, d.h. $p = \lambda \nabla u$ mit $\lambda \in \mathbb{R}$. Um die Beschränkung $|p| \leq 1$ einzuhalten, wählen wir

$$p = \pm \frac{\nabla u}{|\nabla u|}.$$

Damit erhalten wir

$$\max_p \int_\Omega u(\mathrm{div} p) = \max_p -\int_\Omega \nabla u \cdot p = \int_\Omega \nabla u \cdot \frac{\nabla u}{|\nabla u|} = \int_\Omega \frac{|\nabla u|^2}{|\nabla u|} = \int_\Omega |\nabla u| \, .$$

□

Durch diesen Trick können wir die Energie (1.1) umschreiben in

$$E[u] = \max_{|p| \leq 1} \beta \int_\Omega u(\mathrm{div} p) + \int_\Omega (u-f)^2,$$

wobei die Bedingungen an p aus dem Satz mit $|p| \leq 1$ abgekürzt wurden.
Die Minimierungsaufgabe ist nun also

$$\begin{aligned} \min_u E[u] &= \min_u \max_{|p| \leq 1} \beta \int_\Omega u(\mathrm{div} p) + \int_\Omega (u-f)^2 \\ &= \max_{|p| \leq 1} \min_u \beta \int_\Omega u(\mathrm{div} p) + \int_\Omega (u-f)^2 \\ &=: \max_{|p| \leq 1} \min_u F[u]. \end{aligned}$$

Wir können nun die Minimierungsaufgabe $\min_u F[u]$ (für p fest) wie üblich durch berechnen der ersten Variation lösen:

$$\begin{aligned} 0 = \frac{d}{dt} F[u+tv]\Big|_{t=0} &= \frac{d}{dt} \int_\Omega \beta(u+tv)(\mathrm{div} p) + (u+tv-f)^2\Big|_{t=0} \\ &= \int_\Omega \beta v(\mathrm{div} p) + 2(u+tv-f)v\Big|_{t=0} \\ &= \int_\Omega \left(\beta(\mathrm{div} p) + 2(u-f)\right) v. \end{aligned}$$

Als Bedingung erhalten wir also

$$\beta(\mathrm{div} p) + 2(u-f) = 0 \iff u = -\frac{\beta}{2}(\mathrm{div} p) + f. \tag{1.2}$$

Einsetzen liefert

$$\begin{aligned} \min_u E[u] &= \max_{|p|\le 1} \int_\Omega \beta(-\frac{\beta}{2}(\mathrm{div} p) + f)(\mathrm{div} p) + \int_\Omega \frac{\beta^2}{4}(\mathrm{div} p)^2, \\ &= \max_{|p|\le 1} -\int_\Omega \frac{\beta^2}{2}(\mathrm{div} p)^2 + \int_\Omega \beta f(\mathrm{div} p) + \int_\Omega \frac{\beta^2}{4}(\mathrm{div} p)^2 \\ &= \max_{|p|\le 1} \beta \int_\Omega f(\mathrm{div} p) - \int_\Omega \frac{\beta^2}{4}(\mathrm{div} p)^2. \end{aligned}$$

Statt der Minimierung eines nicht differenzierbaren Funktionals haben wir also nun die Maximierung unter Nebenbedingungen eines differenzierbaren Funktionals erhalten. Hat man das Maximum p^* gefunden, so erhält man die Funktion u^* durch (1.2).

Minimierung unter Ungleichungs-Nebenbedingungen

In diesem Abschnitt leiten wir Optimalitätsbedingungen für Minimierungen unter Ungleichungs-Nebenbedingungen (restringierte Optimierung) her. Wir besprechen die Theorie für Funktionen im $\mathbb{R}^n$ und übertragen die Ergebnisse dann auf den Fall unseres Energiefunktionals.
Erinnern wir uns zunächst an Gleichungs-Nebenbedingungen: Zu dem Minimierungsproblem

$$\min_{p\in\mathbb{R}^n} f(p) \quad \text{unter der Nebenbedingung } g(p) = 0$$

heißt

$$L(p,\lambda) := f(p) + \lambda g(p)$$

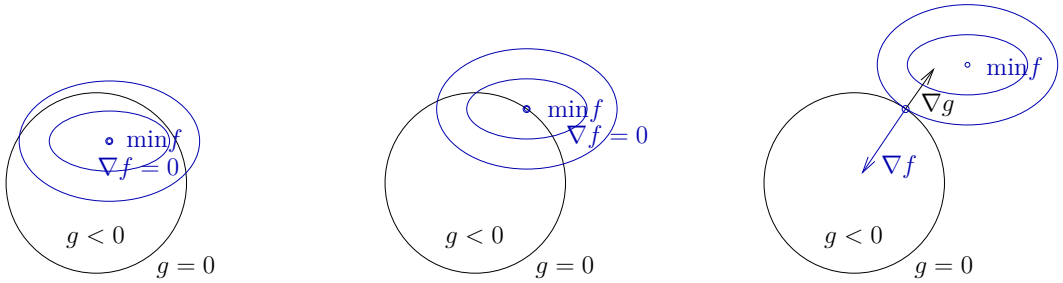
Lagrange-Funktion und $\lambda \in \mathbb{R}$ heißt *Lagrange-Multiplikator*.
Zu einer Minimalstelle p^* gibt es dann stets ein λ^*, so dass

$$\begin{aligned} 0 &= \nabla_p L(p^*,\lambda^*) = \nabla f(p^*) + \lambda^* \nabla g(p^*), \\ 0 &= \nabla_\lambda L(p^*,\lambda^*) = g(p^*). \end{aligned}$$

Die zweite Gleichung gilt, da die Nebenbedingung erfüllt sein muss. Die erste Gleichung besagt, dass $\nabla f(p^*)$ in die selbe Richtung wie $\nabla g(p^*)$ zeigt. Der Anteil von

$\nabla f(p^*)$, der senkrecht zu $\nabla g(p^*)$ ist – also tangential zur durch $g(p) = 0$ beschriebenen Niveaumenge – muss nämlich Null sein.
Die selbe Funktion $L(p, \lambda)$ kann man auch im Fall von Ungleichungs-Nebenbedingungen definieren. Wir betrachten nun zur Veranschaulichung die verschiedenen Fälle, wo das unrestringierte Minimum von f in Bezug auf die zulässige Menge liegt:

Im Inneren	Auf dem Rand	Außen



$\nabla f(p^*) = 0 = -\lambda^* \nabla g(p^*)$,	$\nabla f(p^*) = 0 = -\lambda^* \nabla g(p^*)$,	$\nabla f(p^*) = -\lambda^* \nabla g(p^*)$,
$\lambda^* = 0$,	$\lambda^* = 0$,	$\lambda^* > 0$,
$g(p^*) < 0$,	$g(p^*) = 0$,	$g(p^*) = 0$.

Die Situation im (zweiten und) dritten Fall liegt analog zur Gleichungs-Nebenbedingung, da $g(p^*) = 0$. Zusammen erhalten wir also

$$\begin{aligned} \nabla f(p^*) &= -\lambda^* \nabla g(p^*), \\ \lambda^* &\geq 0, \\ g(p^*) &\leq 0, \\ \lambda^* g(p^*) &= 0. \end{aligned}$$

Nun definieren wir einige Begriffe, die wir bei der Behandung von Ungleichungs-Nebenbedingungen benötigen.

Definition 1.22 (Konvexe Funktion)**.** *Eine Funktion $f : \mathbb{R}^n \to \mathbb{R}$ heißt konvex, wenn*

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda) f(y)$$

für alle $x, y \in \mathbb{R}^n$ und alle $\lambda \in [0, 1]$ gilt.

Lemma 1.23. *Eine differenzierbare Funktion $f : \mathbb{R}^n \to \mathbb{R}$ ist genau dann konvex, wenn für alle $x, y \in \mathbb{R}^n$*

$$f(x) \geq f(y) + \nabla f(y) \cdot (x - y)$$

gilt.

Betrachtet man für eine zweimal stetig differenzierbare Funktion f das entsprechende Restglied der Taylorentwicklung, so sieht man, dass diese Ungleichung jedenfalls dann erfüllt ist, wenn die Hesse-Matrix von f positiv semidefinit ist.

Bemerkung 1.24. *In 1d mit $h := x - y$ heißt das*

$$f(y+h) \geq f(y) + f'(y)h,$$

der Graph der Funktion verläuft also immer oberhalb der Tangente.

Nun kommen wir zu den Optimalitätsbedingungen.

Definition 1.25 (Optimalitätsbedingungen)**.** *Seien $f : \mathbb{R}^n \to \mathbb{R}$ und $g : \mathbb{R}^n \to \mathbb{R}$ differenzierbar und konvex. Zu dem Minimierungsproblem*

$$\min_{p \in \mathbb{R}^n} f(p) \quad \text{unter der Nebenbedingung } g(p) \leq 0 \tag{1.3}$$

heißt

$$L(p, \lambda) := f(p) + \lambda g(p)$$

Lagrange-Funktion *und $\lambda \in \mathbb{R}$ heißt* Lagrange-Multiplikator*. Ein Punkt $(p^*, \lambda^*) \in \mathbb{R}^n \times \mathbb{R}$ heißt* KKT-Punkt *(nach Karush, Kuhn und Tucker) von* (1.3)*, wenn er die Bedingungen*

$$\begin{aligned} \nabla_p L(p^*, \lambda^*) &= 0 \\ \lambda^* &\geq 0 \\ g(p^*) &\leq 0 \\ \lambda^* g(p^*) &= 0 \end{aligned}$$

erfüllt.

Bemerkung 1.26. *Die letzte Bedingung heißt, dass in einem KKT-Punkt immer $\lambda^* = 0$ oder $g(p^*) = 0$ gilt. Im ersten Fall ist p^* auch ein Minimum von f ohne Nebenbedingung, denn $L(p, 0) = f(p)$ und daher $0 = \nabla_p L(p^*, 0) = \nabla f$. Wir sagen in diesem Fall, dass die Nebenbedingung nicht aktiv ist. Im zweiten Fall, wenn $g(p^*) = 0$, liegt p^* also auf dem Rand des zulässigen Gebietes. Wir sagen, die Nebenbedingung ist aktiv.*

Satz 1.27. *Seien $f : \mathbb{R}^n \to \mathbb{R}$ und $g : \mathbb{R}^n \to \mathbb{R}$ konvex und (p^*, λ^*) sei KKT-Punkt von* (1.3)*. Dann ist p^* Minimalpunkt von f unter der Nebenbedingung $g(p) \leq 0$.*

Beweis. Nach Voraussetzung ist

$$\nabla f(p^*) + \lambda^* \nabla g(p^*) = 0 \iff \nabla f(p^*) = -\lambda^* \nabla g(p^*).$$

Sei nun p ein beliebiger zulässiger Vektor, d.h. $g(p) \leq 0$. Da f konvex ist, gilt

$$\begin{aligned} f(p) &\geq f(p^*) + \nabla f(p^*) \cdot (p - p^*) \\ &= f(p^*) - \lambda^* \nabla g(p^*) \cdot (p - p^*). \end{aligned}$$

Falls $\lambda^* = 0$ ist, dann gilt also $f(p) \geq f(p^*)$. Ist andererseits $\lambda^* > 0$, dann muss $g(p^*) = 0$ sein und da g konvex

$$\begin{aligned} g(p) &\geq \underbrace{g(p^*)}_{=0} + \nabla g(p^*) \cdot (p - p^*) \\ \iff \nabla g(p^*) \cdot (p - p^*) &\leq g(p) \leq 0. \end{aligned}$$

Damit erhalten wir auch im zweiten Fall $f(p) \geq f(p^*)$. Da dies für alle zulässigen p gilt, ist p^* ein Minimierer von f. □

Beispiel (in 2D):

$$f(x, y) = \left| \begin{pmatrix} x \\ y \end{pmatrix} - \begin{pmatrix} x_0 \\ 0 \end{pmatrix} \right|^2 \quad \text{mit } x_0 \in \mathbb{R}.$$

Wir berechnen den KKT-Punkt zum Minimierungsproblem

$$\min_{x^2+y^2 \leq 1} f(x, y).$$

Dazu setzen wir $g(x, y) := x^2 + y^2 - 1$. Dann ist g konvex und

$$g(x, y) \leq 0 \iff x^2 + y^2 \leq 1.$$

Die Bedingungen für einen KKT-Punkt sind

$$\nabla f(x^*, y^*) + \lambda^* \nabla g(x^*, y^*) = 0 \iff \begin{aligned} 2(x^* - x_0) + 2\lambda^* x^* &= 0 \\ 2y^* + 2\lambda^* y^* &= 0 \end{aligned}$$

sowie

$$\lambda^* \geq 0, \quad g(x^*, y^*) \leq 0, \quad \text{und} \quad \lambda^* g(x^*, y^*) = 0.$$

Aus der zweiten Gleichung ergibt sich

$$(1 + \lambda^*) y^* = 0,$$

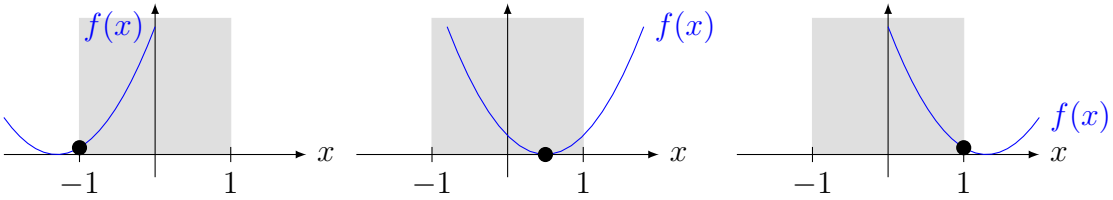
und da $\lambda^* \geq 0$ muss $y^* = 0$ sein. Die Nebenbedingung $g(x^*, y^*) \leq 0$ ist also äquivalent zu $|x^*| \leq 1$.
Falls die Nebenbedingung nicht aktiv ist, also $\lambda^* = 0$, dann folgt aus der ersten Gleichung $x^* = x_0$ und $((x_0, 0), 0)$ ist KKT-Punkt. Hierzu muss $|x_0| \leq 1$ sein. Wie erwartet ist dies auch das Minimum des unrestringierten Minimierungsproblems.
Falls die Nebenbedingung aktiv ist, also $\lambda^* > 0$, dann muss $g(x^*, y^*) = 0$ sein, also $x^* = \pm 1$. Wir unterscheiden

- $x^* = +1$. Dann $(1 + \lambda^*) = x_0 \iff \lambda^* = x_0 - 1$ und $((1, 0), x_0 - 1)$ ist KKT-Punkt falls $x_0 > 1$ (wegen $\lambda^* > 0$).

- $x^* = -1$. Dann $-(1 + \lambda^*) = x_0 \iff \lambda^* = -x_0 - 1$ und $((-1, 0), -x_0 - 1)$ ist KKT-Punkt falls $x_0 < -1$.

Zusammen erhalten wir also für den KKT-Punkt

$$((x^*, y^*), \lambda^*) = \begin{cases} ((-1,0), -x_0 - 1) & x_0 < -1, \\ ((x_0, 0), 0) & -1 \leq x_0 \leq 1, \\ ((1,0), x_0 - 1) & x_0 > 1. \end{cases}.$$



Kommen wir nun zurück zu unserem Minimierungsproblem

$$\min_u E[u] = \max_{|p|\leq 1} \beta \int_\Omega f(\mathrm{div} p) - \int_\Omega \frac{\beta^2}{4} (\mathrm{div} p)^2.$$

Wir können dies leicht in ein Minimierungsproblem umschreiben:

$$\min_u E[u] = -\min_{|p|\leq 1} -\beta \int_\Omega f(\mathrm{div} p) + \int_\Omega \frac{\beta^2}{4} (\mathrm{div} p)^2.$$

Um die KKT-Bedingungen herzuleiten, berechnen wir wieder die erste Variation:

$$\begin{aligned} 0 &= \frac{d}{dt} \int_\Omega -\beta f(\mathrm{div} p + t\mathrm{div} q) + \frac{\beta^2}{4} (\mathrm{div} p + t\mathrm{div} q)^2 \Big|_{t=0} \\ &= -\beta \int_\Omega f(\mathrm{div} q) + \frac{\beta^2}{4} \int_\Omega 2(\mathrm{div} p + t\mathrm{div} q)(\mathrm{div} q) \Big|_{t=0} \\ &= -\beta \int_\Omega f(\mathrm{div} q) + \frac{\beta^2}{2} \int_\Omega (\mathrm{div} p)(\mathrm{div} q) \\ &= \beta \int_\Omega \nabla f \cdot q - \frac{\beta^2}{2} \int_\Omega \nabla \mathrm{div} p \cdot q \\ &= \int_\Omega \left(-\frac{\beta^2}{2} \nabla \mathrm{div} p + \beta \nabla f \right) \cdot q. \end{aligned}$$

Dabei verwenden wir wieder Testfunktionen q, die auf dem Rand Null sind, so dass das Randintegral bei der partiellen Integration wegfällt. Die erste Variation ist also

$$-\beta \nabla \left(\frac{\beta}{2} \mathrm{div} p(x) - f(x) \right).$$

Wenn wir die erste Variation statt des Gradienten verwenden, erhalten wir als Bedingung für einen KKT-Punkt

$$-\beta \nabla \left(\frac{\beta}{2} \mathrm{div} p(x) - f(x) \right) + 2\lambda(x) p(x) = 0. \tag{1.4}$$

Anders als im Fall einer Minimierung in $\mathbb{R}^n$ ist der Lagrange-Multiplikator λ hier keine Zahl, sondern eine Funktion.

Lemma 1.28. *Es ist* $\lambda(x) = \frac{\beta}{2}\left|\nabla\left(\frac{\beta}{2}(\mathrm{div}p(x)) - f(x)\right)\right|$.

Beweis. Wir zeigen die Aussage wieder punktweise. Falls $\lambda(x) > 0$, dann muss $|p| = 1$ sein. Nehmen wir den Betrag von (1.4), so erhalten wir in x

$$\lambda(x) = |\lambda(x)| = |\lambda(x)p(x)| = \frac{\beta}{2}\left|\nabla\left(\frac{\beta}{2}(\mathrm{div}p(x)) - f(x)\right)\right|.$$

Falls $\lambda(x) = 0$, so ist nach (1.4) auch die rechte Seite Null. □

Einsetzen liefert

$$-\nabla\left(\frac{\beta}{2}(\mathrm{div}p(x)) - f(x)\right) + \left|\nabla\left(\frac{\beta}{2}(\mathrm{div}p(x)) - f(x)\right)\right| p(x) = 0. \tag{1.5}$$

Diese Gleichung wollen wir nun mit Hilfe einer Fixpunktiteration numerisch lösen.

Fixpunktiterationen

Allgemein ist eine Fixpunktgleichung von der Form $f(x^*) = x^*$. Gesucht ist also ein Fixpunkt x^* von f. Man kann nun eine Gleichung in eine Fixpunktgleichung umformulieren und dann die Fixpunktgleichung mittels einer Fixpunktiteration lösen. Der einfachste Fall einer Fixpunktiteration ist die *Richardson-Iteration*. Um ein lineares Gleichungssytem $Ax = b$ zu lösen, addiert man Null

$$Ax = b \iff x - x + Ax = b \iff x = x + b - Ax$$

und setzt

$$x^{n+1} = x^n + b - Ax^n \iff x^{n+1} = (\mathrm{id} - A)x^n + b. \tag{1.6}$$

Gesucht ist nun x^* mit $x^* = (\mathrm{id} - A)x^* + b$ und somit $Ax^* = b$. Die Konvergenz der Iteration (1.6) hängt von den Eigenwerten der *Iterationsmatrix* $(\mathrm{id} - A)$ ab.
Um die Iterationsmatrix zu ändern (und damit das Konvergenzverhalten), kann man z.B.

- $x/\tau - x/\tau$ (mit einer Zahl τ) statt $x - x$ addieren,
- A aufteilen in $A = A_1 + A_2$ und die Summanden unterschiedlich behandeln.

Damit ändert sich obige Rechnung zu

$$\tfrac{1}{\tau}x - \tfrac{1}{\tau}x + (A_1 + A_2)x = b \iff x = x + \tau\left(b - (A_1 + A_2)x\right)$$

und man setzt

$$x^{n+1} = x^n + \tau\left(b - A_1x^{n+1} - A_2x^n\right),$$

also

$$(\mathrm{id} + \tau A_1)x^{n+1} = (\mathrm{id} - \tau A_2)x^n + \tau b \tag{1.7a}$$
$$\iff x^{n+1} = (\mathrm{id} + \tau A_1)^{-1}((\mathrm{id} - \tau A_2)x^n + \tau b). \tag{1.7b}$$

Die Iterationsmatrix ist nun $(\mathrm{id} + \tau A_1)^{-1}(\mathrm{id} - \tau A_2)$. Um x^{n+1} zu bestimmen würde man nicht $(\mathrm{id} + \tau A_1)$ invertieren, sondern das Gleichungssystem (1.7a) lösen.

Bevor wir die Fixpunktiteration auf die Gleichung (1.5) anwenden, teilen wie die Gleichung durch $\beta/2$:

$$-\nabla\left((\mathrm{div}p(x)) - \tfrac{2}{\beta}f(x)\right) + \left|\nabla\left((\mathrm{div}p(x)) - \tfrac{2}{\beta}f(x)\right)\right| p(x) = 0.$$

Dann verwenden wir punktweise die folgende Fixpunktiteration:

$$\begin{aligned}\tfrac{1}{\tau}p^{n+1}(x) &= \tfrac{1}{\tau}p^n(x) + \nabla\left((\mathrm{div}p^n(x)) - \tfrac{2}{\beta}f(x)\right) - \left|\nabla\left((\mathrm{div}p^n(x)) - \tfrac{2}{\beta}f(x)\right)\right| p^{n+1}(x) \\ \iff p^{n+1}(x) &= \frac{p^n(x) + \tau\nabla(\mathrm{div}p^n(x) - \tfrac{2}{\beta}f(x))}{1 + \tau\left|\nabla(\mathrm{div}p^n(x) - \tfrac{2}{\beta}f(x))\right|}.\end{aligned}$$

Der Vorteil dieser Formulierung besteht darin, dass

- der neue Wert p^{n+1} einfach und ohne Lösen eines Gleichungssystems berechnet werden kann,
- die Konvergenz des Verfahrens sich durch die Wahl von τ steuern lässt und
- die Iteration für τ klein genug konvergiert.

Wir stoppen die Fixpunktiteration, wenn sich p nicht mehr oder kaum noch ändert, denn wenn $p^{n+1} \approx p^n$, dann ist p^n (fast) ein Fixpunkt.

