Erzeugung glatter Flächen aus Meßpunkten

Ulrich Dietz

Fachbereich Mathematik, Technische Hochschule Darmstadt, D-64289 Darmstadt

Kurzfassung

Es wird eine Methode zur Erzeugung glatter parametrischer Tensorproduktflächen aus diskreten Meßpunkten vorgestellt. Es handelt sich dabei um eine Erweiterung der einfachen 'least square'-Approximation um Glättungs- und Regularisierungsterme. Der iterative Algorithmus besteht aus linearen Approximationsschritten und nichtlinearen Parameterkorrekturen. Aufgrund der Linearität der Approximation ist das Verfahren schnell und numerisch stabil. Es ist insbesondere für extrem dünne Punktmengen oder auch für Punktmengen mit stark schwankender Punktdichte geeignet. Am Ende des Textes werden verschiedene Anwendungsbeispiele präsentiert.

1. Einleitung

In den technischen Anwendungen trifft man immer wieder auf das Problem, zu einer gegebenen Form (Gehäuse, Linse, Autokarosserie, etc.) eine mathematische Repräsentation zu finden für die Weiterverarbeitung in einem CAD-System. Zu diesem Zweck wird das gewünschte Objekt diskret abgetastet; das kann taktil oder berührungslos mit Lasern geschehen.

Das Ergebnis dieser Meßwerterfassung sind i. allg. sehr große unstrukturierte 3D-Punktmengen. Die Punktmengen können von stark schwankender Dichte sein; sind die Messungen teuer oder schwierig, wird die Punktmenge sehr dünn sein.

Die mathematische Aufgabe besteht nun darin, zu diesen Punktmengen glatte Flächen zu konstruieren mit möglichst geringem Abstand. Die klassischen Approximationsverfahren liefern für die oben beschriebenen Punktmengen keine zufriedenstellenden Ergebnisse; die Approximationsflächen sind meist stark wellig.

Sämtliche Verfahren, die das Problem angehen, lassen sich grob in 2 Klassen einteilen:

- Zweischrittverfahren: Ein klassisches Approximationverfahren gefolgt von einem Glättungsverfahren
- Einschrittverfahren: Approximation mit integrierter Glättung

Im industriellen Einsatz findet man vorwiegend Verfahren des ersten Typs; die vorgestellte Methode gehört zum zweiten Typ von Verfahren.

Viele Autoren haben sich schon mit dem Problemkreis glatte Approximation/Glättung befaßt. Arbeiten, die sich dem ersten Typ zuordnen lassen sind z.B.:

- Glättung mit Hilfe von Parameterelimination [Hoschek, 1985]
- Glättung mit Reflexionslinien [Klass, 1980], [Kaufmann & Klass, 1988]
- Glättung über Knotenentfernen [Sapidis & Farin, 1990]
- Lokale Energieglättung [Eck & Hadenfeld, 1995], [Hadenfeld, 1995]

Vertreter der zweiten Klasse von Verfahren sind:

- Kurven-/Netzinterpolation mit Nebenbedingungen [Nowacki et al., 1992]
- Variationsansätze zum Design glatter Flächen [Hagen & Santarelli, 1992]
- Generierung von Flächen mit vorgeschriebener Krümmung [Andersson et al., 1988]
- Glatte Approximation von Flächen über Variationsansatz [Greiner, 1994]
- Glatte Kurven/Flächen als Lösung eines allgemeinen nichtlinearen Optimierungsproblems [Laurent-Gengoux & Mekhilef, 1993]

Diese Zusammenstellung erhebt keinen Anspruch auf Vollständigkeit. Ein Überblick über Glättungsverfahren wird in [Nowacki, 1990] gegeben.

2. Problemstellung und Glättebegriff

Gegeben sind ungeordnete, mit Meßfehlern behaftete Raumpunkte

$$P_m \in I\!R^3, \quad m = 0, \ldots, N-1. \tag{1}$$

Gesucht sind eine parametrische Tensorproduktfläche

$$X(u,v) \in I\!R^3 \tag{2}$$

und zu jedem Datenpunkt P_m Punktparameterwerte

(u_m, v_m), so daß die Abstände

$X(u_m, v_m) - P_m, \quad m = 0, \ldots, N-1$ klein werden

und die Fläche

$X(u,v)$ möglichst 'glatt' ist.

Die Fehler sollen hier bezüglich der Fehlerquadratnorm

$$Q = \sum_m (X(u_m, v_m) - P_m)^2 \tag{3}$$

minimiert werden.

Es stellt sich die Frage nach einem geeigneten Glättebegriff. In [Klass, 1980] wird die Glattheit einer Fläche über deren Reflexionslinien definiert. [Andersson et al., 1988] konstruieren glatte Flächen über die Vorgabe gewünschter Krümmungen. In [Poeschl, 1984] werden Flächenirregularitäten mit Hilfe von Isophoten aufgedeckt. Die Glattheit einer Fläche läßt sich diesem Ansatz folgend auch über den stetigen und gleichmäßigen Verlauf verschiedenster Arten von Isolinien (Isophoten, Isophengen, Linien gleicher gaußscher Krümmung, etc.) einführen.

Ein verbreitetes Vorgehen ist, solche Flächen als glatt zu bezeichnen, die gewisse Energieintegrale bzw. Glättungsfunktionale minimieren [Nowacki, 1990], [Sinha & Schunck, 1992], [Hagen & Santarelli, 1992], [Greiner, 1994].

Diesem Vorgehen folgend, wird in diesem Text eine Fläche als glatt angesehen, wenn sie gewisse Funktionale minimiert. Die verwendeten Funktionale basieren auf Integralen der partiellen Flächenableitungen; sie werden so gewählt, daß sie quadratisch von den Unbekannten abhängen. Dadurch wird das Problem linear.

Nachdem nun der Glättebegriff festgelegt ist, muß noch präzisiert werden, wie minimiert werden soll. Die Bestimmung der 'glattesten' Fläche, die eine vorgegebene Fehlerschranke unterschreitet, führt auf ein nichtlineares Problem (Ansatz mit Lagrange-Multiplikatoren).

