

Regularization

Gemini AI Pro 2.5

Q

Biharmonic regularization. Please explain the concept and the formulas here. The text is in norwegian. (Kompendium 9.10.2)

A

This image explains the **biharmonic smoothness penalty** (or biharmonic regularization), a concept used in mathematics, computer graphics, and machine learning to find an optimally “smooth” function.

Here is a breakdown of the concept and the formulas.

Concept: Biharmonic Smoothness Penalty

In many problems, we want to find a function $f(x)$ that fits some data but is also “smooth.” We enforce smoothness by adding a **penalty term** to our optimization problem.

- **A simple penalty** might use the *first* derivative ($f'(x)$). This penalizes steep **slopes**.
- **The biharmonic penalty** is a “stronger” penalty. As the text says, it “penalizes **curvature** rather than just slope.” It does this by using the *second* derivative ($f''(x)$ in 1D, or the Laplacian $\nabla^2 f$ in multiple dimensions).

Minimizing this penalty results in functions that bend as little as possible, leading to very smooth and natural-looking curves or surfaces (this is the principle behind “thin plate splines”).

Formulas Explained

1.

$$S = \int |\nabla^2 f(x)|^2 dx$$

This is the definition of the **biharmonic energy** (S).

- $f(x)$: The function we are trying to find.

- ∇^2 : The **Laplacian operator**. In 1D, $\nabla^2 f$ is just the second derivative, $f''(x)$. In 2D, it's $\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$.
 - $|\dots|^2$: This squares the magnitude of the curvature.
 - $\int \dots dx$: This integrates (sums up) the squared curvature over the function's entire domain.
 - **In short:** This formula calculates a single number S that represents the total “bending” of the function. The goal is to find the function $f(x)$ that makes S as small as possible (while still fitting the data).
2. **The 1D Case:** $\phi_i(x) = x^i$ The text simplifies the problem to one dimension (from 0 to 1) and assumes the function $f(x)$ can be built from a set of **polynomial basis functions**, $\phi_i(x) = x^i$. (For example, $f(x) = c_0 + c_1x + c_2x^2 + \dots$).

3.

$$\phi_i''(x) = i(i-1)x^{i-2}$$

This is the second derivative of the basis function $\phi_i(x)$, found using the power rule:

- **Basis:** $\phi_i(x) = x^i$
- **First derivative:** $\phi_i'(x) = ix^{i-1}$
- **Second derivative:** $\phi_i''(x) = i(i-1)x^{i-2}$

4.

$$B_{ij} = \int_0^1 \phi_i''(x)\phi_j''(x)dx$$

This formula defines the elements of a matrix, B . This matrix is essential for finding the numerical solution. Each element B_{ij} measures the “interaction” of the curvature of one basis function (ϕ_i) with another (ϕ_j).

5.

$$B_{ij} = \int_0^1 i(i-1)x^{i-2}j(j-1)x^{j-2}dx$$

This is just substituting the formula from step 3 into the integral from step 4 for both ϕ_i'' and ϕ_j'' .

6.

$$B_{ij} = i(i-1)j(j-1) \int_0^1 x^{i+j-4}dx$$

This simplifies the previous step:

- The terms $i(i-1)$ and $j(j-1)$ are constants, so they are pulled outside the integral.
- Inside the integral, x^{i-2} and x^{j-2} are multiplied by adding their exponents: $(i-2) + (j-2) = i+j-4$.

7.

$$B_{ij} = \frac{i(i-1)j(j-1)}{i+j-3}$$

This is the final result after solving the integral:

- The integral of x^n is $\frac{x^{n+1}}{n+1}$.
- Here, $n = i + j - 4$. So $n + 1 = i + j - 3$.
- The integral becomes: $\left[\frac{x^{i+j-3}}{i+j-3} \right]$ evaluated from $x = 0$ to $x = 1$.
- Plugging in $x = 1$: $\frac{1^{i+j-3}}{i+j-3} = \frac{1}{i+j-3}$
- Plugging in $x = 0$: $\frac{0^{i+j-3}}{i+j-3} = 0$
- The result of the integral is $\frac{1}{i+j-3} - 0$.
- Multiplying this by the constants from step 6 gives the final formula.