Ortsdiskretisierung

Wir benutzen wie in Abschnitt 1.3 ein äquidistantes zweidimensionales Gitter mit Gitterweite h. Wie vorher Diskretisieren wir die eindimensionalen Richtungsableitungen ∂_x und ∂_y durch Vorwärtsdifferenzen. Für das Entrauschen eins Bildes schien unsere Randbedingung $u = f$ auf $\partial\Omega$ nicht ideal zu sein, da sich so ein unschöner Rand bildete. Wir verwenden diesmal die Randbedingung

$$\nabla u \cdot \nu = 0,$$

wobei ν die äußere Normale ist. Bei einem rechteckigen Ω heißt dies, dass $\partial_x u = 0$ am linken und rechten Rand und $\partial_y u = 0$ am oberen und unteren Rand. Wir erhalten also für die diskreten, approximierten Gradienten ∂_x^h und ∂_y^h

$$\partial_x^h u(x_i, y_j) = \begin{cases} \frac{1}{h}(u(x_{i+1}, y_j) - u(x_i, y_j)) & \text{falls } i < N \\ 0 & \text{falls } i = N \end{cases}$$

und

$$\partial_y^h u(x_i, y_j) = \begin{cases} \frac{1}{h}(u(x_i, y_{j+1}) - u(x_i, y_j)) & \text{falls } j < N \\ 0 & \text{falls } j = N \end{cases}$$

Das Vektorfeld p speichern wir in zwei Komponenten $p^{(1)}$ und $p^{(2)}$. Die Divergenz von p erhält man dann nach Bemerkung 1.18 durch Negieren und Transponieren der Matrixdarstellung des Gradienten:

$$(\mathrm{div}^h p)(x_i, y_j) = \frac{1}{h}\begin{cases} p^{(1)}(x_i, y_j) & \text{falls } i = 1 \\ p^{(1)}(x_i, y_j) - p^{(1)}(x_{i-1}, y_j) & \text{falls } 1 < i < N \\ -p^{(1)}(x_{i-1}, y_j) & \text{falls } i = N \end{cases}$$
$$+ \frac{1}{h}\begin{cases} p^{(2)}(x_i, y_j) & \text{falls } j = 1 \\ p^{(2)}(x_i, y_j) - p^{(2)}(x_i, y_{j-1}) & \text{falls } 1 < j < N \\ -p^{(2)}(x_i, y_{j-1}) & \text{falls } j = N \end{cases}$$

Das Aufstellen der Matrizen verläuft analog zu Abschnitt 1.3. Statt `spdiags` zu verwenden und die entsprechenden Zeilen und Spalten Null zu setzen, kann man auch die Funktion `sparse` benutzen. Der Aufruf ist

```
S=sparse(i,j,s,n,n);
```

mit Vektoren i, j, s, die für jeden nicht-Null-Eintrag der Matrix einen Eintrag haben. Die Matrix S ist dann eine $n \times n$ Matrix mit $S_{i(k),j(k)} = s(k)$ für alle k. Für den Gradienten in x-Richtung also z.B.

Schema 1.29 (`MATLAB`-Code für x-Ableitung)**.**

```
rows=N*N;
nzs=2*N*(N-1);
row=zeros(nzs,1);
col=zeros(nzs,1);
val=zeros(nzs,1);
count=0;
for i=1:(N-1)
    for j=1:N
        count=count+1;
        row(count)=(i-1)*N+j;
        col(count)=(i-1)*N+j;
        val(count)=-1/h;
        count=count+1;
        row(count)=(i-1)*N+j;
        col(count)=i*N+j;
        val(count)=1/h;
    end
end
dx=sparse(row,col,val,rows,rows);
```

Achtung: Die Vektoren i, j und s in jedem Schritt um 1 zu verlängern wäre sehr langsam. Daher sollte unbedingt vorher mittels `zeros` die richtige Menge Speicher reserviert werden.
Als Abbruchkriterium wählen wir

$$\max_{ij} \left| p^{n+1}(x_i, y_j) - p^n(x_i, y_j) \right| < \theta$$

mit θ klein.

Bemerkung 1.30. *Mit der Energie aus Abschnitt 1.3 war es egal, ob die Helligkeit eines Bildpunktes durch Werte in* $[0,1]$ *oder* $[0,255]$ *(wie* `imread` *es liefert) gerechnet wurde, denn*

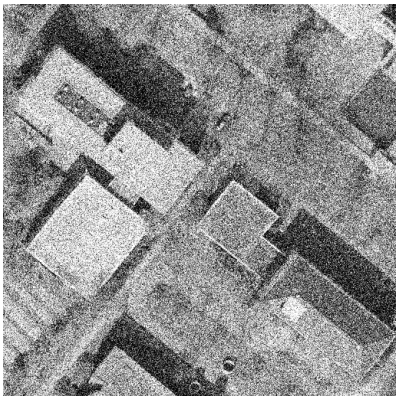
$$E[\lambda u] = \int_\Omega \beta \left|\lambda \nabla u\right|^2 + (\lambda u - \lambda f)^2 = \lambda^2 \int_\Omega \beta \left|\nabla u\right|^2 + (u-f)^2.$$

Dies ist nun anders, denn

$$E[\lambda u] = \int_\Omega \beta \left|\lambda \nabla u\right| + (\lambda u - \lambda f)^2 = \lambda^2 \int_\Omega \hat{\beta} \left|\nabla u\right|^2 + (u-f)^2$$

mit $\hat{\beta} = \beta/\lambda$*. Man muss* $\hat{\beta}$ *also je nach verwendeter Skalierung anpassen.*

Als Ergebnis für $\tau = 7.6 \cdot 10^{-7}, \beta = 8 \cdot 10^{-4}, \theta = 10^{-3}, \Omega = [0,1]^2, h = 1/511$ erhalten wir nach ca. 60 Sekunden




Wie wir sehen, ist das Ergebnis weit weniger verschwommen als zuvor. Ein Nachteil dieses Verfahrens ist allerdings, dass der Kontrast des Bildes verringert wird.

Bemerkung 1.31. *In der Rechnung wurde ein Bild mit den Dimensionen* $\Omega = [0,1]^2$ *angenommen. Da das Bild* 512×512 *Pixel hat, ergibt sich ein Wert von* $h = 1/511$*. Hierbei ist zu beachten, dass die Schrittweite* τ *sich mit* h^2 *verändert. Im Programm wurde* $\tau = 0.2h^2$ *gesetzt.*
Als andere Möglichkeit würde man $h = 1$ *setzen, also* $\Omega = [1,512]^2$*. Dann ist* τ *zwar unabhängig von der Auflösung des Bildes, aber dafür muss man* β *anpassen, wenn sich die Auflösung ändert.*
Der Vorteil der ersten Variante ist also, dass man das passende β *für ein niedrig aufgelöstes Bild suchen kann und dann die höher aufgelöste Variante mit demselben* β *rechnen kann.*

1.4.2 TV-L^1

Um den Nachteil des verringerten Kontrastes zu beheben, betrachten wir folgende weitere Änderung an der Energie

$$E[u] = \int_\Omega \beta \left|\nabla u\right| + \left|u - f\right|.$$

Da der zweite Term der (hier nicht eingeführten) L^1-Norm von $u - f$ entspricht, bezeichnet man diese Energie als *TV-L^1*.
Anders als beim TV-Term, bei dem wir eine äquivalente Formulierung gefunden hatten, wählen wir für den L^1-Term eine Approximation:

$$\tilde{E}[u, v] = \int_\Omega \beta \left|\nabla u\right| + \frac{1}{2\varepsilon} \int_\Omega (u + v - f)^2 + \int_\Omega \left|v\right|,$$

wobei ε sehr klein ist. Die Idee dahinter ist, dass durch das kleine ε der Term $\int_\Omega (u + v - f)^2$ sehr stark bestraft wird. Für einen Minimierer (u^*, v^*) der Energie erwarten wir also $u^* + v^* - f \approx 0$, also $|v^*| \approx |u^* - f|$. Damit

$$\tilde{E}[u^*, v^*] \approx \int_\Omega \beta \left|\nabla u^*\right| + \frac{1}{2\varepsilon} \int_\Omega 0 + \int_\Omega \left|u^* - f\right| = E[u^*].$$

Um den Minimierer (u^*, v^*) zu finden, minimieren wir separat in u und v und iterieren:

1. Für v^n gegeben und fest finde den Minimierer u^{n+1} von

   $$\min_u \int_\Omega \beta \left|\nabla u\right| + \frac{1}{2\varepsilon} \int_\Omega (u + v^k - f)^2 + \int_\Omega \left|v^k\right|.$$

   Da der Minimierer der Gleiche bleibt, wenn wir die Zahl $\int_\Omega \left|v^k\right|$ weg lassen und die Energie mit 2ε multiplizieren, können wir auch den Minimierer von

   $$\min_u \int_\Omega \tilde{\beta} \left|\nabla u\right| + \int_\Omega (u - \tilde{f})^2 \tag{1.8a}$$

   berechnen, wobei $\tilde{\beta} := 2\varepsilon\beta$ und $\tilde{f} = f - v^k$.

2. Für u^{n+1} fest finde den Minimierer v^{n+1} von

   $$\min_v \int_\Omega \frac{1}{2\varepsilon}(v - g)^2 + \int_\Omega \left|v\right|, \tag{1.8b}$$

   wobei $g = f - u^{n+1}$ und wir wieder die Zahl $\int_\Omega \beta \left|\nabla u^{n+1}\right|$ weg lassen konnten.

Um (1.8a) zu lösen, können wir also den TV-L^2-Algorithmus aus dem vorigen Abschnitt verwenden. Der zweite Teil (1.8b) lässt sich leicht lösen.
Dazu betrachten wir das 1d Minimierungsproblem

$$\min_{v \in \mathbb{R}} \frac{1}{2\varepsilon}(v - g)^2 + |v|$$

und nehmen an, v^* sei ein Minimierer. Dann gibt es drei Möglichkeiten

- $v^* < 0$. Dann ist v^* Minimierer von

  $$\min_{v \in \mathbb{R}} \frac{1}{2\varepsilon}(v - g)^2 - v,$$

  erfüllt also

  $$\frac{1}{\varepsilon}(v^* - g) - 1 = 0 \iff v^* = g + \varepsilon$$

  und die Annahme $v^* < 0$ ist äquivalent zu $g < -\varepsilon$.

- $v^* = 0$.
- $v^* > 0$. Dann ist v^* Minimierer von

$$\min_{v \in \mathbb{R}} \frac{1}{2\varepsilon}(v - g)^2 + v,$$

erfüllt also

$$\frac{1}{\varepsilon}(v^* - g) + 1 = 0 \iff v^* = g - \varepsilon$$

und die Annahme $v^* > 0$ ist äquivalent zu $g > \varepsilon$.

Die Lösung von (1.8b) ist also gegeben durch

$$v^{n+1} = \begin{cases} g + \varepsilon & \text{falls } g < -\varepsilon \\ 0 & \text{falls } -\varepsilon < g < \varepsilon \\ g - \varepsilon & \text{falls } g > \varepsilon \end{cases}$$

Die Iteration soll abgebrochen werden, wenn

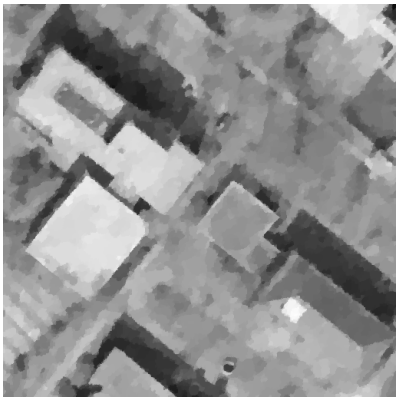
$$\max_{ij} \left| v^{n+1}(x_i, y_j) - v^n(x_i, y_j) \right| < \theta.$$

Programmieraufgabe 3.

1. *Implementieren Sie das TV-L^2-Verfahren zur Bildglättung in* `MATLAB` *als Funktion.*
2. *Implementieren Sie das TV-L^1-Verfahren zur Bildglättung in* `MATLAB` *und benutzen Sie dabei obige Funktion.*
3. *Wenden Sie beide Verfahren auf das Testbild (weißer Kreis auf schwarzem Grund) an. Wie ändert sich das Resultat bei verschiedenen β. Achten Sie auf den Kontrast, hier die Differenz aus dunkelstem und hellstem Grauwert. Gibt es einen Zusammenhang zwischen dem Radius des Kreises und dem Verhalten des TV-L^1-Verfahren?*
4. *Für welche Werte von τ konvergiert die Iteration und wie schnell?*
5. *Wenden Sie beide Verfahren auf das Luftbild an.*

Bemerkung: Wenn Sie das Bild auf $\Omega = [0, 1]^2$ und nicht auf $\Omega = [1, \texttt{\#Pixel}]^2$ skalieren, also den Gradienten mit $h \neq 1$ verwenden, können Sie zuerst die kleinen Testbilder verwenden und mit dem gleichen β dann die großen Testbilder rechnen. Es spart viel Rechenzeit, wenn man das letzte p aus der inneren (TV-L^2) Iteration als Startwert für die nächste innere Iteration verwendet.

Als Ergebnis für $\tau = 7.6 \cdot 10^{-7}, \beta = 2 \cdot 10^{-3}, \theta = 10^{-3}, L = 1, h = 1/511$ und $\varepsilon = 10^{-2}$ erhalten wir nach ca. 35 Minuten



2 Finite Elemente

Finite Differenzen haben den Vorteil, einfach implementierbar und effizient zu sein. Allerdings

- sind sie unflexibel bei der Approximation des Gebiets Ω, falls dieses kein Rechteck bildet (z.B. bei Messpunkten auf einem Dreiecksgitter).
- besitzen sie nur dann eine gute Konvergenzordnung, wenn die kontinuierliche Lösung sehr glatt ist (viermal stetig differenzierbar!)

Abhilfe schafft die sogenannte Finite-Elemente-Methode.

2.1 Schwache Lösungen

Wir wollen nun einen Lösungsbegriff definieren, der auch dann gültig ist, wenn u weniger glatt ist.
Erinnern wir uns nochmal an die Herleitung der notwendigen Bedingung

$$-\beta\Delta u + u = f$$

in Abschnitt 1.3. Es war

$$E[u] = \int \beta \left|\nabla u\right|^2 + (u-f)^2$$

und es sollte gelten

$$0 = \frac{d}{dt} E[u+tv]\Big|_{t=0}$$

für alle stetig differenzierbaren $v : \bar{\Omega} \to \mathbb{R}$ mit $v = 0$ auf $\partial\Omega$, also

$$\begin{aligned} 0 &= \frac{d}{dt}\int_\Omega \beta \left|\nabla u + t\nabla v\right|^2 + (u+tv-f)^2\Big|_{t=0} \\ &= \int_\Omega 2\beta(\nabla u + t\nabla v)\cdot\nabla v + (u+tv-f)^2 v\Big|_{t=0} \\ &= 2\int_\Omega \beta\nabla u\cdot\nabla v + (u-f)v. \end{aligned}$$

Bisher hatten wir weiter gerechnet

$$\begin{aligned} &= 2\int_\Omega -\beta\Delta u\, v + (u-f)v + 2\int_{\partial\Omega} \nabla u\cdot\nu\, v \\ &= 2\int_\Omega \Big(-\beta\Delta u + (u-f)\Big)v. \end{aligned}$$

Stattdessen sagen wir, dass u eine *schwache Lösung* von $-\beta\Delta u + u = f$ ist, wenn für alle „Testfunktionen“ v gilt

$$\int_\Omega \beta \nabla u \cdot \nabla v + (u - f)v = 0.$$

Damit reicht es, wenn u einmal differenzierbar ist. Allerdings brauchen wir gar nicht, dass u stetig differenzerbar ist, sonder nur dass

$$\int_\Omega \nabla u \cdot \nabla v$$

existiert. Daher definieren wir

Definition 2.1. *Sei $\Omega \subset \mathbb{R}^n$ beschränkt, der Rand von Ω sei stückweise glatt. Dann ist $C_0^\infty(\Omega)$ die Menge aller beliebig oft differenzierbarer Funktionen, die in einer Umgebung des Randes von Ω Null sind.*
Wir nennen $w_i = \partial_i u$ schwache Ableitung *bezüglich x_i von u auf Ω, falls für alle $\varphi \in C_0^\infty$ gilt, dass*

$$\int_\Omega w_i \varphi = -\int_\Omega u \partial_i \varphi \,.$$

Bemerkung 2.2.

1. *Schreiben wir $w := (w_1, \dots, w_n)$, so erhalten wir*

$$\int_\Omega w\varphi = -\int_\Omega u \nabla \varphi \qquad \forall \varphi \in C_0^\infty(\Omega)$$

oder

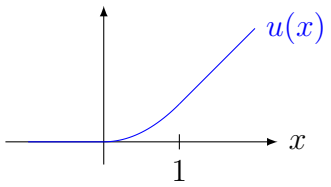
$$\int_\Omega w \cdot \Phi = -\int_\Omega u(\mathrm{div}\Phi) \qquad \forall \Phi : \Omega \to \mathbb{R}^n \text{ mit } \Phi_i \in C_0^\infty(\Omega).$$

2. *Falls u differenzierbar ist, so ist stimmen die übliche („starke“) Ableitung und die schwache Ableitung überein: mit $w := \nabla u$ erhalten wir*

$$\int_\Omega w \cdot \Phi = \int_\Omega \nabla u \cdot \Phi = -\int_\Omega u(\mathrm{div}\Phi) + \int_{\partial\Omega} u \underbrace{\Phi \cdot \nu}_{=0} = -\int_\Omega u(\mathrm{div}\Phi).$$

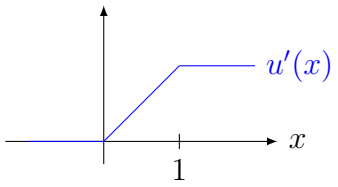
Beispiel 2.3. *Sei $u : (-1, 2) \to \mathbb{R}$ gegeben durch*

$$u(x) = \begin{cases} 0 & -1 \le x \le 0 \\ \frac{1}{2}x^2 & 0 \le x \le 1 \\ x - \frac{1}{2} & 1 \le x \le 2 \end{cases}$$



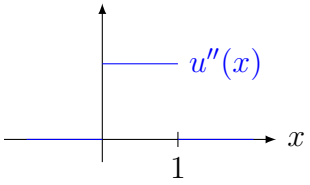
Dann ist die schwache Ableitung von u gegeben durch

$$u'(x) = \begin{cases} 0 & -1 \le x \le 0 \\ x & 0 \le x \le 1 \\ 1 & 1 \le x \le 2 \end{cases}$$



Die schwache Ableitung von u' ist analog

$$u''(x) = \begin{cases} 0 & -1 \le x \le 0 \\ 1 & 0 < x < 1 \\ 0 & 1 \le x \le 2 \end{cases}$$



u'' besitzt keine schwache Ableitung.

Allgemein ist für eine stetige Funktion $u : [a, c] \to \mathbb{R}$, die auf den Intervallen (a, b) und (b, c) differenzierbar ist (im Punkt b jedoch nicht), die schwache Ableitung gegeben durch

$$w(x) = \begin{cases} u'(x) & x \in (a, b) \\ 0 & x = b \\ u'(x) & x \in (b, c) \end{cases}$$

In der Tat ist

$$\begin{aligned} \int_a^c w\varphi &= \int_a^b w\varphi + \int_b^c w\varphi \\ &= \int_a^b u'\varphi + \int_b^c u'\varphi \\ &= -\int_a^b u\varphi' - \int_b^c w\varphi' + u\varphi|_a^b + u\varphi|_b^c \\ &= -\int_a^c u\varphi' + u(c)\underbrace{\varphi(c)}_{=0}\underbrace{-u(b)\varphi(b) + u(b)\varphi(b)}_{=0} - u(a)\underbrace{\varphi(a)}_{=0} \\ &= -\int_a^c u\varphi'. \end{aligned}$$

Dabei ist $u(b)$ einmal als linksseitiger Grenzwert, einmal als rechtsseitiger zu verstehen, die bieden Werte sind also nur gleich, wenn u stetig ist.
Also ist w schwache Ableitung von u, obwohl u im Punkt b nicht differenzierbar ist. Wir wir gesehen haben, „sieht“ das Integral einzelne Punkte nicht, denn

$$\int_a^c u = \int_a^b u + \int_b^c u$$

und somit ist die Festlegung $w(b) = 0$ beliebig und eigentlich auch überflüssig. Wie in obigem Beispiel schon geschehen, werden wir die schwache Ableitung in einzelnen Punkten nicht festlegen.
Wir wollen jetzt noch zeigen, dass die Funktion u aus Beispiel 2.3 schwache Lösung von

$$-\Delta u = f \quad \text{in } (-1, 2) \qquad \text{mit } f(x) = \begin{cases} -1 & x \in (0, 1) \\ 0 & \text{sonst} \end{cases}$$

ist. Dazu ist zu zeigen, dass

$$\int_\Omega \nabla u \cdot \nabla\varphi = \int_\Omega f\varphi \qquad \forall\varphi$$

gilt. Setzen wir die schwache Ableitung von u ein, erhalten wir

$$\int_{-1}^2 u'\varphi' = \int_{-1}^0 0\varphi' + \int_0^1 x\varphi' + \int_1^2 1\varphi'$$

und mittels partieller Integration im zweiten Term

$$
\begin{aligned}
&= -\int_0^1 \varphi + [x\varphi]_0^1 + [\varphi]_1^2 \\
&= -\int_0^1 \varphi + 1\varphi(1) - 0\varphi(0) + \underbrace{\varphi(2)}_{=0} - \varphi(1) = \int_0^1 (-1)\cdot\varphi \\
&= \int_{-1}^2 f\varphi.
\end{aligned}
$$

Definition 2.4 (Sobolevraum)**.** *Sei $\Omega \subset \mathbb{R}^n$ offen mit einem stückweise glatten Rand. Wir bezeichnen mit $H^1(\Omega)$ die Menge aller Funktionen u aus $L^2(\Omega)$, die in alle Richtungen schwach differenzierbar sind deren schwache Ableitung in $L^2(\Omega)$ liegt (also $\int_\Omega (\partial_{x_i} u)^2 < \infty$ für alle $i = 1, \ldots, N$).*
Wir definieren eine Norm auf $H^1(\Omega)$ durch

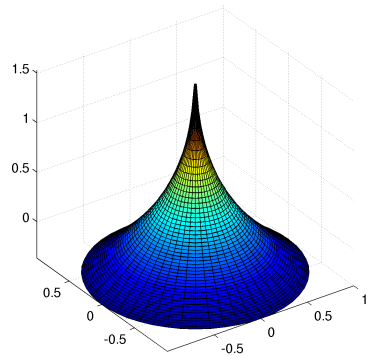
$$
\begin{aligned}
\|u\|^2_{H^1(\Omega)} &:= \|u\|^2_{L^2(\Omega)} + \|\nabla u\|^2_{L^2(\Omega)} = \|u\|^2_{L^2(\Omega)} + \sum_{i=1}^n \|\partial_{x_i} u\|^2_{L^2(\Omega)} \\
&= \int_\Omega \left(u^2 + \sum_{i=1}^n (\partial_{x_i} u)^2 \right).
\end{aligned}
$$

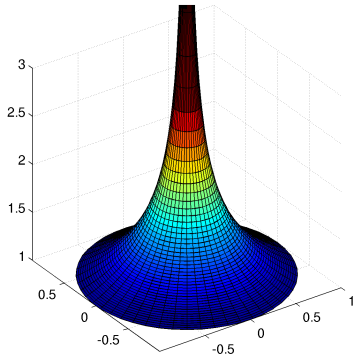
Weiter sei $H^1_0(\Omega)$ die Menge der Funktionen $u \in H^1(\Omega)$, die beliebig gut in $C^\infty_0(\Omega)$ approximiert werden können, d.h. jedem $\varepsilon > 0$ gibt es ein $u_\varepsilon \in C^\infty_0(\Omega)$ mit $\|u - u_\varepsilon\|_{1,\Omega} < \varepsilon$.
Die Mengen $H^1(\Omega)$ und $H^1_0(\Omega)$ sind Vektorräume.