Sucht man dagegen diejenige Fläche, die ein Funktional Q' minimiert, das in einem festen Verhältnis Fehlerquadratsumme und Glättungsterme enthält, so wird das Problem linear; vorausgesetzt wird hier die quadratische Abhängigkeit der Glättungsfunktionale von den Unbekannten.

Ein Verfahren letzteren Typs wird in den nächsten Abschnitten vorgestellt.

3. Approximationsalgorithmus ohne Glättung

Das Problem, zu einer gegebenen Punktmenge $P_m \in \mathbb{R}^3,\ m = 0, \ldots, N-1$ eine Tensorproduktfläche $X(u, v) = \sum_{ij} a_{ij}\varphi_i\sigma_j$ und Parameterwerte $(u_m, v_m),\ m = 0, \ldots, N-1$ zu finden, so daß die Fehlerquadratsumme minimal wird, ist nichtlinear.

Der direkte Ansatz über nichtlineare Minimierungsmethoden hat den Nachteil hoher Rechenzeiten [Laurent-Gengoux & Mekhilef, 1993]. Die geometrische Information wird dabei nicht verwendet.

In [Grossman, 1970] wird ein iterativer Prozeß vorgeschlagen, in dem abwechselnd das lineare Approximationsproblem und eine Parameterkorrektur ausgeführt werden. Diese Grundidee findet man später bei verschiedenen Autoren wieder; die Parameterkorrektur wird dort weiter verbessert [Pratt, 1985], [Hoschek et al., 1988], [Rogers & Fog, 1989].

Der Algorithmus läßt sich schematisch folgendermaßen darstellen:

Algorithmus

(A) Bestimmung einer Startnäherung

$$(u_m, v_m)^0, \quad m = 0, \ldots, N-1$$

für die Punktparameterwerte. Setze $k := 0$

(B) Lösen des linearen Approximationsproblems:

$$\sum_m (X((u_m, v_m)^k) - P_m)^2 \overset{a_{ij}}{\to} \min$$

Setze $k := k + 1$

(C) Berechnung einer neuen Näherung

$$(u_m, v_m)^k, \quad m = 0, \ldots, N-1$$

für die Parameterwerte

(D) Sprung nach (B) bis Abbruchbedingung erfüllt ist

Die Bestimmung der Startnäherung ist im nächsten Abschnitt 'Parametrisierung' beschrieben. Die Formulierung und Lösung des linearen Approximationsproblems wird im Abschnitt 'l^2-Approximation' behandelt; verschiedene Verfahren zur Berechnung neuer Näherungen für die Parameterwerte werden im Abschnitt 'Parameterkorrektur' vorgestellt.

Der obige Algorithmus kann in fast unveränderter Form zur Erzeugung glatter Flächen verwendet werden. Dazu wird in Schritt (B) die Fehlerquadratsumme um Glättungsfunktionale erweitert.

Als Abbruchbedingung in Schritt (D) kommen eine maximale Iterationsanzahl, das Unterschreiten einer vorgegebenen Fehlerschranke oder das Erreichen eines lokalen Minimums in Betracht.

4. Parametrisierung

Unter Parametrisierung soll hier das Finden einer initialen Parameterbelegung $(u_m, v_m)^0$, $m = 0, \ldots, N-1$ für eine Punktmenge P_m, $m = 0, \ldots, N-1$ zur Approximation durch eine Tensorproduktfläche verstanden werden. Diese initiale Parameterbelegung sollte eine Näherung für die Parameterwerte der Lotfußpunkte (Projektion in Richtung der Flächennormalen) auf der unbekannten Approximationsfläche sein.

Denn durch die Zuordnung der Parameterwerte zu den Punkten P_i wird festgelegt, welche

Flächenregion welche Punkte annähert.

Zur Parametrisierung gibt es verschiedene Strategien:

- Hat man zur eigentlichen Punktmenge noch zusätzlich 4 Randpunktmengen gegeben, so bietet sich an, diese als erstes durch 4 Kurven zu approximieren. In diese Kurven spannt man eine bilineare Coons-Fläche ein. Die Parameterwerte der Lotfußpunkte auf der Coons-Fläche dienen dann als Startparametrisierung.
- Sind die Punkte so strukturiert, daß sie sich in einer Matrix anordnen lassen, dann läßt sich die Parameterbelegung auf den Kurvenfall zurückführen. Die Zeilen und die Spalten werden dann jeweils uniform (chordal, zentripetal) parametrisiert. Die Zeilen- bzw. Spaltenparameterwerte ergeben dann jeweils die u- bzw. v-Parameterwerte.
- Kennt man von der betreffenden Punktmenge die ungefähre Gestalt und besitzt diese eine bekannte Parameterdarstellung (z.B. Ebene, Kugel, Zylinder, etc.), so läßt sich durch Projektion der Punkte auf diese Fläche eine Parametrisierung bestimmen. Die Parameterwerte der Lotfußpunkte auf der Fläche dienen dann als Parameterwerte für die Approximationspunkte.

All diese Parametrisierungsstrategien sind mit dem Mangel behaftet, daß sie nicht universell einsetzbar und nicht affin invariant sind. Die Parametrisierung hat einen großen Einfluß auf das spätere Approximationsergebnis. Ist die Startparametrisierung schlecht, dann wird das im vorigen Abschnitt beschriebene iterative Verfahren gegen ein unerwünschtes lokales Minimum streben und darin hängen bleiben.

5. l^2-Approximation

Für eine feste Parametrisierung ist das Flächenapproximationsproblem linear:

Gegeben:

$$P_m \in \mathbb{R}^3, \quad (u_m, v_m) \in \mathbb{R}^2, \quad m = 0, \ldots, N-1 \tag{4}$$

Gesucht:

$$a_{ij} \in \mathbb{R}^3, \quad \text{so daß} \quad \sum_m (X_m - P_m)^2 \to \min \quad \text{mit} \quad X_m = \sum_{i=0}^{n_u} \sum_{j=0}^{n_v} a_{ij} \varphi_i(u_m) \sigma_j(v_m) \tag{5}$$

Die a_{ij} sind die Kontrollpunkte der gesuchten Fläche; die φ_i bzw. σ_j bilden jeweils ein System von linear unabhängigen Basisfunktionen.