Q

What do you mean by “interaction” here in point 4: Each element B_{ij} measures the “interaction” of the curvature of one basis function (f_i) with another (f_j).

A

That’s an excellent question. “Interaction” is a plain-language term I used to describe a mathematical concept, specifically the **correlation of curvature** between the two basis functions.

Let’s break down the formula for B_{ij} to see what it’s physically doing.

$$B_{ij} = \int_0^1 \phi_i''(x) \phi_j''(x) dx$$

Think of the integral \int as a “sum” over every point x from 0 to 1. At each point, we are multiplying two numbers:

1. $\phi_i''(x)$: The curvature of the i -th basis function at point x .
2. $\phi_j''(x)$: The curvature of the j -th basis function at that *same* point x .

The product $\phi_i''(x) \phi_j''(x)$ measures how the curvatures “behave together” at that single point. The integral B_{ij} sums up this behavior over the entire domain.

This leads to three scenarios:

Scenario 1: High Positive “Interaction” (Large B_{ij})

- Imagine a region where $\phi_i(x)$ is very curvy (its second derivative $\phi_i''(x)$ is large and positive).
- If, in that *same region*, $\phi_j(x)$ is *also* very curvy in the *same direction* (its second derivative $\phi_j''(x)$ is also large and positive), their product will be large and positive.

- The same logic applies if both are large and negative (curving the other way).
 - If this happens consistently across the domain, all these large positive products add up, and B_{ij} becomes a large positive number.
 - **Meaning:** This tells us that ϕ_i and ϕ_j **curve in the same way at the same places**. They are highly correlated.
-

Scenario 2: High Negative “Interaction” (Large Negative B_{ij})

- Imagine a region where $\phi_i(x)$ curves strongly one way (e.g., $\phi_i''(x)$ is large and positive).
 - But in that *same region*, $\phi_j(x)$ curves strongly the *opposite way* (e.g., $\phi_j''(x)$ is large and negative).
 - At these points, their product $\phi_i''(x)\phi_j''(x)$ will be large and *negative*.
 - If they consistently curve in opposite ways, the integral B_{ij} will be a large negative number.
 - **Meaning:** ϕ_i and ϕ_j are **anti-correlated**. Where one bends one way, the other bends the opposite way.
-

Scenario 3: No “Interaction” ($B_{ij} \approx 0$)

- This happens if the curvatures are **uncorrelated**.
- For example, in all the regions where $\phi_i(x)$ is very curvy, $\phi_j(x)$ is flat (so $\phi_j''(x) \approx 0$). Their product is always zero.
- Or, the regions where they are curvy simply don’t overlap.
- **Meaning:** The curvature of ϕ_i and the curvature of ϕ_j are **independent** or “orthogonal.” Bending one has no relationship to bending the other.

Why this Matters

The matrix B is essentially a “map” of these curvature overlaps. When we try to find the smoothest possible function $f(x)$, we are trying to find the right combination of all the ϕ functions. The B matrix tells the solver, “Be careful—if you use a lot of ϕ_i , it’s going to have a strong combined curvature cost with ϕ_j (if B_{ij} is large).” It quantifies the “cross-cost” of bending two basis functions simultaneously.

Q

Dirichlet-energy and regularization. In the formulas below, I don’t understand the use of i and j indices, and why we have a product of two terms after the integral sign. Please clarify. (Kompendium 9.10)

A

This is a common point of confusion when expanding vector products. Let's break it down.

Here are the clarifications for your two questions.

1. Why the Product of Two Terms?

The short answer is that the **squared magnitude** of a vector, $|\mathbf{v}|^2$, is defined as the **dot product** of the vector with itself: $\mathbf{v} \cdot \mathbf{v}$.

In your case, the “vector” is the gradient $\nabla f(x)$.

