Partial differential equations and optimization: optimizing forcing terms

Martin Berggren
Department of Information Technology
Uppsala University

Optimization for differential equations

Inverse/control problems

An inverse/control problem for steady heat conduction

- $u$: temperature field in a homogeneous, isotropic solid $\Omega$ (e.g. a metal)
- Held at constant temperature at the boundary $\partial \Omega$
- $f$: Heat sources (e.g. electric wires) distributed in $\Omega$
- $u$ can be measured at each point in $\omega \subset \Omega$

Mathematical model

$$-\Delta u = f \quad \text{in } \Omega, \quad u = 0 \quad \text{on } \partial \Omega.$$ (1)
Inverse/control problems

- An **analysis** problem: given \( f \), compute \( u|_\omega \)
- An **inverse** or **control** problem: given a target temperature \( z \) within \( \omega \), find \( f \) so that \( u|_\omega = z \)
- Thus, properties of the mapping \( A : f \mapsto u|_\omega \) of interest
- Sometimes called the “forward map”: \( A : L^2(\Omega) \to L^2(\omega) \). (In fact, \( A \) is a bounded, linear map.)
- Is the system “controllable”: can we find a \( f \) so that “every” \( u|_\omega \) can be reached?

Cf.: \( u = Af \), \( A \) an \( n \)-by-\( m \) matrix. “Controllable” if for each \( u \in \mathbb{R}^n \) there is an \( f \in \mathbb{R}^m \) such that \( u = Af \)

- What is \( \mathcal{R}(A) = \{ u \mid u = Af \text{ for some } f \in \mathbb{R}^m \} \)?
- \( \mathcal{R}(A) \subseteq \mathbb{R}^n \); controllable if \( \mathcal{R}(A) = \mathbb{R}^n \)

**Method:** check which vectors that are orthogonal to reachable \( u \)'s. If from \( 0 = v^T u = v^T Af \ \forall f \in \mathbb{R}^m \) follows that \( v = 0 \), then \( \mathcal{R}(A) = \mathbb{R}^n \).

**Remark:** We are then really computing the **null space** of \( A^T \): \( \mathbb{R}^n \) can be split in the orthogonal spaces \( \mathcal{R}(A) \) and \( \mathcal{N}(A^T) (\mathbb{R}^n = \mathcal{R}(A) \oplus \mathcal{N}(A^T)) \).
Inverse/control problems

Back to problem (1)

Theorem
If \( v \in L^2(\omega) \) satisfies, for each \( f \in L^2(\Omega) \),

\[
\int_\omega v u = 0 \tag{2}
\]

where \( u \) is corresponding (weak) solution to equation (1), then \( v \equiv 0 \)

Proof. Let \( p \) be the (weak) solution to

\[
-\Delta p = \chi_\omega v \quad \text{in } \Omega
\]

\[
p = 0 \quad \text{on } \partial\Omega \tag{3}
\]

where

\[
\chi_\omega = \begin{cases} 
1 & \text{in } \omega \\
0 & \text{in } \Omega \setminus \omega
\end{cases}
\]

Multiply each side of equation (3) with \( u \) and integrate by parts:

\[
0 = \int_\omega u v \quad \text{[by (2)]}
\]

\[
= -\int_\Omega u \Delta p = -\int_{\partial\Omega} \left( u \frac{\partial p}{\partial n} - \frac{\partial u}{\partial n} p \right) - \int_\Omega \Delta u \ p \quad \text{[by (1), (3)]}
\]

\[
= \int_\Omega f p \quad \forall f \in L^2(\Omega)
\]

implying that \( p \equiv 0 \), from which it follows that \( v \equiv 0 \) (multiply each side of equation (3) with \( \chi_\omega v \) and integrate)
Recall $\mathcal{A} : L^2(\Omega) \to L^2(\omega); u|_{\omega} = \mathcal{A} f$