Die Lösung bestimmt sich aus folgendem linearen Gleichungssystem, den sogenannten Normalgleichungen:

$$\sum_{ij} a_{ij} \sum_m \varphi_i \sigma_j \varphi_r \sigma_s = \sum_m P_m \varphi_r \sigma_s, \quad r = 0, \ldots, n_u; \quad s = 0, \ldots, n_v \tag{6}$$

Die Matrix des Gleichungssystems ist symmetrisch und positiv definit; bei Verwendung der B-Spline-Basis besitzt sie außerdem Bandstruktur. Betrachtet man das überbestimmte System der Fehlergleichungen

$$\sum_{ij} a_{ij}\varphi_i\sigma_j = P_m, \quad m = 0, \ldots, N-1 \tag{7}$$

und bezeichnet dessen Matrix mit M, so kann man die Normalgleichungen formal erhalten, wenn man die Fehlergleichungen von links mit M^T multipliziert:

$$M^T M a = M^T P \tag{8}$$

Die Matrix $M^T M$ des Normalgleichungssystems kann besonders speichereffizient als Summe der N Matrizen $M_i^T M_i, i = 0, \ldots, N-1$ vom Rang 1 aufgebaut werden. M_i ist die i-te Zeile von M und enthält gerade alle Basisfunktionen ausgewertet an der Stelle (u_i, v_i). $M_i^T M_i$ bezeichnet das dyadische Produkt des Vektors M_i mit sich selbst. Von der Matrix M muß also nie mehr als eine Zeile im Speicher stehen.

6. B-Spline-Basis

Die Verwendung der B-Splines als Basisfunktionen erlaubt eine besonders kompakte Beschreibung der Flächen. Eine Darstellung der Grundlagen von B-Spline-Kurven und B-Spline-Flächen findet man in [Hoschek & Lasser, 1992]. Im folgenden wird für B-Splines die Notation dieses Buches verwendet.

Eine Tensorprodukt-B-Spline-Fläche ist dann gegeben durch:

$$X(u,v) = \sum_{i=0}^{n_u} \sum_{j=0}^{n_v} d_{ij} N_{ik}(u) N_{jl}(v) \tag{9}$$

Die vektorwertigen Größen $d_{ij} \in \mathbb{R}^3$ heißen Kontrollpunkte und bilden ein polygonales Netz, das sogenannte Kontrollnetz der Fläche. Dessen Aussehen gibt einen guten Eindruck von der Flächengestalt. Die Funktion N_{ik} ist die i-te B-Spline-Basisfunktion der Ordnung k.

Besondere Vorteile dieser Flächendarstellung sind der lokale Einfluß der Kontrollpunkte auf die Fläche und die Entkopplung von Polynomgrad und Anzahl der Freiheitsgrade. Die B-Splines sind nach Konstruktion an den Segmentgrenzen C^{k-2}-stetig (mit k Ordnung der B-Splines). Diese Stetigkeit überträgt sich auf die TP-Fläche; in den Approximationsprozeß müssen daher keine zusätzlichen Übergangsbedingungen eingearbeitet werden.

Eine Konsequenz der Lokalität der B-Splines ist ein signifikant reduzierter Aufwand beim Aufstellen und Lösen der Normalgleichungen. Baut man die Fehlergleichungsmatrix M auf als

$$M = (m_{ij}), \quad i = 0, ..., N-1; \quad j = 0, \ldots, (n_u+1)(n_v+1) - 1 \tag{10}$$

mit

$$m_{ij} = N_{rk}(u_i)N_{sl}(v_i) \quad \text{und} \quad j = s + r(n_v + 1); \;\; r = 0, \ldots, n_u; \;\; s = 0, \ldots, n_v, \qquad (11)$$

dann ergeben sich für festes Parameterwertpaar (u_i, v_i) genau k Teilvektoren der Länge l, die verschieden von Null sind.

Für die Teilmatrizen $M_i^T M_i$ des vorigen Abschnitts bedeutet das, daß nur k^2 $(l \times l)$-Submatrizen nicht verschwinden. Es müssen also nur $k^2 l^2$ Einträge von $(n_u + 1)^2(n_v + 1)^2$ Einträgen berechnet werden. Für ein anwendungstypisches Beispiel $(k = l = 4, n_u = n_v = 9)$ ist das ein Verhältnis von 1:40.

Die Matrix $M^T M$ des Normalgleichungssystems nimmt darüber hinaus Bandstruktur an. Die Anzahl der nicht verschwindenden oberen Nebendiagonalen ist dabei $(l - 1) + (k - 1)(n_v + 1)$. Daher kann das Normalgleichungssystem schnell und stabil mit Verfahren für symmetrische, positiv definite Bandmatrizen gelöst werden.

Neben all diesen Vorteilen ist die einfache l^2-Approximation mit Tensorprodukt-B-Spline-Flächen mit einigen Nachteilen behaftet. Um kleine Approximationsfehler zu erhalten, muß man i. allg. eine hohe Patchanzahl wählen. Für steigende Patchanzahl zeigt jedoch die B-Spline-Approximation eine Tendenz zu Welligkeiten.

Wird die Patchanzahl so groß gewählt, daß nicht mehr genügend Punktparameter pro Knotenintervall zu liegen kommen, dann treten Spitzen oder Ausreißer in der Approximationsfläche auf. Bedingt wird dieses Verhalten durch die Lokalität der B-Spline-Basisfunktionen; die Matrix des Normalgleichungssystems wird dann fast singulär. Im Grenzfall wird die Matrix singulär; die Approximationsfläche ist nicht mehr eindeutig bestimmt.

Aber auch bei relativ kleiner Patchanzahl kann man Flächen mit Spitzen erhalten. Das ist dann der Fall, wenn die Punktdichte sehr ungleichmäßig ist; lokal treten die gleichen numerischen Probleme auf wie bei hoher Patchanzahl.