Notation 2.5. *Wir schreiben kurz $\|u\|_{0,\Omega} := \|u\|_{L^2(\Omega)}$ (Null-mal schwach differenzierbar) und $\|u\|_{1,\Omega} := \|u\|_{H^1(\Omega)}$ (Ein-mal schwach differenzierbar).*

Wir schauen uns nun einige Beispiele für Funktionen aus $H^1(\Omega)$ an. Das wichtigste Beispiel, nämlich stetige, stückweise differenzierbare Funktionen, haben wir schon gesehen. In 1d lassen alle Funktionen aus $H^1(\Omega)$ stetig ergänzen, indem man sie ggf. auf einer Ausnahmemenge A, die vom Integral „nicht gesehen“ wird ($\int_{\Omega\setminus A} 1 = \int_\Omega 1$), umdefiniert.

In höheren Dimensionen müssen Funktionen aus $H^1(\Omega)$ nicht stetig ergänzbar sein. Betrachten wir z.B. die beiden Funktionen





$$f(x) = \log\left(\log\left(1 + \frac{1}{|x|}\right)\right), \qquad g(x) = \frac{1}{\sqrt{|x|}},$$

die im Nullpunkt eine Singularität besitzen, sich also dort nicht stetig ergänzen lassen. Für beide Funktionen kann man zeigen, dass sie schwache Ableitungen besitzen (die außerhalb des Ursprungs ihren klassichen Ableitungen entsprechen), aber nur die schwache Ableitung von f liegt in $L^2(B_1(0))$. Damit ist $f \in H^1(B_1(0))$ und $g \notin H^1(B_1(0))$. Der Unterschied zwischen beiden Funktionen ist, wie schnell sie gegen Unendlich gehen.
Schließlich sei noch angemerkt, dass Funktionen mit einem Sprung nicht in $H^1(\Omega)$ liegen.
Für uns wird im weiteren Verlauf nur das Beispiel der stetigen, stückweise differenzierbaren Funktion relevant sein.

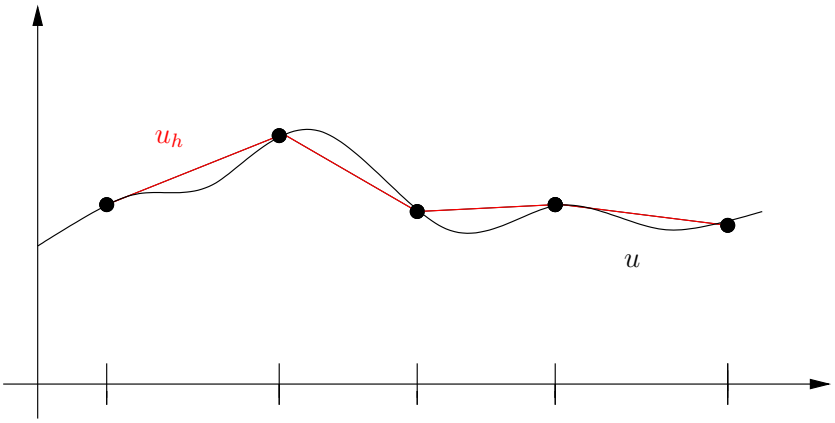
Mit diesen Definitionen können wir nun Problemstellung 1.19 wie folgt verallgemeinern.

Problemstellung 2.6 (Schwache Lösung)**.** *Sei $\Omega \subset \mathbb{R}^n$ beschränkt mit einem glatten Rand. Sei $f \in L^2(\Omega)$. Dann heißt $u \in H_0^1(\Omega)$* schwache Lösung *von $-\beta\Delta u + u = f$ mit Nullrandwerten, falls für alle Testfunktionen $v \in H_0^1(\Omega)$ gilt, dass*

$$\int_\Omega \beta \nabla u \cdot \nabla v + (u - f)v = 0\,. \tag{P^w}$$

2.2 Approximation durch Finite Elemente

Wir kommen nun zur Approximation von schwachen Lösungen. Zur besseren Veranschaulichung betrachten wir zunächst den eindimensionalen Fall ($n = 1$). Der Definitionsbereich einer Funktion u wird in Intervalle unterteilt, auf denen wir u durch affine Funktionen approximieren. Es ergibt sich eine stückweise affine, stetige Funktion u_h:

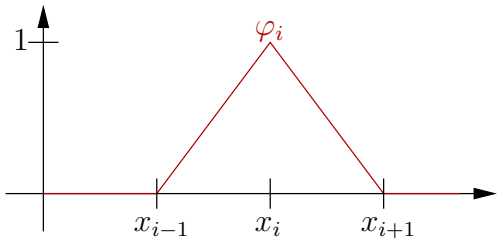


Wie können wir u_h einfach darstellen?

Definition 2.7 (Hütchenbasis in 1D)**.** *Sei* $0 = x_1 < x_2 < \cdots < x_N = L$*. Dann bezeichnet* $\varphi_i(x)$ *diejenige stetige, stückweise affine Funktion, für die*

$$\varphi_i(x_j) := \delta_{ij} = \begin{cases} 1 & \text{falls } i = j\,, \\ 0 & \text{sonst.} \end{cases}$$

gilt.



Satz 2.8. *Jede stetige, in Bezug auf die Gitterpunkte* $x_1 < \ldots < x_N$ *stückweise affine Funktion* u_h *kann als Linearkombination aus Hütchenfunktionen dargestellt werden:*

$$u_h(x) = \sum_{i=1}^{N} U_i \varphi_i(x),$$

wobei $U_i = u_h(x_i)$ *zu setzen ist.*
Die Funktionen $\{\varphi_i\}$ *bilden eine Basis des Vektorraums der stetigen stückweise affinen Funktionen. Diesen Raum bezeichnen wir mit* V_h *bzw.* V_h^0 *(mit Nullrandwerten, d.h. ohne* φ_1, φ_N*). Der Vektorraum* V_h *bzw.* V_h^0 *ist ein Untervektorraum von* $H^1([0, L])$ *bzw.* $H_0^1([0, L])$*, insbesondere sind Funktionen in* V_h *und* V_h^0 *schwach differenzierbar.*

Nun lässt sich das Problem (P^w) einfach approximieren: Wir ersetzen lediglich $H_0^1(\Omega)$ durch V_h^0!

Gesucht ist also $u_h \in V_h^0$, so dass

$$\beta \int_\Omega \nabla u_h \cdot \nabla v_h + \int_\Omega u_h v_h = \int_\Omega f_h v_h \qquad \forall v_h \in V_h^0 \,. \tag{2.1}$$

Die Gleichung (2.1) ist linear in Bezug auf v_h. Es genügt also, statt allen v_h nur $v_h = \varphi_i$ für $i = 2, \ldots, N-1$ zu betrachten, also

$$\beta \int_\Omega \nabla u_h \cdot \nabla \varphi_i + \int_\Omega u_h \varphi_i = \int_\Omega f_h \varphi_i \qquad i = 2, \ldots, N-1.$$

Nun können wir noch u_h und f_h in der Basis darstellen

$$u_h(x) = \sum_{j=2}^{N-1} U_j \varphi_j(x) \qquad \text{und} \qquad f_h(x) = \sum_{j=2}^{N-1} F_j \varphi_j(x).$$

Damit muss für alle Basisfunktionen φ_i, $i = 2, \ldots, N-1$ gelten:

$$\begin{aligned} &\int_\Omega \beta \nabla \left(\sum_{j=2}^{N-1} U_j \varphi_j \right) \cdot \nabla \varphi_i + \int_\Omega \left(\sum_{j=2}^{N-1} U_j \varphi_j \right) \varphi_i = \int_\Omega \left(\sum_{j=2}^{N-1} F_j \varphi_j \right) \varphi_i \\ \iff &\sum_{j=2}^{N-1} U_j \underbrace{\beta \int_\Omega \nabla \varphi_i \cdot \nabla \varphi_j}_{=: L_{ij}} + \sum_{j=2}^{N-1} U_j \underbrace{\int_\Omega \varphi_i \varphi_j}_{=: M_{ij}} = \sum_{j=2}^{N-1} F_j \int_\Omega \varphi_j \varphi_i \end{aligned}$$

Wir nennen die Matrix M *Massematrix* und die Matrix L *Steifigkeitsmatrix*.
Daraus ergibt sich die Matrix-Vektor-Formulierung des Gleichungssystems:

$$\boxed{(L + M)U = MF}$$

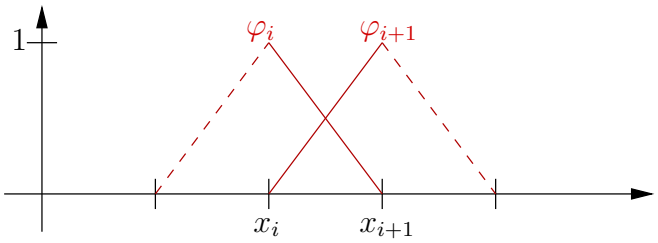
Bemerkung 2.9. *Wir mussten uns bei der Finite-Elemente-Diskretisierung keine Gedanken über die Art der Approximation des Gradienten machen. Nach Wahl der Approximationsfunktionen sind die Matrizen durch die Integrale für M_{ij} und L_{ij} gegeben, dabei können die schwachen Ableitungen der Basisfunktionen* exakt *berechnet werden. Die einzige Approximation findet durch die Wahl des Ansatzraumes V_h bzw. V_h^0 statt.*

Berechnung der Matrizen M und L

Wir berechnen die Matrizen in 1d auf einem äquidistenten Gitter, also $x_i = (i-1)h$.
Für die Hütchenbasis gilt $\varphi_i(x_j) = \delta_{ij} = \begin{cases} 1 & \text{falls } i = j, \\ 0 & \text{sonst.} \end{cases}$
Dazwischen wird affin interpoliert.
Es ergeben sich damit folgende Definitionen für die Funktionen und ihre Ableitungen:

$$\varphi_i(x) = \begin{cases} 0 & \text{falls } x \le x_{i-1} \text{ oder } x \ge x_{i+1}, \\ \dfrac{x - x_{i-1}}{h} & \text{falls } x \in (x_{i-1}, x_i], \\ \dfrac{x_{i+1} - x}{h} & \text{falls } x \in (x_i, x_{i+1}), \end{cases}$$

$$\varphi_i'(x) = \begin{cases} 0 & \text{falls } x \le x_{i-1} \text{ oder } x \ge x_{i+1}, \\ \dfrac{1}{h} & \text{falls } x \in (x_{i-1}, x_i], \\ -\dfrac{1}{h} & \text{falls } x \in (x_i, x_{i+1}). \end{cases}$$



Da sich nur benachbarte Hütchenfunktionen überlappen, sind die meisten Einträge der Matrizen Null. Die Nicht-Null-Einträge für die Massematrix sind

$$
\begin{aligned}
M_{ii} &= \int (\varphi_i)^2 = \frac{1}{h^2}\int_{x_{i-1}}^{x_i} (x - x_{i-1})^2\,dx + \frac{1}{h^2}\int_{x_i}^{x_{i+1}} (x_{i+1} - x)^2\,dx \\
&= \frac{1}{h^2}\left[\frac{1}{3}(x - x_{i-1})^3\right]_{x=x_{i-1}}^{x=x_i} - \frac{1}{h^2}\left[\frac{1}{3}(x_{i+1} - x)^3\right]_{x=x_i}^{x=x_{i+1}} \\
&= \frac{1}{h^2}\frac{1}{3}h^3 + \left(-\frac{1}{h^2}\left(-\frac{1}{3}h^3\right)\right) = \frac{2}{3}h \\
M_{i,i+1} &= \int \varphi_i\varphi_{i+1} = \frac{1}{h^2}\int_{x_i}^{x_{i+1}} \underbrace{(x_{i+1} - x)}_{v}\underbrace{(x - x_i)}_{u'}\,dx \\
&\overset{\text{part. Int.}}{=} -\frac{1}{h^2}\int_{x_i}^{x_{i+1}} (-1)\frac{1}{2}(x - x_i)^2\,dx + \underbrace{\frac{1}{h^2}\left[(x_{i+1} - x)\frac{1}{2}(x - x_i)^2\right]_{x=x_i}^{x=x_{i+1}}}_{=0} \\
&= \frac{1}{h^2}\int_{x_i}^{x_{i+1}} \frac{1}{2}(x - x_i)^2\,dx = \frac{1}{h^2}\left[\frac{1}{6}(x - x_i)^3\right]_{x=x_i}^{x=x_{i+1}} \\
&= \frac{1}{h^2}\frac{1}{6}h^3 = \frac{1}{6}h = M_{i,i-1}
\end{aligned}
$$

also hat die $(N-2)\times(N-2)$-Matrix M die Gestalt

$$
M = \frac{h}{6}\begin{pmatrix}
4 & 1 & & & 0 \\
1 & 4 & 1 & & \\
 & \ddots & \ddots & \ddots & \\
 & & 1 & 4 & 1 \\
0 & & & 1 & 4
\end{pmatrix}.
$$

Für die Steifigkeitsmatrix erhalten wir

$$
\begin{aligned}
L_{ii} &= \beta\int (\varphi_i')^2 = \beta\int_{x_{i-1}}^{x_i} \frac{1}{h^2} + \beta\int_{x_i}^{x_{i+1}} \left(-\frac{1}{h^2}\right) \\
&= \beta h\frac{1}{h^2} + \beta h\frac{1}{h^2} = \frac{2\beta}{h} \\
L_{i,i+1} &= \beta\int \varphi_i'\varphi_{i+1}' = \beta\int_{x_i}^{x_{i+1}} \left(-\frac{1}{h}\right)\left(\frac{1}{h}\right) \\
&= \beta h\left(-\frac{1}{h^2}\right) = -\frac{\beta}{h} = L_{i,i-1}
\end{aligned}
$$

somit hat die $(N-2)\times(N-2)$-Matrix L die Gestalt

$$L = \frac{\beta}{h}\begin{pmatrix} 2 & -1 & & & 0 \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ 0 & & & -1 & 2 \end{pmatrix}.$$

Bemerkung 2.10. *Multipliziert man die Steifigkeitsmatrix mit $\frac{1}{h}$, so erhalten wir genau die selbe Matrix, die wir auch bei der Finite-Differenzen-Diskretisierung des Laplace-Operators erhalten haben. Die Massematrix unterscheidet sich jedoch (auch nach Multiplikation mit $\frac{1}{h}$) von der Matrix, die an dieser Stelle bei Finiten Differenzen auftrat – der Einheitsmatrix.*

2.3 Randwerte

Bisher haben wir implizit Null-Randwerte durch Wahl der Räume $H_0^1(\Omega)$ bzw. V_h^0 vorgeschrieben, also

$$\begin{aligned} -\beta\Delta u + u &= f && \text{in } \Omega \\ u &= 0 && \text{auf } \partial\Omega \end{aligned}$$

Wir wollen aber (zunächst)

$$\begin{aligned} -\beta\Delta u + u &= f && \text{in } \Omega \\ u &= f && \text{auf } \partial\Omega \end{aligned}$$

Dazu führen wir das Problem auf ein Problem mit Nullrandwerten zurück: wir setzen $w := u - f$. Mit dieser Definition erhalten wir $w = 0$ auf $\partial\Omega$ und

$$\begin{aligned} -\beta\Delta w + w &= -\beta\Delta u + \beta\Delta f + u - f \\ &= \underbrace{-\beta\Delta u + u - f}_{=0} + \beta\Delta f. \end{aligned}$$

Wir erhalten also ein äquivalentes Problem mit Nullrandwerten:

$$\begin{aligned} -\beta\Delta w + w &= \beta\Delta f \quad \text{in } \Omega \\ w &= 0 \quad \text{auf } \partial\Omega. \end{aligned}$$

In der schwachen Formulierung erhalten wir (falls $f \in H^1(\Omega)$) aus

$$\int_\Omega \beta\nabla u\cdot\nabla v + (u-f)v = 0$$

mit $u = w + f$ die Gleichung

$$\int_\Omega \beta\nabla w\cdot\nabla v + wv + \beta\nabla f\cdot\nabla v = 0,$$

d.h. nach der Diskretisierung

$$L\,W + M\,W + L\,F = 0.$$

Die (Approximation an die) Funktion u erhält man nach Definition von w durch $u = w + f$.

Wir hatten aber gesehen, dass zumindest für die Bildglättung Nullrandwerte nicht unbedingt ideal sind. Daher betrachten wir wie in Abschnitt 1.4 die Randwerte

$$\nabla u \cdot \nu = 0.$$

Diese erhält man, indem man das Problem (P^w) mit $H^1(\Omega)$ statt $H_0^1(\Omega)$ betrachtet: Gesucht wird eine Funktion $u \in H^1(\Omega)$, so dass für alle $v \in H^1(\Omega)$

$$\int_\Omega \beta \nabla u \cdot \nabla v + (u - f)v = 0$$

gilt. Falls u zweimal differenzierbar ist, erhält man mit Hilfe des Satzes von Gauß

$$\int_\Omega -\beta \Delta u\, v + (u - f)v + \beta \int_{\partial\Omega} \nabla u \cdot \nu\, v = 0.$$

Verwendet man Testfunktionen v mit Nullrandwerten (diese sind zulässig da $H_0^1(\Omega) \subset H^1(\Omega)$), so ergibt sich wie bisher

$$-\beta \Delta u + u = f.$$

Daher ist

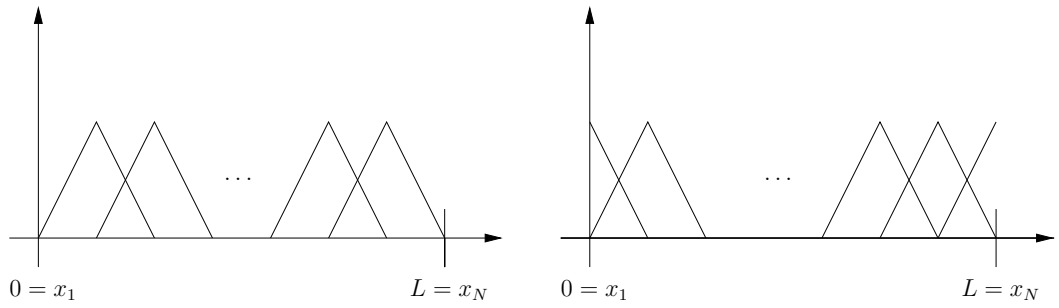
$$\int_{\partial\Omega} \nabla u \cdot \nu\, v = 0$$

für alle v, also

$$\nabla u \cdot \nu = 0.$$

Dies nennt man *natürliche Randwerte*. In der schwachen Formulierung unterscheiden sich natürliche von Null-Randwerten *nur* durch Verwendung von $H^1(\Omega)$ statt $H_0^1(\Omega)$. Analog verwendet man in der Diskretisierung V_h statt V_h^0.
Im eindimensionalen Fall betrachten wir die folgenden Räume:

$V_h^0 = \text{span}\{\varphi_i \,|\, i = 2, \ldots, N-1\}$ $\qquad$ $V_h = \text{span}\{\varphi_i \,|\, i = 1, \ldots, N\}$



Die rechte Abbildung enthält zusätzlich die „abgeschnittenen“ Basisfunktionen zu den Randknoten.