The theorem says that only the zero function is orthogonal to $\mathcal{R}(\mathcal{A})$

As opposed the the matrix case, cannot say that $\mathcal{R}(\mathcal{A}) = L^2(\omega)$

We can only say that $\mathcal{R}(\mathcal{A})$ is dense in $L^2(\omega)$ ($\overline{\mathcal{R}(\mathcal{A})} = L^2(\omega)$)

That is, for each $z \in L^2(\omega)$ there is a sequence $\{ f_n \}_{n=1}^{\infty}$ such that

$$\lim_{n \to \infty} \mathcal{A} f_n = z$$

The generalization of matrix theorem

$$\mathbb{R}^n = \mathcal{R}(\mathcal{A}) \oplus \mathcal{N}(\mathcal{A}^T)$$

to bounded linear operators $\mathcal{A} : W \to V$ between Hilbert spaces $W$ and $V$ is

$$V = \overline{\mathcal{R}(\mathcal{A})} \oplus \mathcal{N}(\mathcal{A}^T)$$

Note that $p = \mathcal{A}^T v$, where $p$ is the solution to equation (3)
We summarize

**Theorem**

\[ u|_{\omega} \] spans a dense subspace of \( L^2(\omega) \) as \( f \) spans \( L^2(\Omega) \)

- This property is sometimes called “approximate controllability”
- Generally the best that can be achieved for elliptic (and parabolic) PDE’s
- Thus, for each given target temperature distribution \( z \in L^2(\omega) \), there is a \( f \in L^2(\Omega) \) so that \( \int_{\omega} (u - z)^2 \) can be made arbitrary small
- Is this good news or bad news?

---

**Almost controllable but ill-posed**

**Problem:** there are many \( z \in L^2(\omega) \) that are not in \( \mathcal{R}(A) \) (any discontinuous function e. g.)

Let \( z \notin \mathcal{R}(A) \), and let \( \{ f_n \}_{n=1}^{\infty} \subset L^2(\Omega) \),

\[
-\Delta u_n = f_n \quad \text{in } \Omega, \\
u_n = 0 \quad \text{on } \partial \Omega.
\]

such that

\[
\| Af_n - z \|^2 = \int_{\omega} (u_n - z)^2 \to 0
\]

Then, \( \int_{\Omega} f_n^2 \) will diverge. (Otherwise would \( z \in \mathcal{R}(A) \))

Also difficult if \( z \in \mathcal{R}(A) \). Any noise would take \( z \) outside of \( \mathcal{R}(A) \) and cause the above problems.
Inverse/control problems

Regularization

The easiest strategy to overcome the above problems: make sure $\|f\|$ does not blow up.

Introduce the Tikhonov regularization parameter $\epsilon > 0$ and the objective function, defined for any target function $z \in L^2(\Omega)$

$$J(f) = \frac{\epsilon}{2} \int_{\Omega} f^2 + \frac{1}{2} \int_{\omega} (u - z)^2$$

and solve the problem

Find $f^* \in L^2(\Omega)$ such that

$$J(f^*) \leq J(f) \quad \forall f \in L^2(\Omega),$$

where $u$ is obtained from $f$ by solving equation (1).

Sensitivity analysis

We will mimic the previous linear-algebra reasoning

Differentiating objective function (4) with respect to an arbitrary variation $\delta f$ yields

$$\delta J = \epsilon \int_{\Omega} \delta f \ f + \int_{\omega} \delta u (u - z)$$

where $\delta u$ is the directional derivative of $u$ with respect to $\delta f$. Differentiating state equation (1) gives an equation for $\delta u$:

$$-\Delta \delta u = \delta f \quad \text{in } \Omega,$$  \(7a\)

$$\delta u = 0 \quad \text{on } \partial \Omega.$$  \(7b\)