Weiterhin geht für hohe Polynomgrade die geometrische Bedeutung des Kontrollnetzes verloren (wahrscheinlich durch die dann schlechte Kondition des Systems). Das Kontrollnetz fängt dann an, wild zu oszillieren.

Durch Einführung von Glättungstermen und Regularisierungstermen lassen sich all diese Nachteile weitgehend beseitigen.

7. Erweiterte Approximation

Bei der einfachen l^2-Approximation wird die Fehlerquadratsumme

$$Q = \sum_m ((X(u_m, v_m) - P_m)^2 \qquad (12)$$

minimiert. Die Fläche X sei im folgenden wieder gegeben durch

$$X(u, v) = \sum_i \sum_j a_{ij} \varphi_i \sigma_j \qquad (13)$$

mit Kontrollpunkten a_{ij} und linear unabhängigen Basisfunktionen φ_i bzw. σ_j.

Das Funktional Q wird nun durch ein Funktional Q' ersetzt:

$$Q' = Q + \sum_i \alpha_i Q_i(X) \tag{14}$$

Die Q_i sind Glättungs- bzw. Regularisierungsterme. Der Einfluß der verschiedenen Terme Q_i wird über die skalaren Design-Parameter α_i gesteuert.

Unter den Funktionalen Q_i sind die quadratisch von den Kontrollpunkten a_{ij} abhängigen besonders interessant, denn diese führen auf lineare Gleichungssysteme.

Ein notwendiges und hinreichendes Kriterium für ein lokales Minimum von Q' ist:

$$\frac{\partial Q'}{\partial d_{rs}} = \frac{\partial(Q + \sum_i \alpha_i Q_i)}{\partial d_{rs}} = 0 \quad \Leftrightarrow \quad \frac{\partial Q}{\partial d_{rs}} + \sum_i \alpha_i \frac{\partial Q_i}{\partial d_{rs}} = 0, \tag{15}$$

$$r = 0, \ldots, n_u; \quad s = 0, \ldots, n_v$$

Das ergibt ein lineares Gleichungssystem der Form:

$$(A + \sum_i \alpha_i A_i)a = b + \sum_i b_i \tag{16}$$

$$A'a = b' \tag{17}$$

Der Vektor a enthält alle unbekannten Kontrollpunkte. Die Vektoren b, b_i, b' sind die rechten Seiten der betrachteten Gleichungssysteme. Die Matrix A' läßt sich also als Summe der Normalgleichungsmatrizen der Funktionale Q und Q_i aufbauen.

Die folgenden Integrale hängen alle quadratisch von den Kontrollpunkten ab und führen somit auf lineare Probleme. Gewichtet mit den α_i's bilden diese Integrale die neue Zielfunktion.

$$Q_1 = \int (grad\ X)^2\, du\, dv = \int (X_u^2 + X_v^2)\, du\, dv \tag{18}$$

Eine Minimierung des Terms Q_1 liefert Flächen mit kurzen partiellen Ableitungsvektoren. Der Flächeninhalt dieser Flächen wird i. allg. klein sein, wenn er auch nicht direkt durch das Funktional minimiert wird.

$$Q_2 = \int (X_{uu} + X_{vv})^2 - 2(1-\mu)(X_{uu}X_{vv} - X_{uv}^2)\, du\, dv \tag{19}$$

Q_2 stellt näherungsweise die Biegeenergie einer dünnen Platte dar und ist schon von verschiedenen Autoren zur Glättung von Flächen eingesetzt worden z.B. [Greiner, 1994], [Hadenfeld, 1995] und dortige Referenzen. Dieser Term eignet sich gut, Welligkeiten der Fläche zu vermindern, rundet aber auch vorgegebene Kanten aus.

$$Q_3 = \int (grad\ div\ grad\ X)^2\, du\, dv = \int (X_{uuu} + X_{uvv})^2 + (X_{uuv} + X_{vvv})^2\, du\, dv \tag{20}$$

Dieses Funktional wurde von Greiner [Greiner, 1994] vorgeschlagen, um auch die 3-ten Ableitungen in den Glättungsprozeß miteinzubeziehen.

$$Q_4 = \sum (d_{i+1j} - d_{ij})^2 + (d_{i-1j} - d_{ij})^2 + (d_{ij+1} - d_{ij})^2 + (d_{ij-1} - d_{ij})^2 \tag{21}$$

Für diesen Term muß vorausgesetzt werden, daß die Kontrollpunkte eine geometrische Bedeutung haben, wie etwa bei Bézier- oder B-Spline-Flächen.

Der Audruck Q_4 wurde in [Laurent-Gengoux & Mekhilef, 1993] zur Formulierung eines allgemeinen Optimierungsproblems benutzt. Die Summe ist so zu verstehen, daß über i, j so aufsummiert wird, daß keines der Quadrate mehrfach auftritt. Die Minimierung von Q_4 führt zu regulären, straffen Kontrollnetzen. Insbesondere bei der Approximation mit hohem Polynomgrad sollte der Term eingesetzt werden, um die sonst auftretenden wild oszillierenden Kontrollnetze zu vermeiden.

Allen Funktionalen Q_i ist gemeinsam, daß sie von der speziellen Parametrisierung einer Fläche abhängen. Wünschenswert wären Funktionale, die parametrisierungsinvariant sind, also nur von der Geometrie der Fläche abhängen. Solche Funktionale hängen aber nicht quadratisch von den Kontrollpunkten ab. Das sich ergebende Gleichungssystem ist dann hochgradig nichtlinear.