Die entstehenden Masse- und Steifigkeitsmatrizen sind $N \times N$-Matrizen. Die Struktur entspricht den Matrizen für Nullrandwerte bis auf den linken oberen und den rechten unteren Eintrag. Bei den Integralen $\int_\Omega \varphi_1\varphi_1$, $\int_\Omega \nabla\varphi_1 \cdot \nabla\varphi_1$ und den analogen Integralen von φ_N wird jeweils nur über „halbe“ Basisfunktionen integriert, so dass sich dort auch jeweils der halbe Wert ergibt. Wir erhalten also

$$M = \frac{h}{6}\begin{pmatrix} 2 & 1 & & & 0 \\ 1 & 4 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & 4 & 1 \\ 0 & & & 1 & 2 \end{pmatrix}, \qquad L = \frac{\beta}{h}\begin{pmatrix} 1 & -1 & & & 0 \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & -1 \\ 0 & & & -1 & 1 \end{pmatrix}.$$

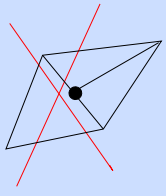
2.4 Finite Elemente auf Dreiecksgittern

Als Beispiel wollen wir die Generalisierung von DHM (digitalen Höhenmodellen) betrachten. Hier besteht die Aufgabe darin, die grobskaligen bzw. die „wichtigen“ Geländeeigenschaften zu finden. Dazu betrachten wir Dreiecksgitter in den Ebenen (z.B. Gauß–Krüger-Koordinaten) und der Funktionswert u stellt die Höhe des Gitterpunktes dar. Dies führt auf die bekannte partielle Differentialgleichung

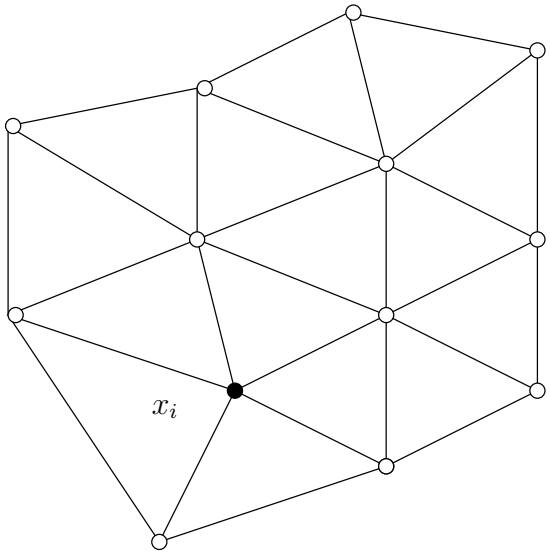
$$\begin{aligned} -\beta\Delta u + u &= f && \text{in } \Omega, \\ \nabla u \cdot \nu &= 0 && \text{auf } \partial\Omega. \end{aligned}$$

Definition 2.11. *Eine* Triangulierung *eines polygonalen Gebietes Ω_h ist eine Unterteilung von Ω_h in Dreiecke, wobei gilt*

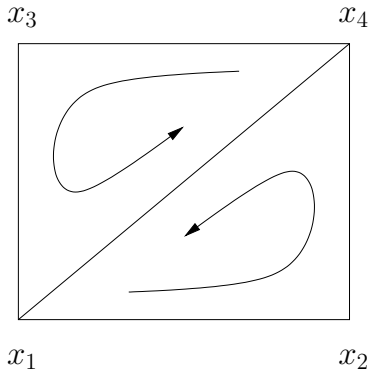
- *ganz Ω_h wird überdeckt*
- *zwei Dreiecke schneiden sich entweder garnicht, in einem Punkt, oder einer kompletten Kante, es gibt also keine „hängenden“ Knoten*
- *die Winkel sind nicht „zu spitz“*



Eine derartige Triangulierung hat beispielsweise die Form:



Um eine Triangulierung im Rechner zu implementieren, speichert man ihre Punkte und Dreiecke ab. Nehmen wir als Beispiel das einfache Gitter



Liste von Punkten:	Liste von Dreiecken:
$x_1 = (0,0)$	$T_1 = (x_1, x_2, x_4)$
$x_2 = (1,0)$	$T_2 = (x_4, x_3, x_1)$
$x_3 = (0,1)$	
$x_4 = (1,1)$	

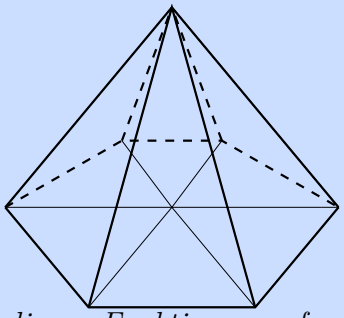
Dabei ist es sinnvoll, eine Konvention für die Orientierung der Dreiecke zu haben, z.B. gegen den Uhrzeigersinn.
In eine Datei schreib man üblicherweise zunächst die Anzahl der Punkte und die Anzahl der Dreiecke und listet diese dann auf:

```
4 2
0 0
1 0
0 1
1 1
1 2 4
4 3 1
```

Auch in 2d definieren wir wieder eine Basis für stetige, sückweise affine Funktionen:

Definition 2.12. *Zu einer Triangulierung wie oben bezeichnet φ_i die stetige stückweise affine Funktion, für die am Knoten x_j gilt:*

$$\varphi_i(x_j) = \delta_{ij} = \begin{cases} 1 & \text{für } i = j \\ 0 & \text{sonst.} \end{cases}$$



Analog zum 1d-Fall bezeichnen wir mit V_h den von diesen Funktionen aufgespannten Vektorraum und setzen $V_h^0 := \operatorname{span}\left\{\varphi_i(x) \;\middle|\; x_i \text{ liegt im Inneren von } \Omega_h\right\}$.

Wie in 1d gilt

$$u_h(x) = \sum_{i=1}^{N} U_i \varphi_i(x).$$

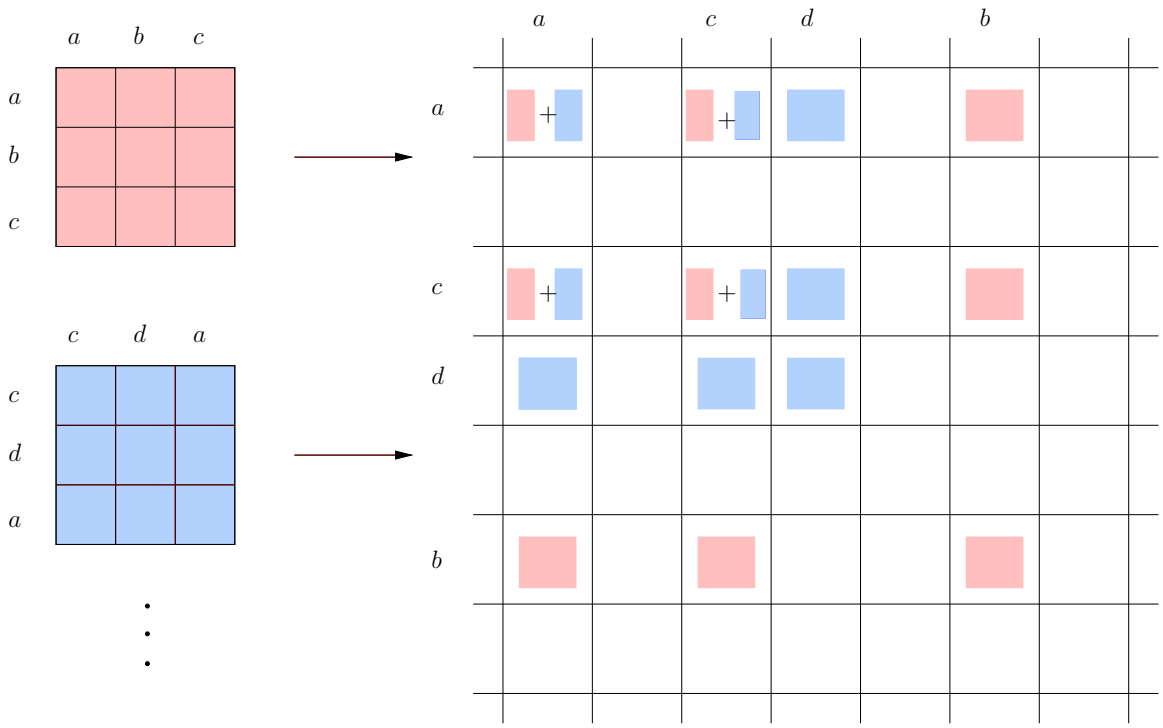
2.5 Assemblierung der Matrizen

Um den (i, j)-ten Eintrag der Massematrix (und analog der Steifigkeitsmatrix) zu ermitteln, berechnen wir $\int_\Omega \varphi_i \varphi_j$ nicht direkt für das gesamte Ω, sondern werten das Integral

$$\int_T \varphi_i \varphi_j$$

für jedes einzelne Dreieck T und für alle φ_i, φ_j, die auf diesem Dreieck nicht Null sind, aus.

Hierbei wissen wir genau, welche φ_i auf dem Dreieck T nicht Null sind: diejenigen, die zu einer Ecke des Dreiecks gehören. Für $T = (x_a, x_b, x_c)$ sind wären das also $i, j \in \{a, b, c\}$. Dies führt zu 3×3-Kombinationen, die auf die entsprechenden Stellen der Matrix verteilt werden, dabei addiert man die Beiträge aus allen beteiligten Dreiecken auf.



Zunächst ein Wort zur Notation: Um die Menge an Bezeichnungen übersichtlich zu halten, sind wir im Folgenden etwas ungenau. Wir unterscheiden nicht explizit zwischen

- lokalen $(1, 2, 3)$ und globalen $(1, \ldots, N)$ Kantennummern,
- der Punktemenge, aus der eine Seite besteht und dem zugehörigen Kantenvektor.

Die Bedeutung ergibt sich jeweils aus dem Kontext.

Steifigkeitsmatrix

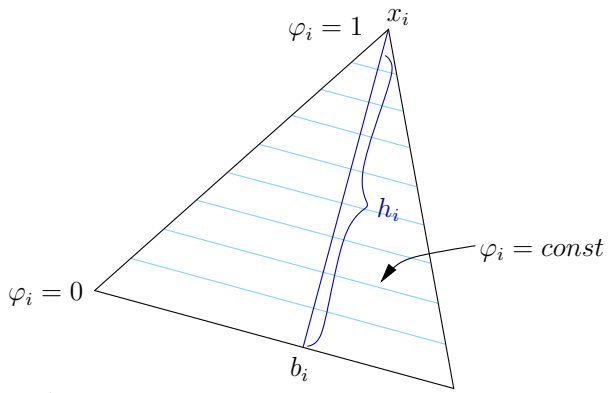
Die Steifigkeitamatrix ist einfacher, denn $\nabla\varphi_i$ und $\nabla\varphi_j$ sind konstante Vektoren auf T, also

$$\int_T \nabla\varphi_i \cdot \nabla\varphi_j = |T| \, \nabla\varphi_i \cdot \nabla\varphi_j.$$

Zur Berechnung dieses Skalarprodukts wollen wir die Formel

$$v \cdot w = |v||w| \cos(\sphericalangle(v, w))$$

verwenden. Hierfür müssen wir die Beträge der Gradienten berechnen:

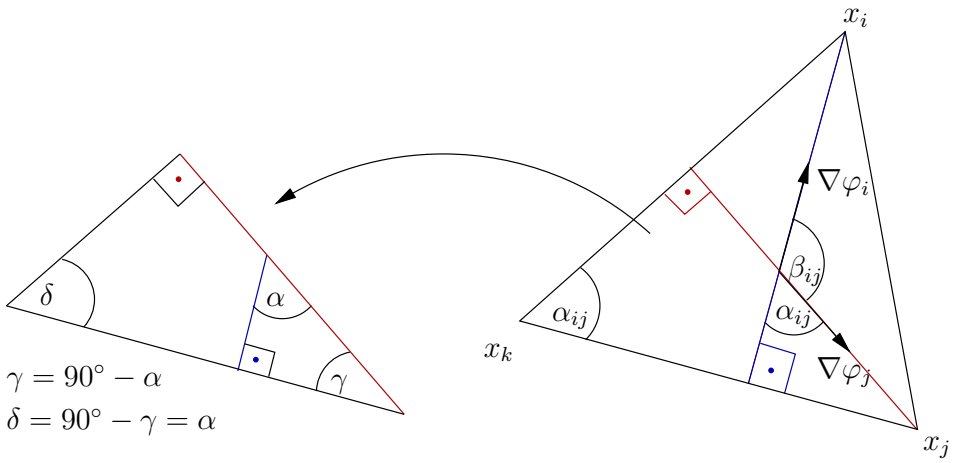


$$|\nabla\varphi_i| = \frac{|\varphi_i(x_i) - \varphi_i(b_i)|}{|x_i - b_i|} = \frac{1}{h_i}$$

Daher

$$\int_T |\nabla\varphi_i|^2 = |T| \frac{1}{h_i^2},$$
$$\int_T \nabla\varphi_i \cdot \nabla\varphi_j = |T| \frac{1}{h_i}\frac{1}{h_j}\cos(\beta_{ij}).$$

Um $\cos\beta_{ij}$ zu berechnen, betrachten wir



so dass wir $\cos(\beta_{ij}) = \cos(180° - \alpha_{ij}) = -\cos(\alpha_{ij})$ erhalten. Letzteres können wir nun einfach aus den Seitenvektoren bestimmen. Wir bezeichnen mit $s_i := x_{i+2} - x_{i+1}$ (und der Konvention, dass die Punkte entgegen des Uhrzeigersinns durchnummeriert sind, Indizes jeweils modulo 3) die Seite gegenüber von x_i. Damit ist

$$-s_i \cdot s_j = |s_i|\,|s_j|\cos(\sphericalangle(-s_i, s_j)) = |s_i|\,|s_j|\cos(\alpha_{ij}) \Longrightarrow \cos(\alpha_{ij}) = \frac{-s_i \cdot s_j}{|s_i|\,|s_j|}.$$

Wir brauchen nun noch die Fläche des Dreiecks und die Länge der Höhe h_i. Nach dem Satz des Heron erhält man mit dem *Semiperimeter* $s = (|s_1| + |s_2| + |s_3|)/2$ die Fläche des Dreiecks durch

$$|T| = \sqrt{s(s - |s_1|)(s - |s_2|)(s - |s_3|)}.$$

Damit kann man dann schließlich durch „Fläche gleich Grundseite mal Höhe“ die Höhe berechnen:

$$|T| = \tfrac{1}{2}\,|s_i|\,h_i \iff h_i = \frac{2\,|T|}{|s_i|}.$$

Damit erhalten wir

$$\int_T |\nabla \varphi_i|^2 = \frac{|s_i|^2}{4\,|T|},$$
$$\int_T \nabla \varphi_i \cdot \nabla \varphi_j = \frac{s_i \cdot s_j}{4\,|T|}.$$

Massematrix

Die Einträge der Massematrix berechnen wir durch eine numerische Quadraturformel, die auf den vorkommenden Funktionen den exakten Wert des Integrals liefert.

Lemma 2.13. *Sei T ein Dreieck mit den Seitenmitten a, b, c und p sei quadratisches Polynom. Dann gilt*

$$\int_T p(x)\,\mathrm{d}x = \frac{|T|}{3}(p(a) + p(b) + p(c)) \tag{2.2}$$

Beweis. Zunächst zeigen wir das Lemma für das Einheitsdreieck $\hat{T}$. Es reicht, die Aussage für die Basis $\{1, x, y, xy, x^2, y^2\}$ der quadratischen Polynome zu zeigen. Aus Symmetriegründen genügt es, die Funktionen $\{1, x, xy, x^2\}$ zu betrachten. Es gilt

- $p(x, y) = 1$

$$\int_T 1\,\mathrm{d}x = \tfrac{1}{2}, \quad \tfrac{\frac{1}{2}}{3}(1+1+1) = \tfrac{1}{2} \qquad \surd$$

- $p(x, y) = x$

$$\begin{aligned}\int_T x\,dx &= \int_0^1 \int_0^{1-y} x\,dx\,dy = \int_0^1 \frac{1}{2}(1-y)^2\,dy = \frac{1}{2}\int_0^1 1 - 2y + y^2\,dy \\ &= \frac{1}{2}\left[1 - y^2 + \frac{1}{3}y^3\right]_0^1 = \frac{1}{2}(1 - 1 + \frac{1}{3}) = \frac{1}{6},\end{aligned}$$
$$\tfrac{\frac{1}{2}}{3}(0 + \frac{1}{2} + \frac{1}{2}) = \frac{1}{6} \qquad \surd$$

- $p(x, y) = xy$

$$\begin{aligned}\int_T xy\,dx &= \int_0^1 \int_0^{1-y} xy\,dx\,dy = \int_0^1 \frac{1}{2}(1-y)^2 y\,dy = \frac{1}{2}\int_0^1 y - 2y^2 + y^3\,dy \\ &= \frac{1}{2}\left[\frac{1}{2}y^2 - \frac{2}{3}y^3 + \frac{1}{4}y^5\right]_0^1 = \frac{1}{2}(\frac{1}{2} - \frac{2}{3} + \frac{1}{4}) = \frac{1}{2}\cdot\frac{6-8+3}{12} = \frac{1}{24},\end{aligned}$$
$$\tfrac{\frac{1}{2}}{3}(0 + 0 + \frac{1}{4}) = \frac{1}{24} \qquad \surd$$

- $p(x,y) = x^2$

$$\begin{aligned}\int_T x^2\,dx &= \int_0^1 \int_0^{1-y} x^2\,dx\,dy = \int_0^1 \frac{1}{3}(1-y)^3\,dy \\ &= \frac{1}{3}\int_0^1 1 - 3y + 3y^2 - y^3\,dy = \frac{1}{3}\left[y - \frac{3}{2}y^2 + y^3 - \frac{1}{4}y^4\right]_0^1 \\ &= \frac{1}{3}(1 - \frac{3}{2} + 1 - \frac{1}{4}) = \frac{1}{3}\cdot\frac{8-12+8-2}{8} = \frac{1}{12},\end{aligned}$$

$$\tfrac{\frac{1}{2}}{3}(0 + \frac{1}{4} + \frac{1}{4}) = \frac{1}{12} \qquad \surd$$

Wir haben also gezeigt, dass

$$\int_{\hat{T}} p(x)\ \mathrm{d}x = \frac{|\hat{T}|}{3}(p(a) + p(b) + p(c)).$$

Wir können ein beliebiges Dreieck T mit einer affinen Abbildung auf das Einheitsdreieck $\hat{T}$ abbilden. Aufgrund des Transformationssatzes wissen wir, dass das Integral dabei genau wie die rechte Seite mit der Flächenänderung skaliert. Da eine affine Abbildung Seitenmitten auf Seitenmitten abbildet und p nach der Transformation ein quadratisches Polynom bleibt, gilt die Formel auch auf T. □

Wie sehen nun die Einträge der (lokalen) Massenmatrix aus? Die Basisfunktionen hat in der Mitte einer Seite den Wert 0, in den beiden anderen der Wert $\frac{1}{2}$. Mit Lemma 2.13 erhalten wir

$$\begin{aligned}\int_T \varphi_i\varphi_i &= \frac{|T|}{6}, \\ \int_T \varphi_i\varphi_j &= \frac{|T|}{12} \quad \text{für } i \neq j\end{aligned}$$

Für die lokale Massenmatrix erhalten wir also

$$M_T = \frac{|T|}{12}\begin{pmatrix}2 & 1 & 1\\ 1 & 2 & 1\\ 1 & 1 & 2\end{pmatrix}.$$

Bemerkung 2.14 (Mass lumping)**.** *Einfacher, aber ungenauer ist die Quadraturformel*

$$\int_T p(x) \approx \frac{|T|}{3}\left(p(x_1) + p(x_2) + p(x_3)\right),$$

wobei x_i die Ecken des Dreiecks sind. Sie ist nur für affine p exakt. Der Vorteil ist, dass wegen $\varphi_i(x_j) = \delta_{ij}$

$$\begin{aligned}\int_T \varphi_i\varphi_i &\approx \frac{|T|}{3}(1+0+0) \\ \int_T \varphi_i\varphi_j &\approx \frac{|T|}{3}(0+0+0) \quad \textit{falls } i \neq j.\end{aligned}$$

Damit ist die lokale Massematrix

$$M_T = \frac{|T|}{3}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

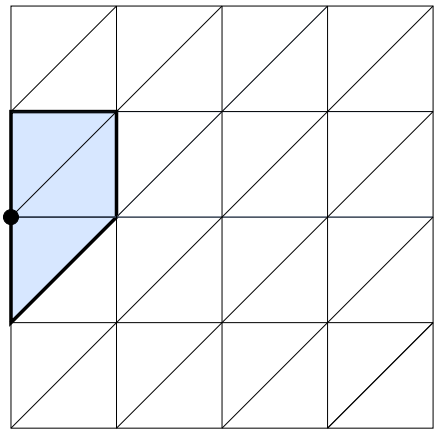
und so ist auch die Massematrix M eine Diagonalmatrix (und damit sehr effizient in der Handhabung). Man bezeichnet diesen Ansatz als „lumped masses“.

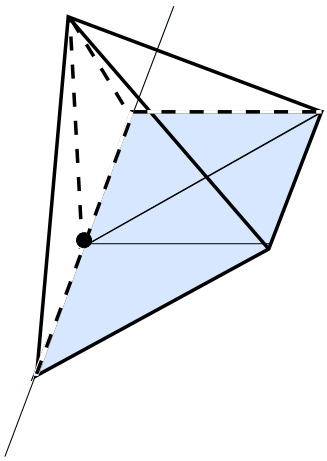
Assemblierung von Matrizen in `MATLAB`
In `MATLAB` kann man das eigentliche Assemblieren dem Aufruf `L = sparse ( I, J, V );` überlassen, wenn man die Vektoren `I`, `J` und `V` nacheinander mit den 9 Beiträgen jedes einzelnen Dreiecks füllt. Beim Aufruf von `sparse` werden dann alle Werte, die zu mehrfach auftretenden Indexpaaren (i, j) gehören, automatisch aufaddiert.

2.6 Randwerte

Natürliche Randwerte

Wie schon in 1d wird für Randknoten nur der innere Teil der Hütchenfunktion betrachtet. Wenn man die Matrix wie oben beschrieben dreicksweise assembliert, geschieht dies automatisch: Es werden für alle Dreiecke im Gebiet die auf dem jeweiligen Dreieck lebenden Teile der Basisfunktionen integriert.





Null-Randwerte

Um Nullrandwerte zu erhalten, betrachten wir wieder den Raum V_h^0, d.h. nur diejenigen Ansatzfunktionen, die zu inneren Knoten gehören. Unbekannte sind also nur diejenigen Knoten, die nicht auf dem Rand liegen. Beim Assemblieren beider Matrizen auf einem Dreieck verwendet man nur die Nicht-Rand-Knoten.
Schwierigkeit: die Nummern der Randknoten sind über die Liste der Knoten verteilt. Um dies zu umgehen, gibt es zwei Möglichkeiten.

- eine Umnummerierung der Knoten

- Zu Randknoten werden nur Beiträge auf die entsprechende Diagonale in der Matrix addiert, alle anderen Assemblierungsbeiträge werden ignoriert. Ebenso werden keine Beiträge zur rechten Seite assembliert. Dies führt dazu, dass für einen Randknoten i alle Einträge der i-ten Zeile und i-ten Spalte 0 sind, ausgenommen dem Diagonalelement (welches z.B. 1 ist). Im Vektor f auf der rechten Seite ist der zugehörigen Eintrag ebenfalls 0. (Statt „1“ ist auf der Diagonalen jeder nicht-Null-Eintrag möglich.)

$$
\begin{array}{r}
\\
i \rightarrow
\end{array}
\begin{pmatrix}
 & & & \overset{i}{\downarrow} & & & \\
 & & & 0 & & & \\
 & & & | & & & \\
 & & & 0 & & & \\
0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\
 & & & 0 & & & \\
 & & & | & & & \\
 & & & 0 & & &
\end{pmatrix}
$$

  Dies ergibt dann $u_i = 0$ ohne jede Kopplung an andere Freiheitsgrade.

  Alternativ: Eine normale Assemblierung (für natürliche Randwerte) durchführen, anschließend für Randknoten i die Zeile i und Spalte i auf Null, den Diagonaleintrag auf 1 setzen (ergibt dieselbe Matrix) und den Eintrag i der rechten Seite ebenfalls Null setzen. Wir haben allerdings schon bemerkt, dass das Nullsetzen von Zeilen langsam sein kann.