Multiplying (7a) with an arbitrary smooth function $p$ and integrating by parts yields
\[
\int_{\Omega} p \delta f = - \int_{\Omega} p \Delta \delta u = - \int_{\partial \Omega} p \frac{\partial \delta u}{\partial n} + \int_{\partial \Omega} \frac{\partial p}{\partial n} \delta u - \int_{\Omega} \Delta p \delta u
\]  
[by (7b)] = - \int_{\partial \Omega} p \frac{\partial \delta u}{\partial n} - \int_{\Omega} \Delta p \delta u
\]

(8)

So far, \( p \) is an arbitrary smooth function. Now choose \( p \) to be the solution to the adjoint equation

\[-\Delta p = \chi_{\omega} (u - z) \quad \text{in} \ \Omega \]
\[p = 0 \quad \text{on} \ \partial \Omega\]  

(9)

so that (8) reduces to

\[
\int_{\Omega} p \delta f = \int_{\omega} (u - z) \delta u
\]  

(10)

Substituting (10) into (6) yields

\[
\delta J = \int_{\Omega} \delta f (\epsilon f + p)
\]  

(11)

Expression (11) reveals that the (Frechét) derivative of \( J \) is the function

\[
DJ = \epsilon f + p
\]

Problem (5) is an unconstrained minimization problem (in fact, a linear least-squares problem), and the first-order optimality conditions is thus

\[
DJ = \epsilon f + p = 0
\]
Optimality system

The state equation, the adjoint equation, and the first-order optimality conditions are together known as the **optimality system**

\[-\Delta u = f \text{ in } \Omega, \quad -\Delta p = \chi_\omega(u - z) \text{ in } \Omega\]

\[u = 0 \text{ on } \partial\Omega, \quad p = 0 \text{ on } \partial\Omega.\]

(12)

\[\epsilon f + p = 0 \text{ in } \Omega\]

- The above derivation used a nested approach: we computed the derivative of \(\phi \mapsto J\)
- An alternative non-nested approach: view state equation as constraints and compute the stationary points of a Lagrangian

Finite-element discretization

FE discretizations are applied on variational forms of the PDE. Multiply equation (1) with an arbitrary smooth function \(v\) that vanishes on \(\partial\Omega\) and integrate by parts:

\[-\int_\Omega v\Delta u = -\int_{\partial\Omega} v\frac{\partial u}{\partial n} + \int_\Omega \nabla v \cdot \nabla u = \int_\Omega \nabla v \cdot \nabla u = \int_\Omega vf\]

Thus, if \(u\) is a solution to equation (1) then \(u\) satisfies the variational form

\[\int_\Omega \nabla v \cdot \nabla u = \int_\Omega vf\]

for each smooth function vanishing on \(\partial\Omega\)
Now “forget” about the PDE (1), define the **energy space**

\[ V = \left\{ v \mid \int_{\Omega} |\nabla u|^2 < +\infty, u = 0 \text{ on } \partial\Omega \right\} \]

(admissible temperature fields are only those with bounded heat energy), and consider the problem

Find \( u \in V \) such that

\[ \int_{\Omega} \nabla v \cdot \nabla u = \int_{\Omega} vf \quad \forall v \in V, \tag{13} \]

which has a unique solution by the Riesz representation theorem. The solution to problem (13) is called a **weak solution** to PDE (1).

We may approximate problem (13) on any finite-dimensional subspace \( V_h \subset V \)

Find \( u_h \in V_h \) such that

\[ \int_{\Omega} \nabla v_h \cdot \nabla u_h = \int_{\Omega} v_h f_h \quad \forall v_h \in V, \tag{14} \]

In FEM, we **triangulate** the domain, and let \( v_h|_T \) be a polynomial for each triangle \( T \).