Am Beispiel der näherungsweisen Plattenenergie Q_2 für $\mu = 1$ soll das zugehörige Normalgleichungssystem berechnet werden.

$$Q_2 = \int (X_{uu} + X_{vv})^2 \, du \, dv \qquad (\mu = 1) \tag{22}$$

Eine notwendige und hinreichende Bedingung für ein lokales Minimum von Q_2 ist:

$$\frac{\partial Q_2}{\partial d_{kl}} \stackrel{!}{=} 0, \quad k = 0, \ldots, n_u; \quad l = 0, \ldots, n_v \tag{23}$$

Es ergibt sich unmittelbar:

$$\begin{aligned} \frac{\partial Q}{2\partial d_{kl}} &= \int (X_{uu} + X_{vv})\left(\frac{\partial X_{uu}}{\partial d_{kl}} + \frac{\partial X_{vv}}{\partial d_{kl}}\right) du \, dv \\ &= \int \Big(\sum_{ij} a_{ij}\varphi_i''\sigma_j + \sum_{ij} a_{ij}\varphi_i\sigma_j''\Big)(\varphi_k''\sigma_l + \varphi_k\sigma_l'') \, du \, dv \\ &= \sum_{ij} a_{ij} \int \varphi_i''\varphi_k''\sigma_j\sigma_l + \varphi_i''\varphi_k\sigma_j\sigma_l'' + \varphi_i\varphi_k''\sigma_j''\sigma_l + \varphi_i\varphi_k\sigma_j''\sigma_l'' \, du \, dv \stackrel{!}{=} 0 \end{aligned} \tag{24}$$

Mit den Abkürzungen

$$\begin{aligned} U_{ik}^{rs} &= \int \varphi_i^{(r)}\varphi_k^{(s)} du \\ V_{jl}^{rs} &= \int \sigma_j^{(r)}\sigma_l^{(s)} dv \end{aligned} \tag{25}$$

können wir schreiben

$$\sum_{ij} a_{ij}(U_{ik}^{22}V_{jl}^{00} + U_{ik}^{20}V_{jl}^{02} + U_{ik}^{02}V_{jl}^{20} + U_{ik}^{00}V_{jl}^{22}) \stackrel{!}{=} 0; \; k = 0, \ldots, n_u; \quad l = 0, \ldots, n_v \tag{26}$$

Für beliebiges μ ergibt sich das folgende lineare Gleichungssystem:

$$\begin{aligned}&\sum_{ij} a_{ij}(U_{ik}^{22}V_{jl}^{00} + \mu U_{ik}^{20}V_{jl}^{02} + \mu U_{ik}^{02}V_{jl}^{20} + U_{ik}^{00}V_{jl}^{22} + 2(1-\mu)U_{ik}^{11}V_{jl}^{11}) = 0;\\ &k = 0,\ldots,n_u;\ \ l = 0,\ldots,n_v\end{aligned} \tag{27}$$

Ganz analog erhält man für Q_1

$$\sum_{ij} a_{ij}(U_{ik}^{11}V_{jl}^{00} + U_{ik}^{00}V_{jl}^{11}) = 0; \quad k = 0,\ldots,n_u;\ \ l = 0,\ldots,n_v \tag{28}$$

und für Q_3

$$\begin{aligned}&\sum_{ij} a_{ij}(U_{ik}^{33}V_{jl}^{00} + U_{ik}^{31}V_{jl}^{02} + U_{ik}^{13}V_{jl}^{20} + U_{ik}^{11}V_{jl}^{22}\\ &+ U_{ik}^{22}V_{jl}^{11} + U_{ik}^{02}V_{jl}^{31} + U_{ik}^{20}V_{jl}^{13} + U_{ik}^{00}V_{jl}^{33}) = 0;\\ &k = 0,\ldots,n_u;\ \ l = 0,\ldots,n_v.\end{aligned} \tag{29}$$

Das lineare Gleichungssystem für Q_4 soll hier nicht angegeben werden. Es ist aber sofort zu sehen, daß die zugehörige Matrix nur die Einträge $0, 1, -1$ enthält und Bandstruktur besitzt.

Die Matrizen A_i der Gleichungssysteme zu Q_1 bis Q_3 sind symmetrisch und positiv semidefinit. Der Defekt der Matrizen ist größer Null, denn die Ableitungen der Basisfunktionen bilden i. allg. keine linear unabhängige Basis. Geometrisch ist auch klar, daß durch die Funktionale weder Position, noch Orientierung und Ausdehnung der Fläche eindeutig festgelegt werden.

Es soll aber nicht jedes Funktional für sich minimiert werden, sondern der Ausdruck (14). Die Matrix des zugehörigen linearen Gleichungssystems (17) ist i. allg. regulär. Dabei wird vorausgesetzt, daß mindestens ein $\alpha_i > 0$ $(i = 1,\ldots,3)$ und genügend Approximationspunkte gegeben sind, um Lage, Orientierung, etc. festzulegen.

Bei Verwendung der B-Spline-Basis besitzen die Matrizen A_i der zu Q_i gehörigen Gleichungssysteme Bandstruktur und zwar mit derselben Bandbreite wie bei der einfachen l^2-Approximation. Ebenso gelten die gemachten Aussagen über die Anzahl von Null verschiedener Einträge. Es muß daher nur ein kleiner Teil der auftretenden Integrale berechnet werden.

Die Integrale können für die B-Spline-Basis (alle polynomialen Basen) formelmäßig exakt berechnet werden. Für niedrige Polynomgrade empfiehlt sich aber die Verwendung eines numerischen Quadraturverfahrens, z.B. der Gauß-Quadratur; mit wenigen Funktionsauswertungen können hier die exakten Integralwerte berechnet werden, ohne explizite Kenntnis der Stammfunktionen.

Die B-Spline-Basis erlaubt auch eine einfache Berücksichtigung von C^0- und C^1-Stetigkeitsforderungen an den Flächenrändern. Die 1. Ableitungen quer zu einem Rand werden nur von 2 Reihen des Kontrollnetzes festgelegt. Sind diese nun bestimmt durch eine Nachbarfläche, stehen sie für den Approximationsprozeß nicht mehr zur Verfügung. Im linearen Gleichungssystem (17) sind die bekannten Kontrollpunkte einfach auf die rechte Seite zu bringen. Das Problem bleibt also linear.

