Eric de Sturler
Integrated Model Reduction Strategies for Nonlinear Parametric Inversion

Department of Mathematics
Virginia Tech
Blacksburg
VA 24061
sturler@vt.edu
Chris Beattie
Serkan Gugercin
Misha E. Kilmer

We will show how reduced order models can significantly reduce the cost of general inverse problems approached through parametric level set methods. Our method drastically reduces the solution of forward problems in diffuse optimal tomography (DOT) by using of interpolatory parametric model reduction. In the DOT setting, these surrogate models can approximate both the cost functional and associated Jacobian with little loss of accuracy and significantly reduced cost.

We recover diffusion and absorption coefficients, $D(\mathbf{x})$ and $\mu(\mathbf{x})$, respectively, using observations, $\boldsymbol{m}(t)$, from illumination by source signals, $\mathbf{u}(t)$. We assume that the unknown fields can be characterized by a finite set of parameters, $\mathbf{p} = [p_1, \ldots, p_\ell]^T$. Discretization of the underlying PDE gives the following, order n, differential algebraic system,

$$\frac{1}{\nu}\mathbf{E}\,\dot{\mathbf{y}}(t;\mathbf{p}) = -\mathbf{A}(\mathbf{p})\mathbf{y}(t;\mathbf{p}) + \mathbf{B}\mathbf{u}(t) \quad \text{with} \quad \boldsymbol{m}(t;\mathbf{p}) = \mathbf{C}\mathbf{y}(t;\mathbf{p}), \qquad (1)$$

where $\mathbf{y}$ denotes the discretized photon flux; $\boldsymbol{m} = [m_1, \ldots, m_{n_{det}}]^T$ is the vector of detector outputs; the columns of $\mathbf{B}$ are discretizations of the source "footprints"; and $\mathbf{A}(\mathbf{p})$ is the discretization of the diffusion and absorption terms, inheriting the parameterizations of these fields.

Let $\mathbf{y}(\omega;\mathbf{p})$, $\mathbf{u}(\omega)$, and $\mathbf{M}(\omega;\mathbf{p})$ denote Fourier transforms of $\mathbf{y}(t;\mathbf{p})$, $\mathbf{u}(t)$, and $\boldsymbol{m}(t;\mathbf{p})$, respectively. Taking the Fourier transform of (1), we find

$$\mathbf{M}(\omega;\mathbf{p}) = \boldsymbol{\Psi}(\imath\omega;\mathbf{p})\,\mathbf{u}(\omega) \quad \text{where} \quad \boldsymbol{\Psi}(s;\mathbf{p}) = \mathbf{C}\left(\frac{s}{\nu}\mathbf{E} + \mathbf{A}(\mathbf{p})\right)^{-1}\mathbf{B}. \qquad (2)$$

$\boldsymbol{\Psi}(s,\mathbf{p})$ is a mapping from sources (inputs) to measurements (outputs) in the frequency domain; it is the *transfer function* of the dynamical system (1). Given a parameter vector, $\mathbf{p}$, and absorption field, $\mu(\cdot,\mathbf{p})$, input source, $\mathbf{U}_i$, and frequency ω_j, $\mathbf{M}_i(\omega_j;\mathbf{p}) \in \mathbb{C}^{n_{det}}$ denotes the vector of observations predicted by

the forward model. For n_{src} sources and n_ω frequencies, we get

$$\mathcal{M}(\mathbf{p}) = [\mathbf{M}_1(\omega_1; \mathbf{p})^T, \mathbf{M}_1(\omega_2; \mathbf{p})^T, \ldots, \mathbf{M}_1(\omega_{n_\omega}; \mathbf{p})^T, \mathbf{M}_2(\omega_1; \mathbf{p})^T \ldots, \mathbf{M}_{n_{src}}(\omega_{n_\omega}; \mathbf{p})^T]^T,$$

which is a vector of dimension $n_{det} \cdot n_{src} \cdot n_\omega$. We obtain the empirical data vector, $\mathbb{D}$, from actual observations, and solve the optimization problem:

$$\min_{\mathbf{p} \in \mathbb{R}^\ell} \|\mathcal{M}(\mathbf{p}) - \mathbb{D}\|_2$$

The computational cost of evaluating $\mathcal{M}(\mathbf{p}) - \mathbb{D}$ is dominated by the solution of the large, sparse block linear systems in (2) for all frequencies ω_j.