**Theorem**

*Let \( v_h|_T \) be a polynomial for each triangle \( T \). Then \( \int_{\Omega} |\nabla v_h|^2 < +\infty \) if and only if \( v_h \) is continuous on \( \bar{\Omega} \).*

Thus, for piecewise polynomials to be a subset of \( V \), they need to be continuous in \( \bar{\Omega} \).
Choose $V_h$ to be a space of continuous functions that are some polynomial on each triangle and that vanish at $\partial \Omega$

Each $u_h \in V_h$, for $V_h$ being a finite-element space, can be expanded in terms of basis functions $N_j$:

$$u_h(x) = \sum_{j=1}^{N} u_j N_j(x) \quad (15)$$

Simplest case: $V_h$ is the space of continuous function, linear on each triangle. Then expansion (15) interpolates the values of $u_h$ at the mesh vertices $x_j$: $u_h(x_j) = u_j$

Inserting expansion (15) into (14), expanding $f_h(x) = \sum_{j=1}^{n} f_j N_j(x)$, and choosing $v_h = N_i, i = 1, \ldots, N$ yields

$$\sum_{j=1}^{N} u_j \int_{\Omega} \nabla N_i \cdot \nabla N_j = \sum_{j=1}^{N} f_j \int_{\Omega} N_i N_j,$$

or

$$Ku = Mf,$$

where

$$K_{ij} = \int_{\Omega} \nabla N_i \cdot \nabla N_j, \quad M_{ij} = \int_{\Omega} N_i N_j$$

$$u = (u_1, \ldots, u_N)^T \quad f = (f_1, \ldots, f_N)^T$$
Discrete inverse/control problem

Discrete objective function:

\[ J_h(f_h) = \frac{\epsilon}{2} \int_{\Omega} f_h^2 + \frac{1}{2} \int_{\omega} (u_h - z)^2 \]  

(16)

where \( u_h \in V_h \) such that

\[ \int_{\Omega} \nabla v_h \cdot \nabla u_h = \int_{\Omega} v_h f_h \quad \forall v_h \in V_h, \]  

(17)

The discrete inverse/control problem

Find \( f_h^* \in V_h \) such that

\[ J_h(f_h^*) \leq J_h(f_h) \quad \forall f_h \in V_h, \]

Sensitivity analysis in the discrete case

Two approaches:

1. Discretize the state equation (1), the adjoint equation (9), and use the discrete quantities in the gradient expression (11). ("differentiate-then-discretize"; "continuous adjoint")

2. Discretize the state equation and the objective function, derive corresponding discrete adjoint equation and gradient expression ("discretize-then-differentiate"; "discrete adjoint")

- In current case: these approaches are equivalent if the same approximations are used throughout
- In general, these approaches may give different results!
Recommendation: if at all possible, use strategy 2.

- Yields the exact gradient of the discrete problem, the one actually solved in the computer.
- Most optimization algorithms very sensitive to gradient accuracy
- Sometimes strategy 2 too complicated. Often easier algebra with strategy 1.

Discrete sensitivity analysis (strategy 2)

Differentiate objective function (16) and state equation (17):

$$\delta J_h = \epsilon \int_{\Omega} \delta f_h f_h + \int_{\omega} \delta u_h (u_h - z)$$  \hspace{1cm} (18)

where

$$\int_{\Omega} \nabla v_h \cdot \nabla \delta u_h = \int_{\Omega} v_h \delta f_h \quad \forall v_h \in V_h$$  \hspace{1cm} (19)

Now let $p_h \in V_h$ be the solution to the discrete adjoint equation

$$\int_{\Omega} \nabla p_h \cdot \nabla w_h = \int_{\omega} w_h (u_h - z) \quad \forall w_h \in V_h$$  \hspace{1cm} (20)