8. Parameterkorrektur

Im eingangs beschriebenen iterativen Algorithmus zur Erzeugung von Approximationsflächen ist ein Teilschritt die Parameterkorrektur. Die einzige Forderung an die neuen Punktparameterwerte ist, daß das Residuum des minimierten Funktionals kleiner werden muß.

Betrachten wir nun als Funktional die Fehlerquadratsumme. Die Länge der Fehlervektoren wird offensichtlich minimal, wenn diese orthogonal auf der Fläche stehen. Für die Punktparameterwerte $(u_i, v_i) \in I\!R^2,\ i = 0, \ldots, N-1$ bedeutet das, daß sie durch die Parameterwerte $(u_i^*, v_i^*) \in I\!R^2,\ i = 0, \ldots, N-1$ der Lotfußpunkte zu ersetzen sind.

Wir fordern also:

$$Q = \sum_m (X(u_m, v_m) - P_m)^2 \overset{(u_k, v_k)}{\rightarrow} \min \tag{30}$$

Eine notwendige Bedingung für ein solches Minimum ist:

$$\frac{\partial Q}{\partial u_k} = \frac{\partial Q}{\partial v_k} \overset{!}{=} 0, \quad k = 0, \ldots, N-1 \tag{31}$$

Das ist formal ein nichtlineares Gleichungssystem in $2N$ Unbekannten. Das Problem zerfällt jedoch in zweidimensionale *nichtlineare* Unterprobleme. Geometrisch ist klar, daß Parameterwerte eines Lotfußpunktes unabhängig von denen der anderen Lotfußpunkte sind. Die Gleichungssysteme der Unterprobleme sind vom Typ:

$$\begin{aligned} (X-P)X_u &= 0 \\ (X-P)X_v &= 0 \end{aligned} \tag{32}$$

Sie können iterativ z.B. mit dem Gauß-Newton-Verfahren gelöst werden. Die Iterationsvorschrift lautet:

$$\begin{pmatrix} u \\ v \end{pmatrix}^{(k+1)} = \begin{pmatrix} u \\ v \end{pmatrix}^{(k)} - \begin{pmatrix} \Delta u \\ \Delta v \end{pmatrix}^{(k)} \tag{33}$$

mit

$$\begin{pmatrix} \Delta u \\ \Delta v \end{pmatrix}^{(k)} = \begin{pmatrix} X_u^2 & X_u X_v \\ X_u X_v & X_v^2 \end{pmatrix}^{-1} \begin{pmatrix} (X-P)X_u \\ (X-P)X_v \end{pmatrix} \tag{34}$$

Es ergeben sich die Parameterkorrekturterme nach Rogers und Fog [Rogers & Fog, 1989]. Es werden nur die partiellen Ableitungen der Fläche benötigt. Das Verfahren ist lokal konvergent von Ordnung ≥ 1. Bei Verwendung des Newton-Verfahrens zur Lösung erhält man eine Konvergenzordnung ≥ 2, muß aber auch die 2-ten partiellen Ableitungen berechnen.

Die Korrekturterme für das Newton-Verfahren lauten:

$$\begin{pmatrix} \Delta u \\ \Delta v \end{pmatrix}^{(k)} = \begin{pmatrix} S_{uu} & S_{uv} \\ S_{uv} & S_{vv} \end{pmatrix}^{-1} \begin{pmatrix} (X-P)X_u \\ (X-P)X_v \end{pmatrix} \tag{35}$$

mit

$$\begin{aligned} S_{uu} &= X_u^2 + (X-P)X_{uu} \\ S_{vv} &= X_v^2 + (X-P)X_{vv} \\ S_{uv} &= X_u X_v + (X-P)X_{uv} \end{aligned} \tag{36}$$

Setzt man in dieser Darstellung $S_{uv} = 0$, erhält man die Parameterkorrekturterme nach Hoschek [Hoschek et al., 1988].

Allen erwähnten Parameterkorrekturverfahren ist gemeinsam, daß sie nur lokal konvergent sind; bei guter Startparametrisierung konvergieren die Verfahren gegen das globale Optimum.

Minimiert man nicht die Fehlerquadratsumme, sondern das Funktional Q' des letzten Abschnitts, so ändert sich die Parameterkorrektur nicht. Denn die notwendigen Bedingungen für ein Minimum

$$\frac{\partial Q'}{\partial u_k} = \frac{\partial Q'}{\partial v_k} \stackrel{!}{=} 0, \quad k = 0, \ldots, N-1 \tag{37}$$

werden zu

$$\frac{\partial Q}{\partial u_k} = \frac{\partial Q}{\partial v_k} \stackrel{!}{=} 0, \quad k = 0, \ldots, N-1, \tag{38}$$

da die Funktionale Q_i unabhängig von den Parameterwerten (u_k, v_k) sind.

Aus dieser Unabhängigkeit der Funktionale Q_i ergibt sich für die Gesamtiteration des ersten Abschnitts ein weiterer Vorteil. Es muß nicht nach jeder Parameterkorrektur die ganze Matrix des Normalgleichungssystems neu aufgestellt werden. Lediglich die Teilmatrix A, die sich aus der Fehlerquadratminimierung ergibt, muß in jedem Gesamtiterationsschritt neu berechnet werden. Der Integralteil, also die Summe der Matrizen A_i, die zu den Funktionalen Q_i gehören, wird vorab einmal berechnet.

9. Anwendungsbeispiele

9.1. *Beispiel a.)*

Der erste Datensatz besteht aus 41×26 digitalisierten Datenpunkten eines Pedaltopfes; siehe Abbildung(1). Er hat eine sehr reguläre Struktur; die x- und y-Koordinaten liegen in einem festen Raster. Dadurch bedingt existieren an den 2 steilen Flanken des Objektes keine Datenpunkte.

In Abbildung (2) ist das Ergebnis der einfachen l^2-Approximation zu sehen, in Abbildung (3) das Ergebnis mit Minimierung der Biegeenergie. Auf beiden Flächen sind die Isophoten zu einer gemeinsamen Lichtquelle dargestellt. Die Zwischenbereiche sind abwechselnd schwarz und weiß eingefärbt.