- The Dirichlet energy is defined as:

$$S = \int |\nabla f(x)|^2 dx$$

- Based on the definition above, we can rewrite

$$|\nabla f(x)|^2 \text{ as } \nabla f(x) \cdot \nabla f(x).$$

- This gives the integral:

$$S = \int (\nabla f(x)) \cdot (\nabla f(x)) dx$$

So, the product of two terms you see is simply the expansion of the $|\nabla f(x)|^2$ term. The first term in the product is $\nabla f(x)$, and the second term is also $\nabla f(x)$.

2. Why Two Different Indices (i and j)?

This is a standard algebraic technique for correctly expanding the product of two sums.

You start with the two terms from above:

$$\nabla f(x) = \sum_i c_i \nabla \phi_i(x)$$

$$\nabla f(x) = \sum_i c_i \nabla \phi_i(x)$$

If you substitute these into the integral, you get:

$$S = \int (\sum_i c_i \nabla \phi_i(x)) \cdot (\sum_i c_i \nabla \phi_i(x)) dx$$

This is a product of two sums. Let's look at a simpler example:

$$(a_1 + a_2) \cdot (b_1 + b_2) = a_1 b_1 + a_1 b_2 + a_2 b_1 + a_2 b_2$$

This is a **double summation**. To write this in sigma notation, you can't use the same index for both sums.

$$(\sum_i a_i) \cdot (\sum_i b_i) \text{ is ambiguous.}$$

The correct way to write it is

$$(\sum_i a_i) \cdot (\sum_j b_j) = \sum_i \sum_j a_i b_j.$$

Applying this to your formula:

1. We have the product:

$$(\sum_i c_i \nabla \phi_i(x)) \cdot (\sum_i c_i \nabla \phi_i(x))$$

2. The index **i** in the first sum is a “dummy variable” that is independent of the index in the second sum.
3. To make this clear and avoid confusion, we **rename the index in the second sum to j**. This doesn’t change its meaning, but it allows us to expand the product correctly.

$$S = \int (\sum_i c_i \nabla \phi_i(x)) \cdot (\sum_j c_j \nabla \phi_j(x)) dx$$

4. Now we can expand this product into a **double summation**, multiplying every term from the first sum (indexed by **i**) by every term from the second sum (indexed by **j**):

$$S = \int \sum_i \sum_j (c_i \nabla \phi_i(x)) \cdot (c_j \nabla \phi_j(x)) dx$$

5. Finally, you can pull the constant coefficients (c_i, c_j) and the summations out of the integral (since integration is a linear operation):

$$S = \sum_i \sum_j c_i c_j \int \nabla \phi_i(x) \cdot \nabla \phi_j(x) dx$$

This final line matches the step in your image right before the definition of B_{ij} . The use of **i** and **j** is essential to get this correct double-summation structure.

Q

Please give an example with simple basis functions, and show all the calculations.

A

Here is a simple, complete example in one dimension (1D).

In 1D, the math simplifies, but the logic is identical:

- The domain is an interval, let’s use $x \in [0, 1]$.
- The gradient $\nabla f(x)$ becomes the first derivative $f'(x)$.
- The dot product $\nabla \phi_i(x) \cdot \nabla \phi_j(x)$ becomes a simple product $\phi'_i(x) \phi'_j(x)$.
- The integral becomes $S = \int_0^1 (f'(x))^2 dx$.

Step 1: Define the Problem

Let’s use a function $f(x)$ built from just **two simple polynomial basis functions**:

- $\phi_1(x) = x$
- $\phi_2(x) = x^2$

Our function is a linear combination of these:

$$f(x) = \sum_i c_i \phi_i(x) = c_1 \phi_1(x) + c_2 \phi_2(x) = c_1 x + c_2 x^2$$

Our goal is to find the matrix \mathbf{B} such that the Dirichlet energy S is given by:

$$S = \mathbf{c}^T \mathbf{B} \mathbf{c} = \begin{bmatrix} c_1 & c_2 \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$$

Step 2: Calculate the Gradients (Derivatives)

We need the derivative of each basis function:

- $\phi'_1(x) = \frac{d}{dx}(x) = 1$
 - $\phi'_2(x) = \frac{d}{dx}(x^2) = 2x$
-

Step 3: Calculate Each Element of the Matrix B

The formula for each element is $B_{ij} = \int_0^1 \phi'_i(x) \phi'_j(x) dx$. We have four elements to calculate.