Choosing $v_h = p_h$ in (19) and using (20), we find that

$$\int_{\omega} \delta u_h (u_h - z) = \int_{\Omega} p_h \delta f_h$$  \hspace{1cm} (21)
Substituting (21) into (18) yields

$$\delta J_h = \int_\Omega \delta f_h (\epsilon f_h + p_h)$$

So, the derivative will again be $D J_h = \epsilon f_h + p_h$

---

Solution algorithms

Objective function (16) can be written

$$J_h(f_h) = \frac{\epsilon}{2} \| f_h \|_{L^2(\Omega)}^2 + \frac{1}{2} \| A_h f_h - z \|_{L^2(\omega)}^2,$$

where $A_h f_h = u_h$. (The matrix representation of $A_h$ is $K^{-1} M$). Also

$$\nabla J_h = \epsilon f_h + A_h^T (A_h f_h - z)$$

(22)
Solution algorithms

Cf. the algebraic least-squares problem

\[
\min_x \frac{1}{2} \| Ax - b \|^2, \tag{23}
\]

where \( x \in \mathbb{R}^n \), \( b \in \mathbb{R}^m \), and \( A \) is \( m \)-by-\( n \); typically \( m \gg n \)

**Theorem**

*Problem (23) has always at least one solution for any matrix \( A \) and vector \( b \), and the solution is unique if the columns of \( A \) are linearly independent.*

▶ A least-squares solution makes \( \| Ax - b \| \) small but \( x \) can be ugly; \( \| x \| \) can become very large e. g.

▶ Tichonov regularization can also here be applied:

\[
\min_x \frac{\epsilon}{2} \| x \|^2 + \frac{1}{2} \| Ax - b \|^2, \tag{24}
\]

▶ Problem (24) has a unique solution for each \( A \) and \( b \). Can be written in previous form

\[
\min_x \frac{1}{2} \| \tilde{A} x - \tilde{b} \|, \tag{25}
\]

with

\[
\tilde{A} = \begin{pmatrix} \epsilon^{\frac{1}{2}} I \\ A \end{pmatrix}, \quad \tilde{b} = \begin{pmatrix} 0 \\ b \end{pmatrix}
\]
A solution to (24) satisfies the \textit{normal equations}

\[
(\epsilon I + A^T A)x = A^T b
\]

(The normal equations for (23) or (25), set \(\epsilon = 0\).)

- The recommended solution method, at least for the full-rank case (linearly independent columns in \(A\)) relies on a \textit{QR-factorization} of \(A\) (or \(\tilde{A}\) in the regularized case). Memory demanding for large \(A\).
- Alternatively, since \(A^T A\) is symmetric and positive definite, we can use the \textit{conjugate gradient} algorithm to solve the normal equation. Then, matrix \(A\) need not to be available explicitly, only the action of \(A\) and \(A^T\) on vectors.

Canned software for the conjugate-gradient algorithm formulated as algorithm for solving \(Sx = r\) with \(S\) symmetric positive definite. The algorithm needs

- the right-hand-side vector \(r\) (at the beginning)
- the vector \(Sw\) for a given vector \(w\) (at each iteration)
- For the inverse problem, \(r\) is the degrees of freedom for \(A_h^T z\)
- For the inverse problem, \(Sw\) is the degrees of freedom for \(\epsilon w_h + A_h^T A_h w_h\)
Right-hand side: \( q_h = \mathcal{A}_h^T z \)
\( q_h \in V_h \) such that
\[
\int_{\Omega} \nabla q_h \cdot \nabla v_h = \int_{\omega} v_h z \quad \forall v_h \in V_h
\]

Matrix–vector product: \( g_h = \epsilon w_h + \mathcal{A}_h^T \mathcal{A}_h w_h \)
Given \( w_h \in V_h \), first compute \( u_h \in V_h \) such that
\[
\int_{\Omega} \nabla v_h \cdot \nabla u_h = \int_{\Omega} v_h w_h \quad \forall v_h \in V_h,
\]
than compute \( p_h \in V_h \) such that
\[
\int_{\Omega} \nabla p_h \cdot \nabla v_h = \int_{\omega} v_h u_h \quad \forall v_h \in V_h
\]
and set \( g_h = \epsilon w_h + p_h \)