Abbildung 1: 41×26 digitalisierte Punkte eines Pedaltopfes

Abbildung 2: l^2-Approximation der Punkte in Abbildung (1) mit 20×13 Patches

Es ist deutlich zu erkennen, daß die zweite Fläche trotz höherer Patchanzahl weit weniger wellig ist. Aufgrund der großen punktfreien Bereiche dieses Datensatzes kann für die einfache l^2-Approximation die Patchanzahl nicht beliebig erhöht werden; das Gleichungssystem wird dann singulär. Bei der erweiterten Approximation mit Energieminimierung kann die Patchanzahl beliebig hoch gewählt werden, um etwaige Fehlerschranken zu halten.

9.2. Beispiel b.)

Im zweiten Beispiel sind 7402 Punkte auf einem Autokotflügel gegeben. Die Punkte verlaufen in gewissen Vorzugsrichtungen, sind aber ansonsten unstrukturiert. Die durch die Punktmenge induzierten Ränder können nicht als Ränder einer Tensorproduktfläche dienen.

Die Approximationsfläche sollte daher über die Punkte hinaus reichen und abschließend getrimmt werden. Die einfache l^2-Approxiamtion kann hierfür nicht eingesetzt werden;

Abbildung 3: Glatte Approximation der Punkte in Abbildung (1) mit 41×26 Patches

Abbildung 4: 7402 Datenpunkte und Patchgrenzen der Approximationsfläche

das zugehörige Gleichungssystem würde singulär.

Die Approximation mit Minimierung der Biegeenergie dagegen liefert eine sehr glatte Fläche mit geringen Fehlern (19×9 Patches bei einem mittleren Fehler von 0.0096 mm). In Abbildung (4) ist die Punktmenge zusammen mit der Patchstruktur zu sehen. In den Bereichen ohne Punkte wird die Fläche allein durch die Energieintegrale gehalten. In Abbildung (5) sind die zugehörigen Reflexionslinien zu sehen; sie verlaufen besonders innerhalb der Trimmkurven sehr gleichmäßig.

9.3. Beispiel c.)

Beim nächsten Datensatz sind Datenpunkte gegeben, die von einer geschlossenen Raumkurve stammen. Es soll nun eine Fläche berechnet werden, die diese Punkte approximiert, die einen kleinen Flächeninhalt hat und glatt ist.

Abbildung 5: Reflexionslinien der Fläche in Abbildung (4)

Abbildung 6: Datenpunkte einer Raumkurve und approximierende 'Minimal'-Fläche

Als Zielfunktion der Approximationsfläche in Abbildung (6) wurde ein gewichteter Term aus Fehlerquadratsumme und $\int (gradX)^2 dudv$ gewählt. Das ganze Flächeninnere wird nur durch das Integral bestimmt. Die Tendenz dieses Funktionals, kleine Flächeninhalte zu generieren, wird in der Abbildung bestätigt.

9.4. Beispiel d.)

Als letztes Beispiel soll eine B-Spline-Approximations-Fläche aus 4 Patches vom Grad 7×7 zu 10×10 Datenpunkten betrachtet werden.

Das Kontrollnetz der einfachen l^2-Approximation ist in Abbildung (7) zu sehen. Es oszilliert wild; von einer geometrischen Bedeutung desselben kann kaum noch gesprochen werden. Die Weiterverarbeitung einer solchen Approximationsfläche in einem CAD-System ist nicht mehr sinnvoll möglich.

Abbildung 7: Oszillierendes Kontrollnetz bei einfacher l^2-Approximation

Abbildung 8: Reguläres Kontrollnetz für die erweiterte Approximation

Verwendet man bei der Approximation den gewichteten Term Q_4 zur Netzregularisierung, erhält man eine Kontrollnetz in der Nähe der Fläche, so wie man es erwarten würde (Abbildung 8). Dabei wird der Approximationsfehler nur wenig größer.

10. Mögliche Erweiterungen

Die eingeführten Glättungsterme wirken auf der ganzen Fläche gleich stark. Das ist jedoch nur für Datensätze geeignet, in denen keine starken Krümmungsänderungen auftreten. Induziert eine Punktmenge eine scharfe Kante in der Fläche, so wird diese mit dem vorgestellten Verfahren 'ausgerundet'. Hier wären datensatz-abhängige Funktionale sinnvoll, die je nach Punktdichte und diskretem Krümmungsverhalten der Punktmenge den Einfluß der Glättungsterme steuern.

Das Verfahren kann leicht erweitert werden, so daß eine vorgegebene Fehlertoleranz ge-

Abbildung 9: Reflexionslinien auf einer Fläche. Aus [Hoschek & Lasser, 1992]

halten werden kann. Dazu wird der ganze Algorithmus wiederholt durchlaufen. In jedem Durchlauf wird die Patchanzahl erhöht, bis die Fehlervorgabe unterschritten wird.

Eine mögliche Verbesserung besteht darin, affin invariante Designparameter einzuführen, die unabhängig von der absoluten Größe der Fehlerquadratsumme und der Skalierung des Datensatzes sind. Auch eine automatische Bestimmung der Designparameter α_i aus dem Datensatz kann das Ziel weiterer Untersuchungen sein.

In der Autoindustrie werden Reflexionslinien eingesetzt, um die Güte einer Fläche zu beurteilen. Als Reflexionslinien bezeichnet man das Bild (meist paraleller) Geraden auf einer Fläche von einem bestimmten Augpunkt aus gesehen. In der Praxis sind die Geraden Leuchtstoffröhren und die Fläche ist eine spiegelnde Autokarosserie.

Die Reflexionslinien könnte man nun direkt zur Erzeugung glatter Flächen einsetzen. Dazu wird ein Objekt entlang von Reflexionslinien abgetastet. Zusätzlich zu den Punktkoordinaten wird der Aug- und der Lichtpunkt erfaßt. In der Optik gilt bekanntermaßen, Einfallswinkel gleich Ausfallswinkel; daher hat man neben den Positionen über die Formel

$$N = \frac{N^*}{|N^*|} \quad \text{mit} \quad N^* = \frac{L-P}{|L-P|} + \frac{A-P}{|A-P|} \tag{39}$$

auch für jeden Punkt eine Normale zur Verfügung.