To reduce costs while maintaining accuracy, we seek a much smaller dynamical system of order $r \ll n$ that replicates the input-output map of (1):

$$\frac{1}{\nu} \widehat{\mathbf{E}} \dot{\widehat{\mathbf{Y}}}_r(t; \mathbf{p}) = -\widehat{\mathbf{A}}(\mathbf{p}) \widehat{\mathbf{Y}}(t; \mathbf{p}) + \widehat{\mathbf{B}} \mathbf{U}(t) \quad \text{with} \quad \widehat{\boldsymbol{m}}(t; \mathbf{p}) = \widehat{\mathbf{C}} \widehat{\mathbf{Y}}(t; \mathbf{p}), \quad (3)$$

where the new state vector $\widehat{\mathbf{Y}}(t; \mathbf{p}) \in \mathbb{R}^r$, $\widehat{\mathbf{E}}, \widehat{\mathbf{A}}(\mathbf{p}) \in \mathbb{R}^{r \times r}$, $\widehat{\mathbf{B}} \in \mathbb{R}^{r \times n_{src}}$, and $\widehat{\mathbf{C}} \in \mathbb{R}^{n_{det} \times r}$ such that $\boldsymbol{m}(t; \mathbf{p}) \approx \widehat{\boldsymbol{m}}(t; \mathbf{p})$. The surrogate transfer function is

$$\widehat{\boldsymbol{\Psi}}(s; \mathbf{p}) = \widehat{\mathbf{C}} \left(\frac{s}{\nu}\widehat{\mathbf{E}} + \widehat{\mathbf{A}}(\mathbf{p})\right)^{-1} \widehat{\mathbf{B}},$$

which requires only the solution of linear systems of dimension $r \ll n$; hence drastically reducing the cost. For a given parameter value $\boldsymbol{\pi}$ used in an optimization step, a reduced (surrogate) model for the necessary function evaluations at the frequency ω_j, involves the construction of a reduced parametric model of the form (3) with transfer function $\widehat{\boldsymbol{\Psi}}(s; \mathbf{p})$ that satisfies

$$\widehat{\boldsymbol{\Psi}}(\imath\omega_j; \boldsymbol{\pi}) = \boldsymbol{\Psi}(\imath\omega_j; \boldsymbol{\pi}) \qquad (4)$$

This is exactly what interpolatory model reduction achieves. Moreover, the transfer function $\widehat{\boldsymbol{\Psi}}(s)$ of the reduced-model also satisfies

$$\nabla_{\mathbf{p}} \boldsymbol{\Psi}(\imath\omega_j, \boldsymbol{\pi}) = \nabla_{\mathbf{p}} \widehat{\boldsymbol{\Psi}}(\imath\omega_j, \boldsymbol{\pi}), \quad and \quad \boldsymbol{\Psi}'(\imath\omega_j, \boldsymbol{\pi}) = \widehat{\boldsymbol{\Psi}}'(\imath\omega_j, \boldsymbol{\pi}).$$

The use of interpolatory projections allows us to match both function and gradient values exactly *without computing them*, requiring instead only that the computed projection spaces defining the reduced model contain particular (stably computable) vectors. If we were to compute a reduced-model for every parameter point $\boldsymbol{\pi}$ and $\imath\omega_j$ for $j = 1, \ldots, N_\omega$ and use these reduced-models in the forward model, the solution of the inverse problem would proceed in *exactly* the same way as in the case of the full-order forward model – the non-linear optimization algorithm would not see the difference between the full and reduced forward problems. Of course, computing a new surrogate-model for every parameter value is infeasible. Hence, we focus on methods of constructing surrogate-models that have high-fidelity over a wide range of parameter values and consider effective approaches for updating the surrogate models. We will present numerical examples that illustrate the effectiveness of these methods.