Nicht-Null-Randwerte

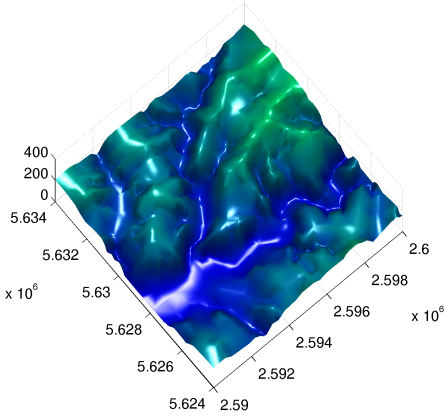
Genau wie in Abschnitt 2.3 wird dieser Fall auf den Fall von Nullrandwerten zurück geführt.

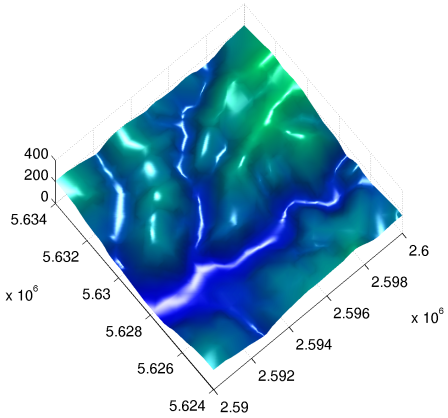
Programmieraufgabe 4. *Implementieren Sie das beschriebene Finite Elemente Verfahren zur Berechnung von*

$$
\begin{aligned}
-\beta\Delta u + u &= f && \text{in } \Omega \\
\nabla u \cdot \nu &= 0 && \text{auf } \partial\Omega.
\end{aligned}
$$

zur Generalisierung digitaler Höhenmodelle auf Dreiecksgittern. Achten Sie auf eine effiziente Assemblierung der Matrix.

Als Ergebnis für $\beta = 5 \cdot 10^4$ (dieser Wert ist sicherlich zu groß, zerstört also zu viele Details) erhalten wir





2.7 Ein wenig Konvergenztheorie

Wir brauchen in diesem Abschnitt mehrfach den

Satz 2.15 (Satz von Hölder)**.** *Seien $u, v \in L^2(\Omega)$. Dann ist*

$$\int_\Omega |uv| \leq \|u\|_{L^2(\Omega)} \|v\|_{L^2(\Omega)} .$$

Beweis. Für den Beweis verwenden wir die Ungleichung

$$(a-b)^2 \geq 0 \iff a^2 + b^2 - 2ab \geq 0 \iff ab \leq \tfrac{1}{2}a^2 + \tfrac{1}{2}b^2.$$

Damit ist

$$\begin{aligned}
\frac{1}{\|u\|_{L^2(\Omega)} \|v\|_{L^2(\Omega)}} \int_\Omega |uv| &= \int_\Omega \frac{|u|}{\|u\|_{L^2(\Omega)}} \frac{|v|}{\|v\|_{L^2(\Omega)}} \\
&\leq \int_\Omega \frac{1}{2} \frac{u^2}{\|u\|^2_{L^2(\Omega)}} + \frac{1}{2} \frac{v^2}{\|v\|^2_{L^2(\Omega)}} \\
&= \frac{1}{2\|u\|^2_{L^2(\Omega)}} \int_\Omega u^2 + \frac{1}{2\|v\|^2_{L^2(\Omega)}} \int_\Omega v^2 \\
&= \tfrac{1}{2} + \tfrac{1}{2} = 1.
\end{aligned}$$

□

Da wir die rechte Seite f durch eine Funktion f_h approximiert haben, ist es wichtig zu wissen, dass kleine Änderungen der rechten Seite nur kleine Änderungen der Lösung nach sich ziehen.

Satz 2.16 (Stabilität)**.** *Sei u_h die diskrete Lösung von*

$$\begin{aligned} -\beta\Delta u + u &= f && \text{in } \Omega, \\ u &= f && \text{auf } \partial\Omega. \end{aligned}$$

Dann gibt es $C > 0$, so dass gilt

$$\|u_h\|_{H^1(\Omega)} \leq C \|f_h\|_{H^1(\Omega)}.$$

Beweis. Wir können den Beweis auf den Fall von Nullrandwerten zurückführen. Mit $w_h = u_h - f_h$ war w_h Lösung des Problems mit Nullrandwerten (und anderer rechter Seite). Wenn wir zeigen können, dass für w_h

$$\|w_h\|_{H^1(\Omega)} \leq C_1 \|f_h\|_{H^1(\Omega)} \tag{2.3}$$

gilt, dann ist

$$\|u_h\|_{H^1(\Omega)} = \|w_h + f_h\|_{H^1(\Omega)} \leq \|w_h\|_{H^1(\Omega)} + \|f_h\|_{H^1(\Omega)} \leq (C_1 + 1) \|f_h\|_{H^1(\Omega)}.$$

Mit $C := C_1 + 1$ folgt dann die Aussage. Nun zu (2.3). Es galt

$$\beta \int_\Omega \nabla w_h \cdot \nabla v_h + \int_\Omega w_h v_h = -\beta \int_\Omega \nabla f_h \cdot \nabla v_h \qquad \forall v_h \in V_h.$$

Wählen wir $v_h := w_h$, so ergibt sich

$$\beta \int_\Omega |\nabla w_h|^2 + \int_\Omega w_h^2 = -\beta \int_\Omega \nabla f_h \cdot \nabla w_h$$

und daher mit $\tilde{\beta} := \min\{\beta, 1\}$ (also $\tilde{\beta} \leq \beta$ und $\tilde{\beta} \leq 1$)

$$\begin{aligned} \tilde{\beta} \|w_h\|^2_{H^1(\Omega)} &= \tilde{\beta} \int_\Omega |\nabla w_h|^2 + \tilde{\beta} \int_\Omega w_h^2 \leq \beta \int_\Omega |\nabla w_h|^2 + \int_\Omega w_h^2 \\ &= -\beta \int_\Omega \nabla f_h \cdot \nabla w_h \leq \left| -\beta \int_\Omega \nabla f_h \cdot \nabla w_h \right| \\ &\leq \beta \int_\Omega |\nabla f_h| \, |\nabla w_h| \leq \beta \|\nabla f_h\|_{L^2(\Omega)} \|\nabla w_h\|_{L^2(\Omega)} \\ &\leq \beta \|f_h\|_{H^1(\Omega)} \|w_h\|_{H^1(\Omega)} \end{aligned}$$

nach dem Satz von Hölder. Teilen wir nun durch $\tilde{\beta} \|w_h\|_{H^1(\Omega)}$, erhalten wir

$$\|w_h\|_{H^1(\Omega)} \leq \frac{\beta}{\tilde{\beta}} \|f_h\|_{H^1(\Omega)}.$$

Schließlich $C_1 := \beta / \min\{\beta, 1\}$. □

Wir zeigen nun, dass die Genauigkeit der numerischen Lösung wesentlich davon abhängt, wie gut die Lösung u in V_h approximiert werden kann.

Lemma 2.17 (Lemma von Céa)**.** *Sei u die exakte und u_h die numerische Lösung von*

$$\begin{aligned} -\beta\Delta u + u &= f && \text{in } \Omega, \\ u &= 0 && \text{auf } \partial\Omega. \end{aligned}$$

Dann ist

$$\|u - u_h\|_{H^1(\Omega)} \leq \frac{\max\{\beta, 1\}}{\min\{\beta, 1\}} \inf_{w_h \in V_h^0} \|u - w_h\|_{H^1(\Omega)}.$$

Dies ist die *Bestapproximationseigenschaft*. Bis auf eine Konstante ist die numerische Lösung so gut, wie die exakte Lösung durch stückweise affine Funktionen (auf dem gegebenen Gitter) approximiert werden kann. Nun zum

Beweis. Der Einfachheit halber betrachten wir hier nur den Fall, dass f stückweise affin ist, also $f_h = f$. Da u und u_h die exakte und numerische Lösung sind, gilt

$$\beta \int_\Omega \nabla u_h \cdot \nabla v_h + \int_\Omega u_h v_h = \int_\Omega f v_h \qquad \forall v_h \in V_h^0, \tag{2.4}$$

$$\beta \int_\Omega \nabla u \cdot \nabla v + \int_\Omega uv = \int_\Omega fv \qquad \forall v \in H_0^1(\Omega). \tag{2.5}$$

Da $V_h^0 \subset H^1(\Omega)$ gilt (2.5) insbesondere auch für alle $v_h \in V_h^0$, also

$$\beta \int_\Omega \nabla u \cdot \nabla v_h + \int_\Omega uv_h = \int_\Omega f v_h \qquad \forall v_h \in V_h^0. \tag{2.6}$$

Subtrahieren wir (2.4) von (2.6), so erhalten wir

$$\beta \int_\Omega (\nabla u - \nabla u_h) \cdot \nabla v_h + \int_\Omega (u - u_h) v_h = 0 \qquad \forall v_h \in V_h^0. \tag{2.7}$$

Damit ist mit einem beliebigen $w_h \in V_h^0$

$$\begin{aligned}
\min\{\beta,1\}\,\|u-u_h\|^2_{H^1(\Omega)} &\le \beta\int_\Omega(\nabla u-\nabla u_h)\cdot(\nabla u-\nabla u_h)+\int_\Omega(u-u_h)(u-u_h)\\
&= \beta\int_\Omega \nabla(u-u_h)\cdot\nabla(u-w_h+w_h-u_h)\\
&\quad+\int_\Omega(u-u_h)(u-w_h+w_h-u_h)\\
&= \beta\int_\Omega \nabla(u-u_h)\cdot\nabla(u-w_h)+\int_\Omega(u-u_h)(u-w_h)\\
&\quad+\beta\int_\Omega \nabla(u-u_h)\cdot\nabla\underbrace{(w_h-u_h)}_{=:v_h\in V_h^0}+\int_\Omega(u-u_h)\underbrace{(w_h-u_h)}_{=v_h}\Bigg\}\overset{(2.7)}{=}0\\
&\le \max\{\beta,1\}\Big(\|\nabla(u-u_h)\|_{L^2(\Omega)}\|\nabla(u-w_h)\|_{L^2(\Omega)}\\
&\qquad+\|u-u_h\|_{L^2(\Omega)}\|u-w_h\|_{L^2(\Omega)}\Big)\\
&\le \max\{\beta,1\}\left(\|u-u_h\|^2_{L^2(\Omega)}+\|\nabla(u-u_h)\|^2_{L^2(\Omega)}\right)^{1/2}\\
&\qquad\cdot\left(\|u-w_h\|^2_{L^2(\Omega)}+\|\nabla(u-w_h)\|^2_{L^2(\Omega)}\right)^{1/2}\\
&= \max\{\beta,1\}\,\|u-u_h\|_{H^1(\Omega)}\,\|u-w_h\|_{H^1(\Omega)}\,.
\end{aligned}$$

Teilen wir nun durch $\min\{\beta,1\}\,\|u-u_h\|_{H^1(\Omega)}$ erhalten wir

$$\|u-u_h\|_{H^1(\Omega)} \le \frac{\max\{\beta,1\}}{\min\{\beta,1\}}\,\|u-w_h\|_{H^1(\Omega)}\,.$$

Da w_h beliebig war, können wir auch dasjenige wählen, das $\|u-w_h\|_{H^1(\Omega)}$ minimiert. □

Bemerkung 2.18. *Ist die Lösung u glatt genug, so gilt*

$$\inf_{w_h\in V_h^0}\|u-w_h\|_{H^1(\Omega)} \le Ch$$

für eine von u abhängige Konstante $C>0$ und die Gitterweite h. Folglich

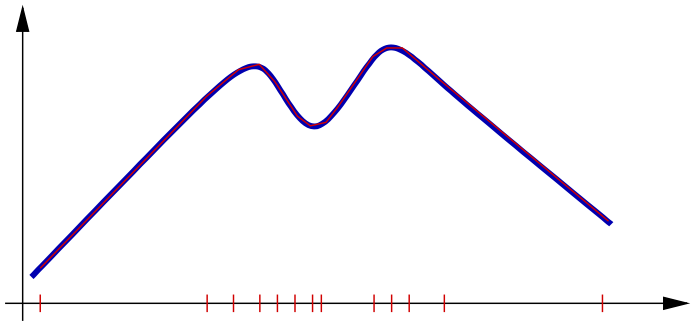
$$\|u-u_h\|_{H^1(\Omega)} \le \tilde{C}h$$

für ein $\tilde{C}>0$. Weiterhin kann man für glattes u zeigen

$$\|u-u_h\|_{L^2(\Omega)} \le \hat{C}h^2.$$

2.8 Adaptive Methoden

Wir haben gesehen, dass es wichtig ist, die Lösung u auf dem Gitter gut approximieren zu können. Dazu könnte man einfach das gesamte Gitter feiner machen. Dies treibt aber die Anzahl der Unbekannten und damit die Rechenzeit nach oben. Es ist auch gar nicht nötig, das Gitter überall fein zu machen. Betrachten wir als Beispiel folgendes u (blaue, dicke Kurve):



Schon mit relativ wenigen Gitterpunkten, die sich dort häufen wo sich u stark ändert, kann die Kurve mit einer stückweise affinen Funktion (rot) gut approximiert werden. Das Problem ist natürlich, dass wir die Lösung u vor der Berechnung mit einem gegebenen Gitter gar nicht kennen. Daher verwenden wir folgende Iteration

1. Starte mit grobem Gitter und löse das Problem
2. Markiere mit Hilfe eines *Fehlerschätzers* die Dreiecke, die verfeinert werden müssen
3. *Verfeinere* die markierten Dreiecke
4. Löse das Problem auf dem neuen Gitter
5. Gehe zu 2

2.8.1 Fehlerschätzer

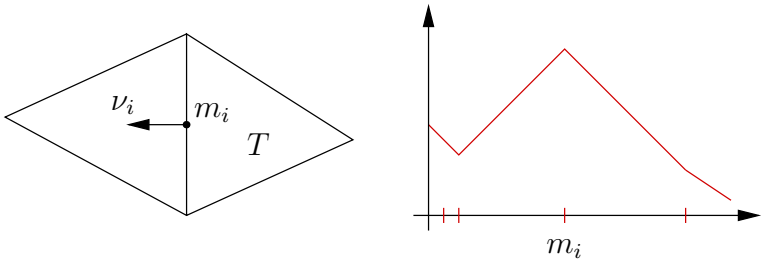
Als einfachen Fehlerindikator η_T auf einem Dreieck verwenden wir den mit der Seitenlänge gewichteten Sprung der Normalenableitung über die Kanten:

$$\eta_T^2 := \sum_{i=1}^{3} \left(|s_i| \left[\frac{\partial u_h}{\partial \nu_i} \right]_{s_i} \right)^2,$$

wobei s_i wieder die Kanten von T sind, $|s_i|$ deren Länge und $[\partial u_h/\partial \nu_i]_{s_i}$ ist der Sprung der Normalenableitung $\nabla u_h \cdot \nu_i =: \partial u_h/\partial \nu_i$ über die Kante s_i:

$$\left[\frac{\partial u_h}{\partial \nu_i} \right]_{s_i} = \lim_{\varepsilon \searrow 0} \frac{\partial u_h}{\partial \nu_i}(m_i + \varepsilon \nu_i) - \lim_{\varepsilon \searrow 0} \frac{\partial u_h}{\partial \nu_i}(m_i - \varepsilon \nu_i),$$

wobei ν_i die zur Kante s_i gehörende Normale ist (z.B. s_i um 90° gedreht) und m_i ein Punkt auf s_i ist.



In 1d entspricht dies gerade dem Sprung der Ableitung am Gitterpunkt m_i.
Es gibt nun verschiedene Strategien, nach denen man Dreiecke markieren kann. Unter anderem

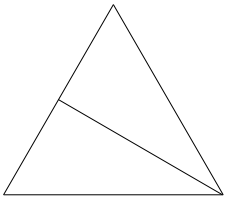
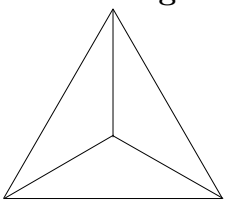
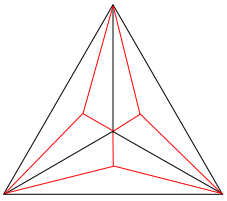
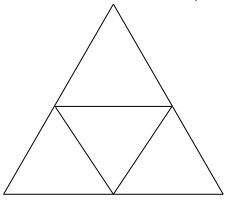
- Wähle Fehlerschranke η
  Markiere T zur Verfeinerung, falls $\eta_T > \eta$
- Wähle $p \in (0, 1)$
  Markiere diejenigen $p \cdot N_T$ Dreiecke mit dem größten η_T

2.8.2 Verfeinerung

Bei der Verfeinerung einer Triangulierung müssen wir darauf achten, dass das neue Gitter

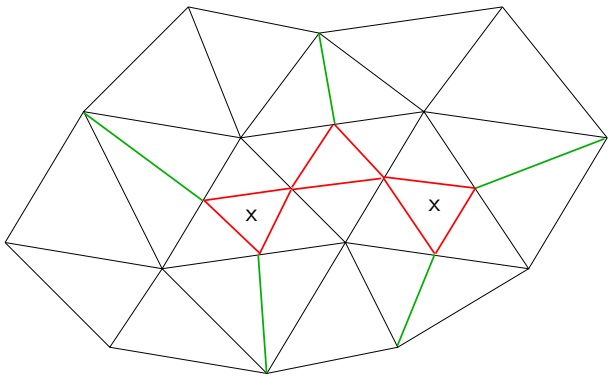
- keine hängenden Knoten hat und
- keine zu spitzen Winkel bekommt.

Zunächst diskutieren wir elementare Verfeinerungsoperationen, d.h. Verfahren um ein Dreieck in mehrere zu unterteilen.

- **Zweiteilung, Bisektion** (durch die Kantenmitte)
  
  Ein Winkel wird halbiert, ein hängender Knoten entsteht
- **Dreiteilung**
  
  keine hängenden Knoten
  
  wiederholte Dreiteilung führt jedoch zu sehr kleinen Winkeln
- **Vierteilung** (durch die Kantenmitten)
  
  Winkel bleiben gleich, es entstehen jedoch hängende Knoten

Keine dieser Operationen ist alleine gesehen optimal. Wir müssen uns also *Verfeinerungsstrategien* anschauen, die die obigen Operationen kombinieren. Da die wiederholte Dreiteilung die Winkel zwangsläufig immer weiter verkleinert, werden nur die beiden anderen verwendet.
Wir diskutieren hier zwei weit verbreitete Strategien. Dazu seien die Dreiecke markiert, die unterteilt werden sollen (in den Bildern durch ein x dargestellt).

1. **Rot-Grün-Verfeinerung**



   Für jedes markierte Dreieck

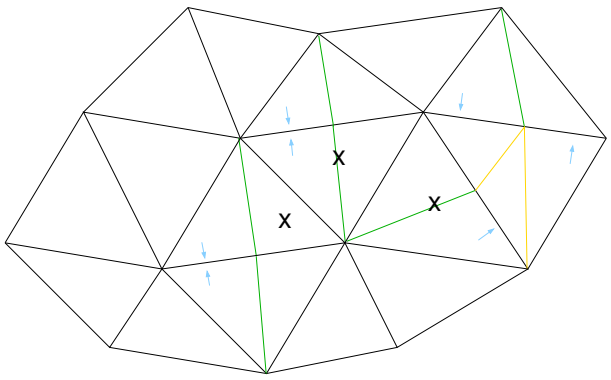
   a) Vierteile das Dreieck („rote Verfeinerung“)

   b) Überprüfe alle Nachbarn des Dreiecks.

      - Falls das Dreieck nicht markiert ist, beseitige hängende Knoten durch Bisektion („grüne Verfeinerung“)
      - Falls das Dreieck grün verfeinert ist, entferne die grüne Verfeinerung und verfeinere rot (incl. Nachbar-Test)

   Vor einer weiteren Verfeinerung wird die grüne Verfeinerung wieder entfernt. Dadurch wird ein Dreieck stets nur einmal grün verfeinert, und nur diese Operation verschlechtert die Winkel.

2. **Verfeinerung der längsten Kante** (longest edge bisection)



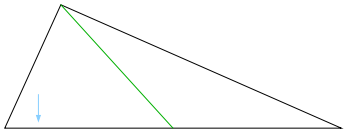
(die blauen Pfeile zeigen die längste Kante an)

a) Für jedes markierte Dreieck teile die längste Kante

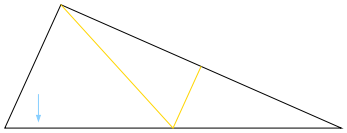
b) Für jedes (unabhängig von der Markierung) Dreieck, das eine geteilte Kante besitzt, teile die längste Kante (falls die längste Kante schon geteilt war, ist hier nichts zu tun). Wiederhole diesen Schritt so lange, bis keine neuen Kanten mehr geteilt werden.

c) Erstelle neue Dreiecke

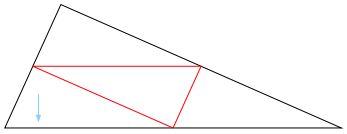
- Falls nur eine Kante geteilt ist, verbinde deren Mittelpunkt mit der gegenüberliegenden Ecke.



- Falls zwei Kanten geteilt sind, verbinde den Mittelpunkt der längsten Kante mit dem Mittelpunkt der zweiten Kante sowie der gegenüberliegenden Ecke.



- Falls alle drei Kanten geteilt sind, vierteile das Dreieck.



Für beide Verfeinerungsstrategien gilt, dass

- sie terminieren (d.h. keine Endlosschleifen)
- die Winkel nach unten beschränkt bleiben
- keine hängenden Knoten entstehen

Um die Verfeinerungsstrategien in MATLAB zu implementieren, müssen weitere Informationen gespeichert werden, z.B. Nachbarschaftsrelationen, Markierungen der Dreiecke, etc.

2.9 Lineare Elastizitätstheorie

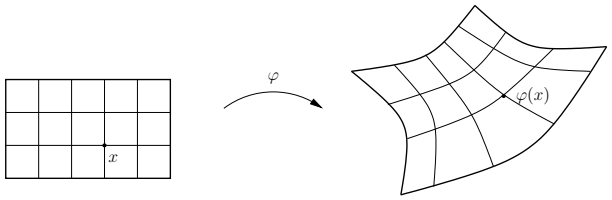
Wir betrachten in diesem Abschnitt einen (zweidimensionalen) Gegenstand auf den Kräfte wirken und zwar

- Volumenkräfte, die auf alle Teile des Gegenstandes wirken, z.B. Gravitation, sowie
- Oberflächenkräfte, die nur auf den Rand wirken, z.B. ein Glas auf einem Tisch oder ein Auto auf einer Brücke.