Element B_{11} (i=1, j=1):

$$B_{11} = \int_0^1 \phi'_1(x) \phi'_1(x) dx = \int_0^1 (1) \cdot (1) dx = \int_0^1 1 dx$$

$$B_{11} = [x]_0^1 = 1 - 0 = \mathbf{1}$$

Element B_{12} (i=1, j=2):

$$B_{12} = \int_0^1 \phi'_1(x) \phi'_2(x) dx = \int_0^1 (1) \cdot (2x) dx = \int_0^1 2x dx$$

$$B_{12} = [x^2]_0^1 = 1^2 - 0^2 = \mathbf{1}$$

Element B_{21} (i=2, j=1):

$$B_{21} = \int_0^1 \phi'_2(x) \phi'_1(x) dx = \int_0^1 (2x) \cdot (1) dx = \int_0^1 2x dx$$

$$B_{21} = [x^2]_0^1 = 1^2 - 0^2 = \mathbf{1}$$

(Note: $B_{12} = B_{21}$, as expected. The matrix is symmetric.)

Element B_{22} (i=2, j=2):

$$B_{22} = \int_0^1 \phi'_2(x) \phi'_2(x) dx = \int_0^1 (2x) \cdot (2x) dx = \int_0^1 4x^2 dx$$

$$B_{22} = [\frac{4}{3}x^3]_0^1 = \frac{4}{3}(1)^3 - 0 = \frac{\mathbf{4}}{\mathbf{3}}$$

Step 4: Assemble the Final Matrix Form

Plugging these values in, we get our “discretized smoothness matrix” \mathbf{B} :

$$\mathbf{B} = \begin{bmatrix} 1 & 1 \\ 1 & 4/3 \end{bmatrix}$$

The Dirichlet energy for *any* function of the form $f(x) = c_1x + c_2x^2$ on the domain $[0, 1]$ can now be calculated with this matrix:

$$S = \mathbf{c}^T \mathbf{B} \mathbf{c} = \begin{bmatrix} c_1 & c_2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 4/3 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$$

If we multiply this out, we get:

$$S = c_1(c_1 + c_2) + c_2(c_1 + \frac{4}{3}c_2)$$

$$S = c_1^2 + c_1c_2 + c_1c_2 + \frac{4}{3}c_2^2$$

$$S = c_1^2 + 2c_1c_2 + \frac{4}{3}c_2^2$$

Verification (Checking our work)

We can verify this by calculating the energy directly from the original definition, $S = \int_0^1 (f'(x))^2 dx$.

1. $f(x) = c_1x + c_2x^2$

2. $f'(x) = c_1 + 2c_2x$

3. $(f'(x))^2$

$$= (c_1 + 2c_2x)^2$$

$$= c_1^2 + 2(c_1)(2c_2x) + (2c_2x)^2$$

$$= c_1^2 + 4c_1c_2x + 4c_2^2x^2$$

4. Now, integrate:

$$S = \int_0^1 (c_1^2 + 4c_1c_2x + 4c_2^2x^2) dx$$

$$S = [c_1^2x + 4c_1c_2\frac{x^2}{2} + 4c_2^2\frac{x^3}{3}]_0^1$$

$$S = [c_1^2x + 2c_1c_2x^2 + \frac{4}{3}c_2^2x^3]_0^1$$

5. Evaluate at the bounds (the 0 bound gives 0):

$$S = (c_1^2(1) + 2c_1c_2(1)^2 + \frac{4}{3}c_2^2(1)^3) - (0)$$

$$S = c_1^2 + 2c_1c_2 + \frac{4}{3}c_2^2$$

This result is identical to the one from the matrix calculation. This demonstrates how the \mathbf{B} matrix, built using the i and j indices, correctly pre-computes all the “interaction terms” between the basis function gradients.