Wählt man für die Normalenapproximation den direkten Weg

$$Q_n = \sum (X_N - N)^2 \rightarrow \min \tag{40}$$

mit

$$X_N = X_u \times X_v, \tag{41}$$

wird man auf ein nichtlineares Problem geführt. Minimiert man jedoch

$$Q_N = \sum (X_u N)^2 + (X_v N)^2, \tag{42}$$

so erhält man ein lineares Gleichungssystem:

$$\sum_{i=0}^{n_u}\sum_{j=0}^{n_v}\sum_{c=0}^{2} a_{ijc}\sum_m N_{mc}N_{md}(\varphi_i'\sigma_j\varphi_r'\sigma_s+\varphi_i\sigma_j'\varphi_r\sigma_s')=0 \tag{43}$$
$$r=0,\ldots,n_u;\quad s=0,\ldots,n_v;\quad d=0,\ldots,2$$

Mit a_{ijc} wird die c-te Koordinate des Kontrollpunktes a_{ij} bezeichnet; mit N_{mc} die c-te Koordinate der Normale N_m. Die Matrix dieses Gleichungssystems ist ebenfalls symmetrisch und positiv semidefinit. Sie ist singuär, da durch die Normalen eine Fläche nicht eindeutig festgelegt ist. Vom Typ her paßt der Term Q_N zu den betrachteten Glättungs- und Regularisierungstermen, so daß er als weiterer Term direkt in das iterative Verfahren integriert werden kann.

In wie weit damit glatte Flächen erzeugt werden können und wie sich Meßfehler bei der Normalenbestimmung auswirken, wird zur Zeit noch untersucht.

Abschließend soll noch eine Problemstellung aus der optischen Industrie angerissen werden. Bei der Entwicklung von optischen Linsen fallen Datensätze an, in denen zu jedem Meßpunkt noch eine Gaußkrümmung oder mitttlere Krümmung vorliegt. Die Approximation von zusätzlichen Krümmungswerten läßt sich nicht im Rahmen des entwickelten Verfahrens realisieren. Das Problem läßt sich vermutlich auch nicht auf ein lineares Problem zurückführen. Hier müssen geeignete iterative, nichtlineare Methoden, etwa vom Gauß-Newton-Typ, entwickelt werden.

Literatur

[Andersson et al., 1988] Andersson, E.; Andersson, R.; Boman, M.; Elmroth, T.; Dahlberg, B.; Johansson, B.: Automatic construction of surfaces with prescribed shape. CAD 20 (1988), 317-324.

[Eck & Hadenfeld, 1995] Eck, M.; Hadenfeld, J.: Local Energy Fairing of B-Spline Curves. To appear in Farin, G.; Hagen, H. and Noltemeier, H. (eds.): Computing, Supplementum 10, Springer (1995).

[Greiner, 1994] Greiner, G.: Variational Design and Fairing of Spline Surfaces. Computer Graphics Forum 13:3 (1994), 143-154.

[Grossman, 1970] Grossman, M.: Parametric curve fitting. The Computer Journal, vol 14 (1970), 169-172.

[Hadenfeld, 1995] Hadenfeld, J.: Local Energy Fairing of B-Spline Surfaces. To appear in Dæhlen, M.; Lyche, T.; Schumaker, L. L. (eds.): Mathematical Methods in CAGD III, (1995).

[Hagen & Santarelli, 1992] Hagen, H.; Santarelli, P.: Variational Design of Smooth B-Spline Surfaces. Hagen, H. (ed.): Topics in Surface Modeling. SIAM (1992), 85-94.

[Hoschek, 1985] Hoschek, J.: Smoothing of curves and surfaces. CAGD 2 (1985), 97-105.

[Hoschek & Lasser, 1992] Hoschek, J.; Lasser, D.: Grundlagen der geometrischen Datenverarbeitung. Teubner Verlag 1992

[Hoschek et al., 1988] Hoschek, J.; Schneider, F.-J.; Wassum, P.: Optimal approximate conversion of spline surfaces. CAGD 6 (1989), 293-306.

[Klass, 1980] Klass, R.: Correction of local surface irregularities using reflection lines. CAD 12 (1980), 73-77

[Kaufmann & Klass, 1988] Kaufmann, E.; Klass, R.: Smoothing surfaces using reflection lines for families of splines. CAD 20 (1988), 312-316.

[Laurent-Gengoux & Mekhilef, 1993] Laurent-Gengoux, P.; Mekhilef, M.: Optimization of a NURBS representation. CAD 25 (1993), 699-710.

[Nowacki, 1990] Nowacki, H.: Mathematische Verfahren zum Glätten von Kurven und Flächen. Encarnacao, J. L.; Hoschek, J.; Rix, J. (Hrsg.): Geometrische Verfahren der Graphischen Datenverarbeitung. Springer (1990), 22-45.

[Nowacki et al., 1992] Nowacki, H.; Kaklis, P.; Weber, J.: Curve Mesh Fairing and GC^2 Surface Interpolation. Mathematical Modelling and Numerical Analysis 26 (1992), 113-136.

[Poeschl, 1984] Poeschl, T.: Detecting surface irregularities using isophotes. CAGD 1 (1984), 163-168.

[Pratt, 1985] Pratt, M. J.: Smooth parametric surface approximations to discrete data. CAGD 2 (1985), 165-171.

[Rogers & Fog, 1989] Rogers, D. F.; Fog, N. G.: Constrained B-Spline curve and surface fitting. CAD 21 (1989), 641-648.

[Sapidis & Farin, 1990] Sapidis, N.; Farin, G.: Automatic fairing algorithm for B-spline curves. CAD 22 (1990), 121-129.

[Sinha & Schunck, 1992] Sinha, S. S.; Schunck, B. G.: A Two-Stage Algorithm for Discontinuity-Preserving Surface Reconstruction. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol 14, no 1 (1992), 36-55.