Wir wollen dann berechnen, wie sich der Gegenstand verformt und welche Spannungen auftreten.
Elastizitätstheorie (in 3d) hat Anwendungen in vielen verschiedenen Gebieten, z.B. bei der Konstruktion von Bauteilen, Häusern, Brücken, etc. oder bei der Ausbreitung von seismischen Wellen. In letzterem Fall müsste man allerdings die zeitabhängigen Gleichungen betrachten.
Wir beschäftigen uns hier nur mit der linearen Theorie, gehen also von geringen Verformungen des Gegenstandes aus.

Bemerkung 2.19. *Wir sprechen von elastischer Verformung, wenn der Gegenstand bei Nachlassen der Kraft wieder in seinen ursprünglichen Zustand zurückkehrt. Bei einer dauerhaften Verformung sprechen wir von plastischer Verformung.*

Wir gehen wie folgt vor: zunächst definieren wir einige grundlegende Begriffe der Elastizitätstheorie sowie die Energie eines elastischen Gegenstandes bei gegebener Verformung. Diese Energie wird dann unter gegebenen Kräften minimiert, um die Verformung zu berechnen.
Sei also $\Omega \subset \mathbb{R}^2$ ein Gebiet mit Rand Γ, das den Gegenstand in seiner Referenzkonfiguration, also ohne einwirkende Kräfte und frei von Spannungen zeigt.
Der Gegenstand werde durch die *Deformation* φ verformt, d.h. der Punkt x aus der Referenzkonfiguration befindet sich nach der Verformung an der Position $\varphi(x)$.



Dabei fordern wir für φ, dass

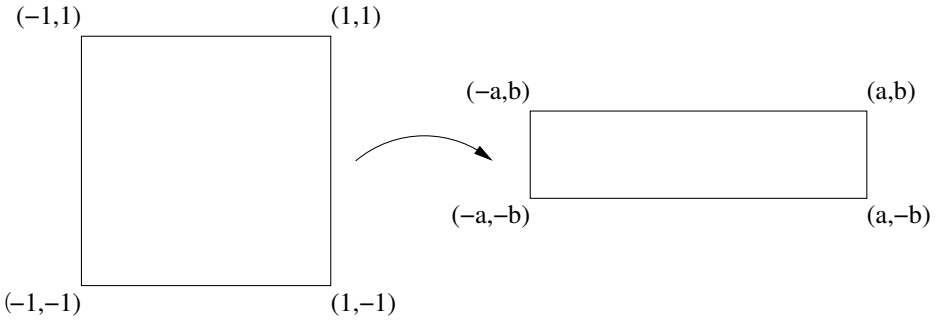
- φ stetig differenzierbar,
- φ injektiv ($x \neq y \Rightarrow \varphi(x) \neq \varphi(y)$), es gibt also keine Selbstüberschneidung, sowie
- $\det(D\varphi) > 0$ (orientierungserhaltend)

ist. Hierbei ist

$$D\varphi = \begin{pmatrix} \partial_1\varphi_1 & \partial_2\varphi_1 \\ \partial_1\varphi_2 & \partial_2\varphi_2 \end{pmatrix}.$$

Beispiel 2.20. *Sei* $\Omega = [-1,1]^2$ *und*

$$\varphi(x_1, x_2) = \begin{pmatrix} \varphi_1(x_1, x_2) \\ \varphi_2(x_1, x_2) \end{pmatrix} := \begin{pmatrix} ax_1 \\ bx_2 \end{pmatrix}.$$



Dann ist

$$D\varphi = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}.$$

Was ist nun die Wirkung einer Deformation φ?
Wir betrachten ein Längenstück in Ω, beschrieben durch den Pfad

$$c : [0,1] \to \Omega, \quad c(t) = \begin{pmatrix} c_1(t) \\ c_2(t) \end{pmatrix}.$$

Die Länge von c ist gegeben durch

$$\int_0^1 \|\dot{c}(t)\| \, \mathrm{d}t = \int_0^1 \sqrt{\dot{c}_1(t)^2 + \dot{c}_2(t)^2} = \int_0^1 \sqrt{\dot{c}(t)^T \dot{c}(t)}.$$

Der verformte Pfad ist $\gamma(t) := \varphi(c(t))$ mit der Ableitung

$$\dot{\gamma}(t) = D\gamma(t) = D(\varphi(c(t))) = D\varphi(c(t))\dot{c}(t).$$

Daher ist die Länge des verformten Pfades durch

$$\begin{aligned} \int_0^1 \|\dot{\gamma}(t)\| &= \int_0^1 \sqrt{\dot{\gamma}(t)^T \dot{\gamma}(t)} = \int_0^1 \sqrt{(D\varphi\dot{c}(t))^T D\varphi\dot{c}(t)} \\ &= \int_0^1 \sqrt{\dot{c}(t)^T D\varphi^T D\varphi \, \dot{c}(t)}. \end{aligned}$$

gegeben. Die Matrix $D\varphi^T D\varphi$ heißt *Verzerrungstensor* (siehe auch der metrische Tensor einer parametrisierten Fläche) und kann als Maß für Längenänderungen im Material angesehen werden.
Im Fall $D\varphi^T D\varphi = \mathrm{id}$ ändert sich die Länge nicht. Man kann zeigen, dass die Deformation in diesem Fall eine Starrkörperbewegung, also eine Drehung und/oder Veschiebung ist.
Wir betrachten die über die Starrkörpertransformation hinausgehenden Veränderung und bezeichnen mit

$$V := \tfrac{1}{2}(D\varphi^T D\varphi - \mathrm{id})$$

die *Verzerrung*. Ebenso betrachten wir statt der Deformation φ die *Verschiebung*

$$u := \varphi - \mathrm{id}.$$

Um V mit u auszudrücken, berechnen wir $Du = D\varphi - \mathrm{id} \iff D\varphi = Du + \mathrm{id}$. Damit ist

$$\begin{aligned} D\varphi^T D\varphi - \mathrm{id} &= (Du + \mathrm{id})^T (Du + \mathrm{id}) - \mathrm{id} \\ &= (Du^T + \mathrm{id})(Du + \mathrm{id}) - \mathrm{id} \\ &= Du^T Du + Du^T + Du + \mathrm{id} - \mathrm{id} \\ &= Du^T + Du + Du^T Du, \end{aligned}$$

also

$$V = \tfrac{1}{2}(Du^T + Du) + \tfrac{1}{2} Du^T Du.$$

Wenn Du klein ist, dann ist der quadratische Anteil $Du^T Du$ viel kleiner als der lineare Anteil Du und kann vernachlässigt werden. Wir betrachten im Folgenden nur den Fall *linearer Elastizität* und definieren

Definition 2.21 (Linearisierter Verzerrungstensor)**.** *Sei u eine Verschiebung. Dann bezeichnet*

$$\varepsilon(u) := \tfrac{1}{2}(Du^T + Du)$$

den linearisierten Verzerrungstensor.

Bemerkung 2.22. *ε ist der symmetrische Anteil der Ableitung von u:*

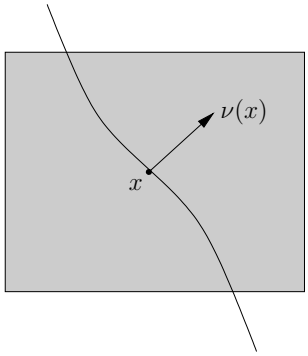
$$Du = \tfrac{1}{2} \underbrace{(Du^T + Du)}_{\text{symmetrisch}} + \tfrac{1}{2} \underbrace{(Du^T - Du)}_{\text{schiefsymmetrisch}}.$$

Der schiefsymmetrische Teil charakterisiert die infinitesimalen Drehungen, die für die Verzerrung keine Rolle spielen.
Die Flächenänderung wird nach dem Transformationssatz durch den Faktor $\det D\varphi$ beschrieben. Analog zur Längenänderung suchen wir eine Approximation für die Differenz $\det D\varphi - 1$, wobei 1 der Faktor einer Starrkörpertransformationen ist, unter der Annahme dass Du klein ist.

$$\begin{aligned} \det D\varphi - 1 &= \det \begin{pmatrix} \partial_1 u_1 + 1 & \partial_2 u_1 \\ \partial_1 u_2 & \partial_2 u_2 + 1 \end{pmatrix} - 1 \\ &= (\partial_1 u_1 + 1)(\partial_2 u_2 + 1) - \partial_2 u_1 \partial_1 u_2 - 1 \\ &= \partial_1 u_1 + \partial_2 u_2 + \partial_1 u_1 \partial_2 u_2 - \partial_2 u_1 \partial_1 u_2 \\ &\approx \mathrm{div} u \end{aligned}$$

Also gibt $\mathrm{div} u$ näherungsweise die Flächenänderung an.

Uns interessieren vor allem die Spannungen, die im Material auftreten. Um die im Material auftretenden Spannungen zu quantifizieren, stellen wir uns eine durch den elastischen Körper laufende Fläche vor und untersuchen die Kraft, die auf diese Fläche wirkt.



Dazu nehmen wir an, dass der Körper entlang dieser Fläche zerschneiden. Um ihn genauso zu deformieren wie den ursprünglichen Körper, müssen wir auf diese Fläche in der Regel eine Kraft ausüben. Diese Kraft lässt sich punktweise durch den Spannungstensor und die Normale an die Fläche beschreiben:

Satz 2.23 (Spannungstensor)**.** *Sei u die Verschiebung eines (isotropen) linearelastischen Körpers Ω zu gegebenen Volumen- und Randkräften. Die Kraft, die entlang der gedachten Fläche vorgegeben werden muss, um die beiden Teile von Ω entsprechend der Verschiebung u zu deformieren, beträgt im Punkt x*

$$F(x) = \sigma(u)\nu(x),$$

wobei ν die Normale an die gedachte Fläche durch den Körper ist. Die Matrix $\sigma(u)$ ist der Spannungstensor*, gegeben durch*

$$\sigma(u) = 2\mu\varepsilon(u) + \lambda(\operatorname{tr}\varepsilon(u))\mathrm{id},$$

wobei $\operatorname{tr}\varepsilon(u) \in \mathbb{R}$ die Spur der Matrix $\varepsilon(u)$ ist. Dieser proportionale Zusammenhang zwischen Spannung und Verformung heißt auch Hookesches Gesetz*. Die* Lamé-Konstanten *λ und μ sind materialabhängig.*

Mit obiger Formel können wir also bei gegebener Verschiebung die im Material auftretenden Kräfte berechnen. Unser Ziel war aber der umgekehrte Fall: wir haben auf den Körper wirkende Kräfte vorgegeben und wollen die entstehende Verschiebung berechnen. Dazu gehen wir wie folgt vor: wir betrachten die *elastische Energie* eines Körpers bei gegebenen Kräften und suchen dann diejenige Verschiebung mit der minimalen Energie.
Wir brauchen zur Formulierung noch die

Notation 2.24. *Für zwei $n \times n$-Matrizen A und B schreiben wir (analog zum Skalarprodukt zweier Vektoren)*

$$A : B = \sum_{i,j=1}^{n} A_{ij}B_{ij}.$$

Ebenso wie die Länge oder Norm eines Vektors a durch $\|a\| = \sqrt{a \cdot a}$ gegeben ist, definieren wir die Frobenius-Norm *als*

$$\|A\|_F = \sqrt{A : A} = \sqrt{\sum_{i,j=1}^{n} A_{ij}^2}.$$

Satz 2.25 (Elastische Energie)**.** *Die elastische Energie eines (isotropen, linear elastischen) Körpers $\Omega \subset \mathbb{R}^2$ bei gegebener Verschiebung u ist gegeben durch*

$$E(u) = \mu \int_\Omega \varepsilon(u) : \varepsilon(u) + \frac{\lambda}{2} \int_\Omega (\operatorname{tr} \varepsilon(u))^2 - \int_\Omega f \cdot u - \int_{\Gamma_1} g \cdot u,$$

wobei $f : \Omega \to \mathbb{R}^2$ die auf den Körper wirkende Volumenkraft und $g : \Gamma_1 \to \mathbb{R}^2$ die auf das Randstück $\Gamma_1 \subset \Gamma$ wirkende Flächenkraft ist.

Zur Bedeutung der Terme:

- In der linearen Approximation ist $\varepsilon(u) \approx V$, wobei V ein Maß für Längenänderungen. Daher „bestraft“ der erste Term, $\int_\Omega \|\varepsilon(u)\|_F^2$, Längenänderungen im Material.

- Um den zweiten Term besser zu verstehen, bemerken wir

$$\begin{aligned} \operatorname{tr} \varepsilon(u) &= \sum_k \varepsilon(u)_{kk} = \sum_k \tfrac{1}{2}(Du^T + Du)_{kk} = \sum_k (Du)_{kk} = \sum_k \partial_k u_k \\ &= \operatorname{div} u. \end{aligned}$$

  Der zweite Term ist also $\|\operatorname{div} u\|_{L^2(\Omega)}$ und misst die Volumenkontraktion.

Die beiden Terme sind nicht unabhängig voneinander. Es gibt zwar Längenänderungen ohne Volumenkontraktion, aber nicht umgekehrt, denn

$$\|\varepsilon(u)\|_F = 0 \Rightarrow \varepsilon(u) = 0 \Rightarrow \operatorname{tr} \varepsilon(u) = 0.$$

Um das Minimum der Energie zu finden, berechnen wir wie in den vorigen Abschnitten die erste Variation.

$$\begin{aligned} E(u + tv) =& \mu \int_\Omega \varepsilon(u + tv) : \varepsilon(u + tv) + \frac{\lambda}{2} \int_\Omega (\operatorname{div} u + t \operatorname{div} v)^2 \\ & - \int_\Omega f \cdot (u + tv) - \int_{\Gamma_1} g \cdot (u + tv). \end{aligned}$$

Wegen

$$\varepsilon(u + tv) : \varepsilon(u + tv) = \varepsilon(u) : \varepsilon(u) + 2t\varepsilon(u) : \varepsilon(v) + t^2 \varepsilon(v) : \varepsilon(v)$$

erhalten wir

$$\frac{d}{dt}E(u+tv)\Big|_{t=0} = 2\mu\int_\Omega \varepsilon(u):\varepsilon(v) + \lambda\int_\Omega (\mathrm{div} u)(\mathrm{div} v) - \int_\Omega f\cdot v - \int_{\Gamma_1} g\cdot v.$$

Die zu lösende schwache Gleichung ist also

$$2\mu\int_\Omega \varepsilon(u):\varepsilon(v) + \lambda\int_\Omega (\mathrm{div} u)(\mathrm{div} v) - \int_\Omega f\cdot v - \int_{\Gamma_1} g\cdot v = 0 \tag{2.8}$$

für alle v. Bevor wir zur Diskretisierung dieser Gleichung kommen, leiten wir noch die starke Form der Gleichung her und rechnen ein Beispiel.
Um die starke Form herzuleiten, integrieren wir in der schwachen Form partiell. Wir verwenden die folgende

Notation 2.26 (Zeilenweise Divergenz)**.** *Für eine matrixwertige Abbildung $A:\Omega\to\mathbb{R}^{2\times 2}$ schreiben wir*

$$\mathrm{Div}\, A(x) := \left(\sum_{j=1}^{2}\partial_j A_{ij}(x)\right)_{i=1,\dots,2}.$$

$\mathrm{Div}\, A(x)$ ist also ein Vektor, bei dem im i-ten Eintrag die Divergenz der i-ten Zeile von $A(x)$ steht.

Mit dieser Notation können wir die partielle Integration auf matrixwertige Abbildungen verallgemeinern.

Lemma 2.27. *Sei A matrixwertige Abbildung und v eine vektorwertige Funktion. Dann gilt*

$$\int_\Omega A : Dv = -\int_\Omega \mathrm{Div}\, A\cdot v + \int_{\partial\Omega} A^T v\cdot\nu,$$

wobei wie üblich ν die äußere Normale an Ω ist.

Beweis.

$$\begin{aligned}
\int_\Omega A:Dv &= \sum_{i,j=1}^{2}\int_\Omega A_{ij}\partial_j v_i = \sum_{i,j=1}^{2}\left(-\int_\Omega \partial_j A_{ij} v_i + \int_{\partial\Omega} A_{ij} v_i \nu_j\right)\\
&= -\int_\Omega \sum_{i=1}^{2} v_i\left(\sum_{j=1}^{2}\partial_j A_{ij}\right) + \int_\Omega\sum_{i=1}^{2} v_i\left(\sum_{j=1}^{2} A_{ij}\nu_j\right)\\
&= -\int_\Omega \sum_{i=1}^{2} v_i\left(\mathrm{Div}\, A\right)_i + \int_\Omega\sum_{i=1}^{2} v_i\left(A\nu\right)_i\\
&= -\int_\Omega \mathrm{Div}\, A\cdot v + \int_{\partial\Omega} A^T v\cdot\nu.
\end{aligned}$$

□

Damit können wir nun in Gleichung (2.8) partiell integrieren. Für den ersten Term erhalten wir

$$\begin{aligned}\int_\Omega \varepsilon(u) : \varepsilon(v) &= \int_\Omega \varepsilon(u) : \tfrac{1}{2}(Dv^T + Dv) = \tfrac{1}{2}\int_\Omega \varepsilon(u) : Dv^T + \tfrac{1}{2}\int_\Omega \varepsilon(u) : Dv \\ &= \tfrac{1}{2}\int_\Omega \varepsilon(u)^T : Dv + \tfrac{1}{2}\int_\Omega \varepsilon(u) : Dv = \tfrac{1}{2}\int_\Omega \varepsilon(u) : Dv + \tfrac{1}{2}\int_\Omega \varepsilon(u) : Dv \\ &= \int_\Omega \varepsilon(u) : Dv = -\int_\Omega \operatorname{Div}\varepsilon(u)\cdot v + \int_{\partial\Omega} \varepsilon(u)v\cdot\nu\end{aligned}$$

und der zweite Term ergibt

$$\int_\Omega (\operatorname{div} u)(\operatorname{div} v) = -\int_\Omega \nabla \operatorname{div} u \cdot v + \int_{\partial\Omega} (\operatorname{div} u) v \cdot \nu.$$

Zusammen erhalten wir also

$$\begin{aligned}0 = &\int_\Omega (-2\mu \operatorname{Div}\varepsilon(u) - \lambda\nabla\operatorname{div} u - f)\cdot v \\ &+ \int_{\partial\Omega} (2\mu\varepsilon(u) + \lambda(\operatorname{div} u)\mathrm{id})\,\nu\cdot v - \int_{\Gamma_1} g\cdot v\end{aligned}$$

für alle v. Daher ist die starke Form der Gleichung gegeben durch

$$\begin{aligned}-2\mu\operatorname{Div}\varepsilon(u) - \lambda\nabla\operatorname{div} u &= f && \text{in } \Omega \\ (2\mu\varepsilon(u) + \lambda(\operatorname{div} u)\mathrm{id})\,\nu &= g && \text{auf } \Gamma_1 \\ (2\mu\varepsilon(u) + \lambda(\operatorname{div} u)\mathrm{id})\,\nu &= 0 && \text{auf } \Gamma_0 \\ u &= 0 && \text{auf } \Gamma_D\end{aligned}$$

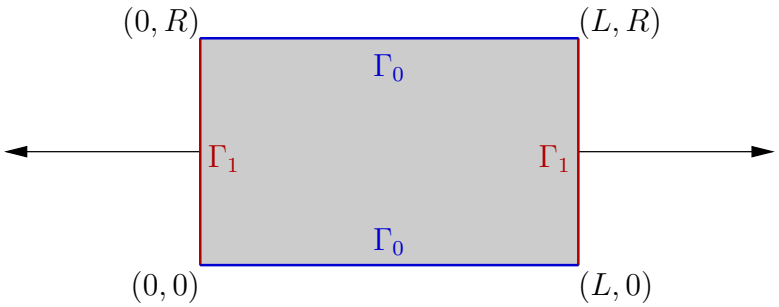
Mit Hilfe des Spannungstensors lässt sich dies vereinfachen zu

$$-\operatorname{Div}\sigma(u) = f \quad \text{in } \Omega \tag{2.9a}$$
$$\sigma(u)\nu = g \quad \text{auf } \Gamma_1 \tag{2.9b}$$
$$\sigma(u)\nu = 0 \quad \text{auf } \Gamma_0 \tag{2.9c}$$
$$u = 0 \quad \text{auf } \Gamma_D \tag{2.9d}$$

Beispiel 2.28. *Wir betrachten einen rechteckigen Gegenstand, $\Omega = (0, L) \times (0, R)$, und ziehen an zwei Enden des Gegenstandes nach außen, also*

$$g = \begin{cases} \begin{pmatrix} -1 \\ 0 \end{pmatrix} & \text{auf } \left\{ (0, y) \;\middle|\; y \in (0, R) \right\} \\ \begin{pmatrix} 1 \\ 0 \end{pmatrix} & \text{auf } \left\{ (L, y) \;\middle|\; y \in (0, R) \right\} \end{cases}$$

und ohne Schwerkaft, also $f = 0$.



Da nach Definition der Energie diese nicht verändert wird, wenn wir den Körper verschieben, müssen wir ihn noch „festhalten“, setzen also z.B. $\Gamma_D := \{(0,0)\}$. *Wir betrachten nun die Abbildung*

$$u(x,y) = \underbrace{\begin{pmatrix} \alpha & \gamma_1 \\ \gamma_2 & \beta \end{pmatrix}}_{=:A} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}.$$

Diese erfüllt (2.9a), *denn*

$$Du = A \quad \Rightarrow \quad \varepsilon(u) = \begin{pmatrix} \alpha & \frac{1}{2}(\gamma_1+\gamma_2) \\ \frac{1}{2}(\gamma_1+\gamma_2) & \beta \end{pmatrix},$$

also ist $\varepsilon(u)$ *konstant und damit* $\operatorname{Div}\varepsilon(u) = 0$. *Ebenso ist* $\operatorname{tr}\varepsilon(u) = \alpha+\beta$ *konstant und daher* $\nabla \operatorname{tr}\varepsilon(u) = 0$. *Nun müssen wir überprüfen, ob diese Abbildung die Randbedingungen erfüllen kann.*
Festhalten der Körpers (2.9d) *liefert* $b_1 = b_2 = 0$. *Für den Spannungstensor erhalten wir*

$$\begin{aligned} \sigma(u) &= 2\mu\varepsilon(u) + \lambda \operatorname{tr}\varepsilon(u)\mathrm{id} \\ &= 2\mu \begin{pmatrix} \alpha & \frac{1}{2}(\gamma_1+\gamma_1) \\ \frac{1}{2}(\gamma_1+\gamma_1) & \beta \end{pmatrix} + \lambda \begin{pmatrix} \alpha+\beta & 0 \\ 0 & \alpha+\beta \end{pmatrix} \\ &= \begin{pmatrix} 2\mu\alpha + \lambda(\alpha+\beta) & \mu(\gamma_1+\gamma_2) \\ \mu(\gamma_1+\gamma_2) & 2\mu\beta + \lambda(\alpha+\beta) \end{pmatrix}. \end{aligned}$$

Die Randbedingungen (2.9b) *liefern*

$$\begin{pmatrix} 2\mu\alpha + \lambda(\alpha+\beta) \\ \mu(\gamma_1+\gamma_2) \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

und (2.9c) *liefern*

$$\begin{pmatrix} \mu(\gamma_1+\gamma_2) \\ 2\mu\beta + \lambda(\alpha+\beta) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Wir bekommen also die zwei Gleichungen

$$\begin{aligned} (2\mu+\lambda)\alpha + \lambda\beta &= 1 \\ \lambda\alpha + (2\mu+\lambda)\beta &= 0 \end{aligned}$$

für die beiden Unbekannten α *und* β. *Im Fall* $\mu = 0$ *haben diese Gleichungen keine Lösung. Für* $\mu \neq 0$ *erhalten wir*

$$\alpha = \frac{2\mu+\lambda}{4\mu(\mu+\lambda)}, \qquad \beta = -\frac{\lambda}{4\mu(\mu+\lambda)}.$$

Für die beiden anderen Unbekannten γ_1 *und* γ_2 *erhalten wir*

$$\mu(\gamma_1+\gamma_2) = 0 \qquad \Longrightarrow \qquad \gamma_2 = -\gamma_1.$$

Damit erhalten wir also eine Familie von Lösungen

$$u_\gamma(x,y) = \frac{1}{4\mu(\mu+\lambda)} \begin{pmatrix} 2\mu+\lambda & 0 \\ 0 & -\lambda \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \gamma \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

Den zweiten Term kann man wie folgt interpretieren: sei $\tilde{\varphi}$ eine Drehung um den Winkel γ, also

$$\tilde{\varphi}(x,y) = \begin{pmatrix} \cos(\gamma) & -\sin(\gamma) \\ \sin(\gamma) & \cos(\gamma) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \qquad \Longrightarrow \qquad \tilde{u}(x,y) = \begin{pmatrix} \cos(\gamma)-1 & -\sin(\gamma) \\ \sin(\gamma) & \cos(\gamma)-1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

Für sehr kleine Winkel γ ist $\cos(\gamma) \approx 1$ und $\sin(\gamma) \approx \gamma$ (vgl. Taylor-Reihe). Daher

$$\tilde{u}(x,y) = \begin{pmatrix} 0 & -\gamma \\ \gamma & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

und so kann dieser Term als Linearisierung einer Drehung angesehen werden. Betrachten wir die Lösungen ohne Drehung, also

$$u(x,y) = \frac{1}{4\mu(\mu+\lambda)} \begin{pmatrix} 2\mu+\lambda & 0 \\ 0 & -\lambda \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix},$$

so sehen wir, dass der Körper wie erwartet in x-Richtung gedehnt wird und in y-Richtung gestaucht wird. Die Stärke der Stauchung hängt von λ ab. Im Fall $\lambda = 0$ wirken keine Kräfte quer zur Zugrichtung und der Körper wird in y-Richtung nicht gestaucht.

2.9.1 Diskretisierung

Als Basis für den Finite-Elemente-Raum wählen wir die $2N$ Basisfunktionen $\psi_{ik}(x)$, definiert durch

$$\psi_{ik}(x_j) = \delta_{ij} e_k \qquad i,j = 1,\ldots,N, k = 1,2,$$

wobei

$$e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \text{und} \qquad e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Es bietet sich an, die Steifigkeitsmatrix in vier Unter-Matrizen zu zerlegen,

$$L = \begin{pmatrix} L^{11} & L^{12} \\ L^{21} & L^{22} \end{pmatrix},$$

wobei die Einträge gegeben sind durch

$$L_{ij}^{kl} = 2\mu \int_\Omega \varepsilon(\psi_{ik}) : \varepsilon(\psi_{jl}) + \lambda \int_\Omega (\mathrm{div}\psi_{ik})(\mathrm{div}\psi_{jl}).$$

Die Berechnung der Einträge basiert auf dem Transformationssatz und der Kettenregel.

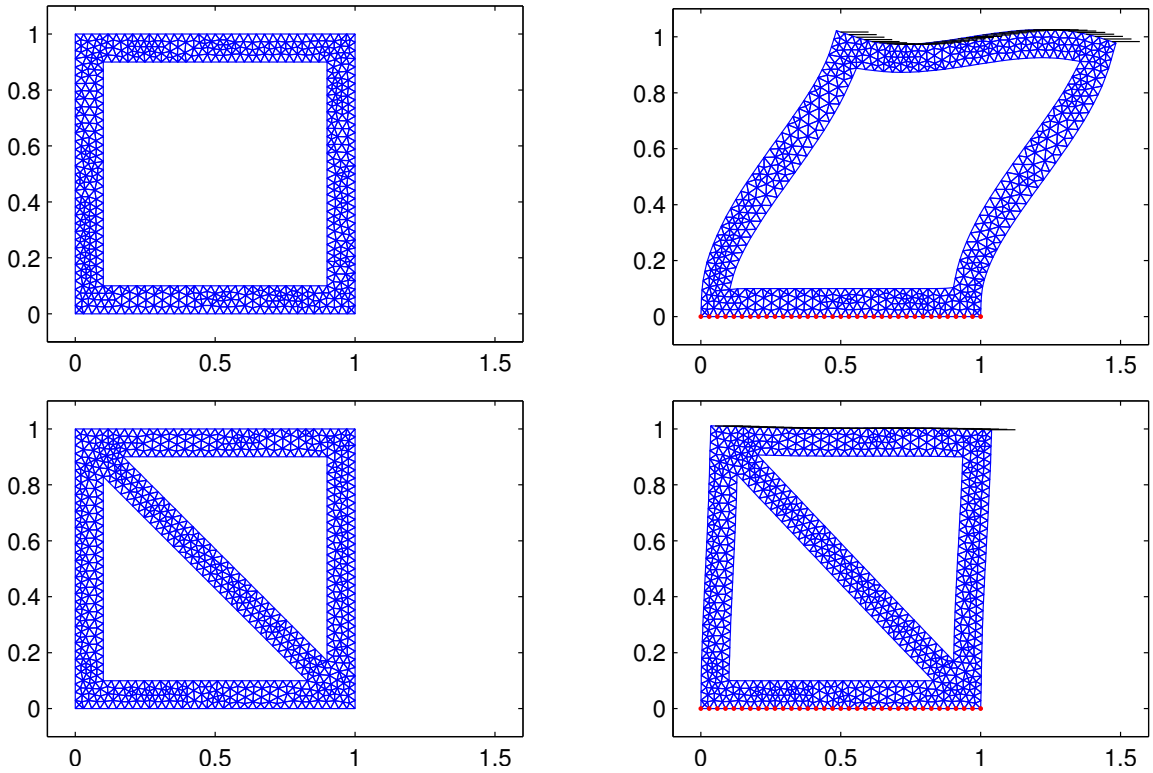
Programmieraufgabe 5. *Programmieren Sie einen Löser für lineare Elastizität und wenden Sie ihn auf folgende Beispiele an*

1. *Einem Block, an dem an beiden Seiten gezogen wird. Vergleiche Sie das Ergebnis mit der analytischen Lösung aus der Vorlesung.* Bemerkung. *Das Ergebnis aus der Vorlesung erhalten Sie nur für $\lambda = 0$ und diesen Dirichlet-Randwerten. Können Sie erklären warum?*

2. *Einem Bauteil mit und ohne Querverstrebung. Finden Sie sinnvolle Werte für μ und λ und vergleichen Sie die Verschiebungen.* Bemerkung. *Versuchen Sie z.B. $\lambda = 1600$ und $\mu = 270$.*

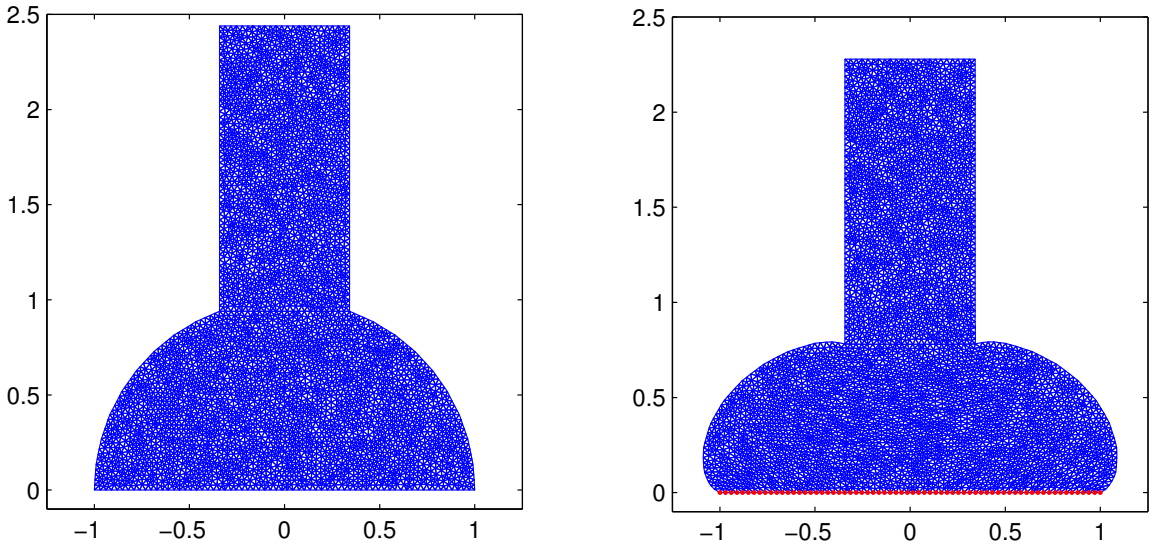
3. *Erweitern Sie den Löser auf nicht-konstante Werte für die Volumenkraft f sowie für die Parameter λ und μ. Wenden Sie den Löser auf das Beispiel eines Hauses (hohe Steifigkeit, große Masse) auf einem Hügel (geringere Steifigkeit, niedrigere Masse bzw. Dichte) an.*

Die Ergebnisse sind für (2):



Deutlich sichtbar ist, dass bei gleicher Last die Querverstrebung zu einer viel geringeren Verschiebung führt.

Für (3) erhalten wir



Auf Grund des weicheren Materials sinkt das Haus in den Hügel ein.
Hinweis: Damit man überhaupt etwas sieht, wurden die Parameter so gewählt, dass die Verschiebung groß ist. Eigentlich ist dann das lineare Modell natürlich nicht mehr gültig.

3 Randelemente

Wir betrachten folgendes Modellproblem. Sei $\Omega \subset \mathbb{R}^n$, $n = 2, 3$, beschränkt (z.B. Erde) und Γ der Rand von Ω. Wir betrachten das Gravitationspotential $u : \mathbb{R}^n \to \mathbb{R}$ im Außenraum

$$-\Delta u = 0 \quad \text{in } \mathbb{R}^n \setminus \Omega \tag{3.1}$$

und u soll klein werden, wenn $|x| \to \infty$, z.B.

$$|u| \leq c \, \|x\|^{-1} \quad \text{für } |x| \to \infty, c \in \mathbb{R}.$$

Wir haben gegeben $\nabla u \cdot \nu$, wobei ν die äußere Normale an $\mathbb{R}^n \setminus \Omega$ also die innere Normale an Ω ist, und suchen u auf $\mathbb{R}^n \setminus \Omega$.
Da $\mathbb{R}^n \setminus \Omega$ nicht beschränkt ist, können wir nicht wie bisher das Gebiet triangulieren.

3.1 Lösung im Außenraum

Wir multiplizieren (3.1) mit einer Testfunktion φ und integrieren über $\mathbb{R}^n \setminus \Omega$

$$-\int_{\mathbb{R}^n \setminus \Omega} \Delta u \varphi = 0 \, .$$

Partielle Integration liefert

$$\int_{\mathbb{R}^n \setminus \Omega} \nabla u \cdot \nabla \varphi - \int_\Gamma \varphi \nabla u \cdot \nu = 0 \, .$$

Nochmalige partielle Integration ergibt

$$-\int_{\mathbb{R}^n \setminus \Omega} u \Delta \varphi + \int_\Gamma u \nabla \varphi \cdot \nu - \varphi \nabla u \cdot \nu = 0 \, . \tag{3.2}$$

Idee: Wenn $\Delta \varphi = 0$, dann bleibt nur eine Gleichung auf Γ übrig. Um diese zu diskretisieren, benötigen wir lediglich eine Triangulierung des Randes Γ statt des Gebietes $\mathbb{R}^n \setminus \Omega$.

Lemma 3.1 (Fundamentallösung)**.** *Sei*

$$\varphi_y(x) := \frac{1}{4\pi} \frac{1}{\|x - y\|} \qquad \text{falls } n = 3,$$
$$\varphi_y(x) := -\frac{1}{2\pi} \log \|x - y\| \qquad \text{falls } n = 2.$$

Dann gilt für $x \neq y$

$$\Delta \varphi_y(x) = 0.$$

Beweis. Zunächst ist $\varphi_y(x) = \varphi_0(x-y)$. Es reicht also zu zeigen, dass $\Delta\varphi_0(z) = 0$ für $z \neq 0$, denn dann

$$\Delta\varphi_y(x) = \Delta\left(\varphi_0(x-y)\right) = \Delta\varphi_0(x-y) = 0.$$

Für $n = 2$ erhalten wir

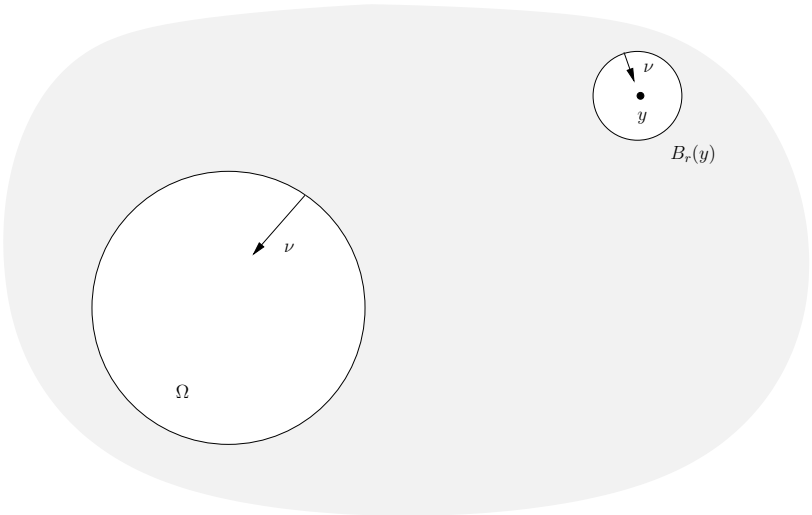
$$-2\pi\nabla\varphi_0(x) = \nabla\left(\log|x|\right) = \frac{1}{|x|}\nabla(|x|) = \frac{1}{|x|}\frac{x}{|x|} = \frac{x}{|x|^2}.$$

und damit

$$\begin{aligned}
-2\pi\mathrm{div}\nabla\varphi_0(x) &= \mathrm{div}\left(x\frac{1}{|x|^2}\right) = \frac{1}{|x|^2}\mathrm{div}x + x\cdot\nabla\left(|x|^{-2}\right) \\
&= \frac{2}{|x|^2} + x\cdot\left(-2\,|x|^{-3}\,\frac{x}{|x|}\right) = \frac{2}{|x|^2} - \frac{2x\cdot x}{|x|^4} \\
&= \frac{2}{|x|^2} - \frac{2}{|x|^2} = 0.
\end{aligned}$$

Ebenso für $n = 3$. □

Wir wollen nun $\varphi = \varphi_y$ in (3.2) wählen und zwar mit $y \in \mathbb{R}^n\backslash\Omega$. Wegen der Bedingung $x \neq y$ müssen wir einen Ball $B_r(y)$ um y aus dem Integrationsgebiet heraus schneiden. Hierbei sei r so klein, dass $B_r(y) \subset \mathbb{R}^n \setminus \Omega$.



Wir erhalten also

$$\begin{aligned}
&-\int_{\mathbb{R}^n\setminus\Omega\setminus B_r(y)} \Delta u\varphi_y = 0\\
\iff &\int_{\mathbb{R}^n\setminus\Omega\setminus B_r(y)} \nabla u\cdot\nabla\varphi_y - \int_{\Gamma\cup\partial B_r(y)} \varphi_y\nabla u\cdot\nu = 0\\
\iff &-\int_{\mathbb{R}^n\setminus\Omega\setminus B_r(y)} u\underbrace{\Delta\varphi_y}_{=0} + \int_{\Gamma\cup\partial B_r(y)} u\nabla\varphi_y\cdot\nu - \int_{\Gamma\cup\partial B_r(y)} \varphi_y\nabla u\cdot\nu = 0\\
\iff &\int_{\Gamma\cup\partial B_r(y)} u\nabla\varphi_y\cdot\nu - \varphi_y\nabla u\cdot\nu = 0\\
\iff &\int_{\Gamma} u\nabla\varphi_y\cdot\nu - \varphi_y\nabla u\cdot\nu + \underbrace{\int_{\partial B_r(y)} u\nabla\varphi_y\cdot\nu}_{=:\mathrm{II}} - \underbrace{\int_{\partial B_r(y)} \varphi_y\nabla u\cdot\nu}_{=:\mathrm{I}} = 0.
\end{aligned}$$

Lemma 3.2. *Seien $B_r(y)\subset\mathbb{R}^n$ und ν innere Normale an $B_r(y)$. Dann gilt für $r\to 0$*

$$\begin{aligned}
&(I) && \int_{\partial B_r(y)} \varphi_y\nabla u\cdot\nu \,\mathrm{d}x \to 0.\\
&(II) && \int_{\partial B_r(y)} u\nabla\varphi_y\cdot\nu \,\mathrm{d}x \to u(y).
\end{aligned}$$

Für den Beweis brauchen wir den folgenden Hilfssatz

Lemma 3.3. *Sei u Lipschitz-stetig. Dann gilt für $r\to 0$*

$$\frac{1}{2\pi r}\int_{\partial B_r(y)} u(x)\,\mathrm{d}x \to u(y).$$

Beweis. Nach Voraussetzung ist

$$\frac{|u(x)-u(y)|}{|x-y|} \le L.$$

Wir zeigen nun

$$\frac{1}{2\pi r}\int_{\partial B_r(y)} u(x)\,\mathrm{d}x - u(y) \to 0.$$

Dazu

$$\begin{aligned}
\left|\frac{1}{2\pi r}\int_{\partial B_r(y)} u(x)\,\mathrm{d}x - u(y)\right| &= \frac{1}{2\pi r}\left|\int_{\partial B_r(y)} u(x)-u(y)\,\mathrm{d}x\right|\\
&\le \frac{1}{2\pi r}\int_{\partial B_r(y)} |u(x)-u(y)|\,\mathrm{d}x \le \frac{L}{2\pi r}\int_{\partial B_r(y)} \underbrace{|x-y|}_{=r}\,\mathrm{d}x\\
&= Lr\frac{1}{2\pi r}\int_{\partial B_r(y)} \mathrm{d}x = Lr \to 0 \text{ für } r\to 0.
\end{aligned}$$

□

Beweis von Lemma 3.2. Für das Integral (I) erhalten wir

$$\begin{aligned}\left|\int_{\partial B_r(y)} \varphi_y \nabla u \cdot \nu\right| &\leq \max_{\partial B_r(y)} |\nabla u \cdot \nu| \int_{\partial B_r(y)} |\varphi_y| \\ &= C\frac{1}{2\pi}\int_{\partial B_r(y)} |\log(r)| = \frac{C}{2\pi}|\log(r)| \int_{\partial B_r(y)} 1 \\ &= Cr\,|\log(r)| \to 0 \text{ für } r \to 0\end{aligned}$$

nach de l'Hospital. Für (II) erinnern wir uns, dass

$$\nabla\varphi_y(x) = -\frac{1}{2\pi}\frac{x-y}{|x-y|^2} \quad \text{und} \quad \nu(x) = -\frac{x-y}{|x-y|}.$$

Damit

$$\begin{aligned}\int_{\partial B_r(y)} u\nabla\varphi_y \cdot \nu &= \frac{1}{2\pi}\int_{\partial B_r(y)} u\frac{x-y}{|x-y|^2}\cdot\frac{x-y}{|x-y|} = \frac{1}{2\pi}\int_{\partial B_r(y)} u\frac{1}{|x-y|} \\ &= \frac{1}{2\pi r}\int_{\partial B_r(y)} u(x)\,\mathrm{d}x \to u(y)\end{aligned}$$

nach Lemma 3.3. □

Damit ergibt sich für $y \in \mathbb{R}^n \setminus \Omega$

$$\int_\Gamma u\nabla\varphi_y \cdot \nu - \varphi_y \nabla u \cdot \nu \,\mathrm{d}x + u(y) = 0,$$

also

$$u(y) = \int_\Gamma \varphi_y \nabla u \cdot \nu - u\nabla\varphi_y \cdot \nu. \tag{3.3}$$

Wir können also u überall in $\mathbb{R}^n \setminus \Omega$ ausrechnen, wenn wir u und $\nabla u \cdot \nu$ auf Γ kennen! In unserem Modellproblem kennen wir bereits $\nabla u \cdot \nu$ auf Γ, wir müssen also noch u auf Γ berechnen. Bevor wir dazu kommen, müssen wir uns kurz mit der Integration singulärer Funktionen beschäftigen

3.2 Exkurs: uneigentlich Integrale

Wir untersuchen zwei Typen von Integralen mit Singularitäten

3.2.1 Schwach singuläre Integranden

Das Integral über den Integranden existiert und ist stetig.
Beispiel: $f(x) = \log|x|$.

- Für $x > 0$ haben wir $f(x) = \log(x)$, die Stammfunktion ist

$$F^+(x) = x\log(x) - x + c.$$

- Für $x < 0$ haben wir $f(x) = \log(-x)$, die Stammfunktion ist

$$F^-(x) = x\log(-x) - x + c.$$

Zusammen also für $x \neq 0$:

$$F(x) = x\log|x| - x + c.$$

Nach de l'Hospital ist $F(x)$ stetig in Null:

$$\lim_{x\to 0} x\log|x| = \lim_{x\to 0} \frac{\log|x|}{x^{-1}} = \lim_{x\to 0} \frac{x^{-1}}{-x^{-2}} = \lim_{x\to 0} -x = 0.$$

Wir können also

$$\int_a^b f(x) = F(b) - F(a)$$

für alle Werte von a und b berechnen.

3.2.2 Stark singuläre Integranden

Hier ist auch die Stammfunktion singulär.
Beispiel: $f(x) = x^{-1}$. Die Stammfunktion

$$F(x) = \log|x| + c$$

ist in Null singulär, beispielsweise existiert also

$$\int_0^b \frac{1}{x} = \lim_{\varepsilon_1\to 0} \int_{\varepsilon_1}^b \frac{1}{x} = \lim_{\varepsilon_1\to 0} F(b) - F(\varepsilon_1)$$

nicht. Allerdings könnten wir für $a, b > 0$ berechnen

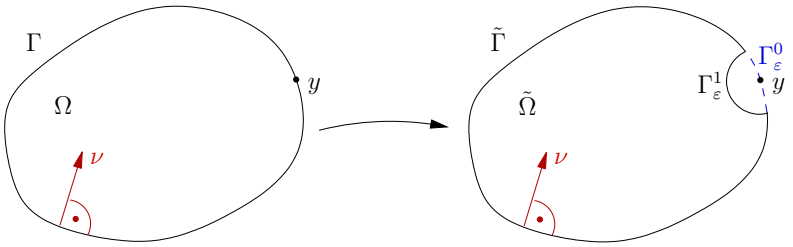
$$\begin{aligned}\int_{-a}^b \frac{1}{x}\,\mathrm{d}x &= \lim_{\varepsilon_1,\varepsilon_2\to 0} \left(\int_{-a}^{-\varepsilon_1} \frac{1}{x}\,\mathrm{d}x + \int_{\varepsilon_2}^b \frac{1}{x}\,\mathrm{d}x\right) \\ &= \lim_{\varepsilon_1,\varepsilon_2\to 0} \left(\left[\log|x|\right]_{-a}^{-\varepsilon_1} + \left[\log|x|\right]_{\varepsilon_2}^b\right) \\ &= \lim_{\varepsilon_1,\varepsilon_2\to 0} \left(\log\frac{\varepsilon_1}{\varepsilon_2}\right) - \log(a) + \log(b).\end{aligned}$$

Wenn wir nun einen bestimmten Grenzwert betrachten, nämlich $\varepsilon_1 = \varepsilon_2 = \varepsilon$ wählen, dann erhalten wir $\log(\varepsilon_1/\varepsilon_2) = \log(1) = 0$. Man bezeichnet den so erhaltenen Wert als den *Cauchy'schen Hauptwert* und schreibt

$$⨍_{-a}^b \frac{1}{x}\,\mathrm{d}x = \log(b) - \log(a),$$

3.3 Lösung auf dem Rand

Nun kommen wir zur Berechnung von u auf Γ. Dazu wollen wir ähnlich wie vorher verfahren, wählen aber nun $y \in \Gamma$. Wie oben müssen wir das Integrationsgebiet modifizieren.



Wir ersetzen den Rand Γ durch $\tilde{\Gamma} := \Gamma \setminus \Gamma_\varepsilon^0 \cup \Gamma_\varepsilon^1$, wobei

$$\Gamma_\varepsilon^0 = \Gamma \cap B_\varepsilon(y) \quad \text{und} \quad \Gamma_\varepsilon^1 = \partial B_\varepsilon(y) \cap \Omega.$$

In dem neuen Gebiet $\tilde{\Omega}$ ist nun $y \in \mathbb{R}^n \setminus \tilde{\Omega}$ und daher können wir das Resultat (3.3) verwenden:

$$u(y) = \int_{\tilde{\Gamma}} \varphi_y \nabla u \cdot \nu - u \nabla \varphi_y \cdot \nu$$

für alle $\varepsilon > 0$. Wir brauchen also die Werte von $\nabla u \cdot \nu$ auf $\tilde{\Gamma}$. Da wir $\nabla u \cdot \nu$ jedoch nur auf Γ kennen, betrachten wir den Limes für $\varepsilon \to 0$ von

$$u(y) = \int_{\Gamma \setminus B_\varepsilon(y)} \varphi_y \nabla u \cdot \nu - u \nabla \varphi_y \cdot \nu + \underbrace{\int_{\Gamma_\varepsilon^1} \varphi_y \nabla u \cdot \nu}_{\text{(I')}} - \underbrace{\int_{\Gamma_\varepsilon^1} u \nabla \varphi_y \cdot \nu}_{\text{(II')}}$$

Wir erhalten ähnlich zu den Rechnungen vorher für $\varepsilon \to 0$

$$\text{(I')} \quad \int_{\Gamma_\varepsilon^1} \varphi_y \nabla u \cdot \nu \to 0$$

$$\text{(II')} \quad \int_{\Gamma_\varepsilon^1} u \nabla \varphi_y \cdot \nu \to -\frac{\theta}{2\pi} u(y).$$

Hierbei gibt $\theta \in (0, 2\pi]$ den Limes der Größe des Kreisbogens Γ_ε^1 an. Für einen vollen Kreis, also $\theta = 2\pi$ erhalten wir das Ergebnis aus Lemma 3.2 zurück. Das Vorzeichen ist anders, da die Normale hier die äußere Normale an $B_\varepsilon(y)$ ist, in Lemma 3.2 jedoch die innere Normale war.
Ist der Rand im Punkt y glatt, so konvergiert Γ_ε^1 gegen einen (kleiner werdenden) Halbkreis, also ist $\theta = \pi$ und der Koeffizient vor $u(y)$ ist $\frac{1}{2}$.
Damit

$$u(y) = \lim_{\varepsilon \to 0} \left(\int_{\Gamma \setminus B_\varepsilon(y)} \varphi_y \nabla u \cdot \nu - \int_{\Gamma \setminus B_\varepsilon(y)} u \nabla \varphi_y \cdot \nu \right) + \frac{\theta}{2\pi} u(y).$$

Da $-2\pi\varphi_y(x) = \log(|x - y|)$ ist, ist das Integral von $\varphi_y \nabla u \cdot \nu$ nur schwach singulär in y, also

$$\lim_{\varepsilon \to 0} \int_{\Gamma \setminus B_\varepsilon(y)} \varphi_y \nabla u \cdot \nu = \int_\Gamma \varphi_y \nabla u \cdot \nu$$

Hingegen ist $2\pi\nabla\varphi_y \cdot \nu = |x-y|^{-1}$ und daher ist das zweite Integral stark singulär in y. Als Grenzwert verwenden wir den Cauchy'schen Hauptwert

$$\lim_{\varepsilon\to 0} \int_{\Gamma\setminus B_\varepsilon(y)} u\nabla\varphi_y \cdot \nu = \mathchoice{\dashint}{\dashint}{\dashint}{\dashint}_\Gamma u\nabla\varphi_y \cdot \nu.$$

Zusammen also

$$\left(1 - \frac{\theta(y)}{2\pi}\right) u(y) = \int_\Gamma \varphi_y \nabla u \cdot \nu - \mathchoice{\dashint}{\dashint}{\dashint}{\dashint}_\Gamma u\nabla\varphi_y \cdot \nu. \tag{3.4}$$

Die Lösung für u erhalten wir also in zwei Schritten:

1. Mit (3.4) können wir u auf Γ berechnen
2. Mit (3.3) können wir dann u in ganz $\mathbb{R}^n \setminus \Omega$ berechnen.

Wir wollen nun (3.4) numerisch lösen.

3.4 Diskretisierung

Es sei Γ eine Fläche im $\mathbb{R}^3$ (oder eine Kurve im $\mathbb{R}^2$). Wir approximieren Γ durch Dreiecke (oder Geradenstücke), $\Gamma_h = \{T_i\}$. Für die Approximation von u wählen wir stückweise konstante Ansatzfunktionen

$$\chi_i(x) = \begin{cases} 1 & : x \in T_i, \\ 0 & : x \notin T_i, \end{cases}.$$

Wir setzen dann

$$u_h = \sum_i U_i\chi_i \quad \text{und} \quad (\nabla u \cdot \nu)_h =: q_h = \sum_i Q_i\chi_i.$$

Für jedes Dreieck T_i wählen wir einen Punkt $y_i \in T_i$. Der Punkt liege im Inneren von T_i, so also ist für ε klein genug $\theta(y_i) = \pi$. Damit erhalten wir

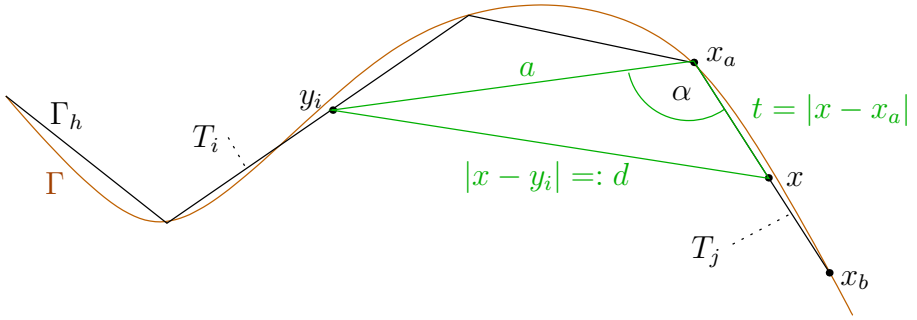
$$\begin{aligned}
\frac{1}{2}U_i &= \int_{\Gamma_h} \varphi_{y_i}(x) \sum_j Q_j\chi_j(x)\,\mathrm{d}x - \mathchoice{\dashint}{\dashint}{\dashint}{\dashint}_{\Gamma_h} \nabla\varphi_{y_i}(x)\cdot\nu \sum_j U_j\chi_j(x)\,\mathrm{d}x \\
&= \sum_j Q_j \int_{\Gamma_h} \chi_j(x)\varphi_{y_i}(x)\,\mathrm{d}x - \sum_j U_j \mathchoice{\dashint}{\dashint}{\dashint}{\dashint}_{\Gamma_h} \chi_j(x)\nabla\varphi_{y_i}(x)\cdot\nu\,\mathrm{d}x \\
&= \sum_j Q_j \underbrace{\int_{T_j} \varphi_{y_i}(x)\,\mathrm{d}x}_{=:V_{ij}} - \sum_j U_j \underbrace{\mathchoice{\dashint}{\dashint}{\dashint}{\dashint}_{T_j} \nabla\varphi_{y_i}(x)\cdot\nu\,\mathrm{d}x}_{=:K_{ij}} \\
&= (VQ)_i - (KU)_i
\end{aligned}$$

Also als Matrixgleichung

$$VQ - (K + \tfrac{1}{2}\mathrm{id})U = 0.$$

3.4.1 Berechnung der Matrixeinträge

Wir berechnen nun die Einträge von V und K.
Für $i \neq j$ hat φ_{y_i} auf T_j keine Singularität, so dass wir im 3-d-Fall eine numerische Quadratur verwenden können. In 2-d können wir die Integrale sogar analytisch berechnen. Mit den folgenden Bezeichnungen



können wir das zweidimensionale Integral über x durch ein eindimensionales Integral über t transformieren. Nach einer Drehung und Verschiebung können wir annehmen, dass x_a im Nullpunkt liegt und dass die Gerade $\overline{x_a x_b}$ auf der x-Achse verläuft, also

$$x_a = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \qquad \text{und} \qquad x_b = \begin{pmatrix} l \\ 0 \end{pmatrix}, l = |x_b - x_a| \,.$$

Nach dem Kosinussatz ist

$$d^2 = a^2 + t^2 - 2at\cos\alpha,$$

wobei alle Werte außer t für die gewählten $i \neq j$ fest sind. Daher ist

$$\begin{aligned} 2\pi V_{ij} &= \int_{T_j} \log \frac{1}{|x - y_i|} \,\mathrm{d}x = \int_{T_j} \log d(x)^{-1} \,\mathrm{d}x = -\tfrac{1}{2} \int_{T_j} \log d(x)^2 \,\mathrm{d}x \\ &= -\tfrac{1}{2} \int_0^l \log(a^2 + t^2 - 2at\cos(\alpha)) \,\mathrm{d}t. \end{aligned}$$

Dieses Integral kann man explizit berechnen und erhält

$$\begin{aligned} 2\pi V_{ij} = -\frac{1}{2}(l - a\cos(\alpha)) \log\left(a^2 - 2al\cos(\alpha) + l^2\right) - \frac{1}{2} a \log\left(a^2\right) \cos(\alpha) \\ + a\sin(\alpha)\tan^{-1}\left(\cot(\alpha) - \frac{l\csc(\alpha)}{a}\right) - a\sin(\alpha)\tan^{-1}(\cot(\alpha)) + l \end{aligned}$$

Für K_{ij} führen wir weitere Bezeichnungen ein:

$$\vec{a} = x_a - y_i, \quad \vec{l} = x_b - x_a \quad \text{und} \quad \gamma^2 = a^2 l^2 - (\vec{a} \cdot \vec{l})^2.$$

Damit ist

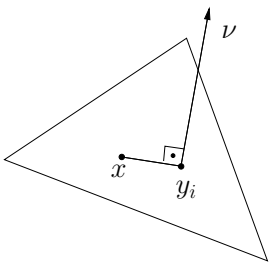
$$K_{ij} = \begin{cases} \dfrac{\vec{a} \cdot \nu}{a(a+1)} & \text{falls } \alpha = 0, \\ \dfrac{l}{\gamma}(\vec{a} \cdot \nu) \left(\arctan\left(\dfrac{l^2 + \vec{a} \cdot \vec{l}}{\gamma}\right) \arctan\left(\dfrac{\vec{a} \cdot \vec{l}}{\gamma}\right) \right) & \text{sonst.} \end{cases}$$

Für $i = j$ haben φ_{y_i} und $\nabla\varphi_{y_i} \cdot \nu$ eine Singularität auf T_i. Wir betrachten zunächst

$$K_{ii} = \oint_{T_i} \nabla\varphi_{y_i}(x) \cdot \nu(x)\, \mathrm{d}x.$$

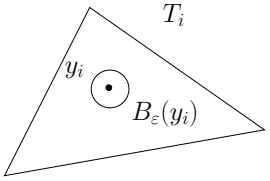
und stellen fest dass in 3-d

$$\nabla\varphi_{y_i}(x) \cdot \nu = -\frac{1}{4\pi r^3}(x - y_i) \cdot \nu = 0,$$



da $\nu \perp (x - y_i)$. Auch im 2-d-Fall ist $\nu \perp (x - y_i)$ und damit verschwindet auch hier das Integral.
Es bleibt also noch V_{ii}. Für ε klein genug passt $B_\varepsilon(y_i)$ in T_i, dann ist

$$\int_{T_i} \varphi_{y_i}(x) = \int_{T_i \setminus B_\varepsilon(y_i)} \varphi_{y_i}(x) + \int_{T_i \cap B_\varepsilon(y_i)} \varphi_{y_i}(x)$$



Im 3-d-Fall ist $T_i \cap B_\varepsilon$ eine Kreisscheibe und der Integrand kann analytisch berechnet werden:

$$\begin{aligned}\int_{T_i \cap B_\varepsilon(y_i)} \varphi_{y_i}(x)\, \mathrm{d}x &= \int_{(T_i - y_i) \cap B_\varepsilon(0)} \varphi_0(x)\, \mathrm{d}x \\ &= \int_{\{z=0\} \cap B_\varepsilon(0)} \varphi_0(x)\, \mathrm{d}x,\end{aligned}$$

nach Verschiebung und wegen Rotationssymmetrie und weiter

$$\begin{aligned}&= \int_0^\varepsilon \int_0^{2\pi} \frac{1}{4\pi r}\, d\alpha\, r\, \mathrm{d}r \\ &= \int_0^\varepsilon \frac{2\pi r}{4\pi r}\, \mathrm{d}r \\ &= \tfrac{1}{2}\varepsilon\end{aligned}$$

Das Integral über $T \setminus B_\varepsilon(y_i)$ ist dagegen problemlos numerisch integrierbar, wobei wir das Dreieck dafür wieder auf ein Referenzdreieck zurück führen.
Im 2-d-Fall kann V_{ii} vollständig analytisch berechnet werden:

$$\begin{aligned}-2\pi V_{ii} = \int_{T_i} \log|x - y_i| &= \int_{-t_1}^{t_2} \log|t| = \int_0^{t_1} \log t + \int_0^{t_2} \log t \\ &= t_1(\log(t_1) - 1) + t_2(\log(t_2) - 1),\end{aligned}$$

wobei $t_1 = |y_i - x_a|$ und $t_2 = |y_i - x_b|$.

Programmieraufgabe 6. *Programmieren Sie einen Randelementelöser. Wenden Sie ihn auf das auf der Vorlesungs-Homepage gegebene Beispiel an.*

4 Finite Differenzen: Wellenfronten in der Seismik

Wir betrachten die Ausbreitung einer Welle in einem elastischen Material. Uns interessiert nur die Bewegung der Wellenfront.

4.1 Die Eikonalgleichung

Idee: Zu $x \in \mathbb{R}^d$ für $d = 2$ oder 3 sei $T(x)$ die (erste) Ankunftszeit der Welle im Punkt x.
Zunächst betrachte ein homogenes Medium.
Beobachtung: Die Niveaumengen von T sind parallel zueinander. Ihr Abstand (der Weg, den die Welle zwischen zwei Zeitpunkten zurückgelegt hat) geteilt durch die Zeitdifferenz ist gleich der Ausbreitungsgeschwindigkeit v.

$$\frac{\Delta x}{\Delta T} = v$$

Aus dieser Beobachtung folgt sofort

$$\boxed{\text{1D:}} \qquad |T'(x)| = \frac{1}{v}.$$

In höheren Dimensionen ist die Steigung senkrecht zur Niveaumenge gleich der Norm des Gradienten, also

$$\boxed{\text{2D/3D:}} \qquad \|\nabla T(x)\| = \frac{1}{v}.$$

Im allgemeinen hängt v vom Material ab, ist also eine (gegebene) Funktion, die von x abhängt.

Problemstellung 4.1 (Eikonalgleichung)**.** *Gegeben sei eine Funktion $v : \Omega \subset \mathbb{R}^d \to \mathbb{R}$ (die Ausbreitungsgeschwindigkeit am jeweiligen Ort) und $\Gamma_0 \subset \Omega$ (die Startfront, z.B. das punktförmige Epizentrum).*
Gesucht ist eine differenzierbare Funktion $T : \Omega \to \mathbb{R}$ mit

$$\|\nabla T(x)\| = \frac{1}{v(x)} \qquad \text{für alle } x,$$
$$T(x) = 0 \qquad \text{für } x \in \Gamma_0.$$

Die Front zur Zeit t ist dann

$$\Gamma_t = \{x \in \Omega \mid T(x) = t\}.$$

4.2 Die Fast-Marching-Methode

Ansatz: Approximiere Ω durch ein kartesisches Gitter der Gitterweite h, zerlege die Gitterpunkte in die drei Teilmengen

- **fertig:** Punkte, für die T bekannt ist;
- **nah:** deren Nachbarn, für diese ist eine obere Schranke für T bekannt;
- **weit:** der Rest, hier ist keine Abschätzung für T bekannt.

Dabei entspricht die Abschätzung von T für Punkte in **nah** der Ankunftszeit der Wellenfront, wenn man nur mögliche Wege der Welle berücksichtigt, die vollständig durch **fertig** verlaufen.

In jedem Schritt wird ein neuer Punkt nach **fertig** verschoben. Dies ist der Punkt in **nah** mit kleinster Zeit. Da für diesen keine kürzeren Wege mehr hinzukommen können, ist dessen Zeitabschätzung korrekt. Dadurch entstehen jedoch für die anderen Punkte in **nah** neue mögliche Wege, so dass deren Zeiten ggf. verringert werden.
Berücksichtigt werden müssen nur die Nachbarn dieses Punktes, da für alle anderen Punkte die Ankunftszeit zu einem späterem Zeitpunkt noch einmal aktualisiert wird.

Start:

- **fertig:** Punkte $x \in \Gamma_0$, setze $T(x) = 0$.
- **nah:** deren Nachbarn x, setze zunächst $T(x) = \frac{h}{v(x)}$.

Schritt:

- Finde den Punkt in **nah** mit kleinstem Wert für T.
- Verschiebe diesen nach **fertig** und fixiere seinen Wert für T.
- Verschiebe dessen Nachbarn, die in **weit** liegen, nach **nah**.
- Aktualisiere Ankunftszeit für alle seine Nachbarn in **nah**.

Wie sieht der Aktualisierungsschritt aus?

Sei T_N **nah** und T_F **fertig**.

1D: $T_N = T_F + \frac{h}{v_N}$
Falls es zwei fertige Punkte gibt, nimmt man den mit kleinerem T_F, denn von dort kommt die Welle zuerst an.

2D:
Falls in einer Richtung beide Nachbarn (links und rechts bzw. oben und unten) fertig sind, nimm jeweils den mit kleinerer Zeit. Verbleibt nur ein Nachbar, verfahre wie in 1D.

Verbleiben zwei Nachbarn mit Zeiten T_X und T_Y:
Falls $T_X + \frac{h}{v_N} < T_Y$, setze $T_N = T_X + \frac{h}{v_N}$.
Falls $T_Y + \frac{h}{v_N} < T_X$, setze $T_N = T_Y + \frac{h}{v_N}$.
Sonst: Der Ansatz (Approximation von $\|\nabla T\|$)

$$\sqrt{\left(\frac{T_N - T_X}{h}\right)^2 + \left(\frac{T_N - T_Y}{h}\right)^2} = \frac{1}{v_N}$$

liefert (die andere Lösung der quadratischen Gleichung ist offensichtlich nicht relevant, da T_N dann kleiner als eine der beiden Zeiten T_X oder T_Y wäre)

$$T_N = \frac{T_X + T_Y}{2} + \sqrt{\frac{h^2}{v_N^2} - \left(\frac{T_X - T_Y}{2}\right)^2}.$$

Wegen der Fallunterschiedung ist $|T_X - T_Y| < \frac{h}{v_N}$ und damit die Diskriminante unter der Wurzel positiv.

Die Fast-Marching-Methode ist eine Variante des Dijkstra-Algorithmus.
In der Implementierung ist es wichtig, für die Menge der nahen Knoten eine geeignete Datenstruktur zu verwenden, die ein effizientes Finden des kleinsten Elements unterstützt (z.B. ein Heap).