# LECTURE 29: GRADIENT DESCENT METHODS

Alternatives to exploiting problem structure by operator splitting are suggested by the least action principle. The action

$$S(q,\dot{q}) = \int_{t_0}^{t_1} L(t,q(t),\dot{q}(t))\,\mathrm{d}t,$$

is a functional over the phase space of the system, e.g., $S\colon \mathbb{R}^{2n} \to \mathbb{R}$ for a system composed of $n$ point masses. The least action principle states that the observed trajectory minimizes the action, hence it is to be expected that optimization algorithms that solve the problem

$$\min_{q,\dot{q}} S$$

would be of interest. This indeed is the case and leads to a class of methods that can exhibit remarkably fast convergence and more readily generalize to variable physical properties and arbitrary domain geometry.

## 1. Spatially dependent diffusivity

A first step in considering linear operators that still exhibit structure but are more complex than the constant-coefficient discretization of the Laplacian $\nabla^2$ is to consider spatially-varying diffusivity in which case the steady-state heat equation in domain $\Omega$ becomes

$$-\nabla\cdot(\alpha\,\nabla u) = f, \tag{1}$$

again with Dirichlet boundary conditions $u = b$ on $\partial\Omega$. Maintaining simple domain geometry for now, the centered finite-difference discretization of (1) on $\Omega = [0,1]\times[0,1]$ with grid points $(x_i = ih, y_j = jh, h = 1/(n+1))$ becomes

$$-\alpha_{i+1/2,j}u_{i+1,j} - \alpha_{i-1/2,j}u_{i-1,j} - \alpha_{i,j+1/2}u_{i,j+1} - \alpha_{i,j-1/2}u_{i,j-1} + 4\bar{\alpha}_{i,j}u_{i,j} = c_{i,j}, \tag{2}$$

where $\bar{\alpha}_{i,j} = (\alpha_{i+1/2,j} + \alpha_{i-1/2,j} + \alpha_{i,j+1/2} + \alpha_{i,j-1/2})/4$ denotes a diffusivity average at $(i,j)$ and $\boldsymbol{c}$ contains the boundary conditions and forcing term as before, $\boldsymbol{c} = \boldsymbol{b} + h^2\boldsymbol{f}$. The sparsity pattern is the same as in the constant diffusivity case, but the system $\boldsymbol{A}\boldsymbol{u} = \boldsymbol{c}$ has a system matrix with variable coefficients. The matrix $\boldsymbol{A}$ expresses a self-adjoint operator through a symmetric discretization, namely centered finite differences on a uniform grid. It can be expected to be symmetric $\boldsymbol{A} = [a_{k,r}] = \boldsymbol{A}^T = [a_{r,k}]$, as verified by considering row $k = (j-1)n + i$, that has non-zero components in columns $k$, $k \pm 1$, $k \pm n$. It is sufficient to verify symmetry for entries within the lower triangle of $\boldsymbol{A}$. The $k, k-1$ component is the coefficient of $u_{i-1,j}$ in (2) $a_{k,k-1} = -\alpha_{i-1/2,j}$. Symmetry of $\boldsymbol{A}$ would require $a_{k,k-1} = a_{k-1,k}$. The $a_{k-1,k}$ component arises from row $k-1$

$$-\alpha_{i-1/2,j}u_{i,j} - \alpha_{i-3/2,j}u_{i-2,j} - \alpha_{i-1,j+1/2}u_{i-1,j+1} - \alpha_{i-1,j-1/2}u_{i-1,j-1} + 4\bar{\alpha}_{i-1,j}u_{i-1,j} = c_{i-1,j}.$$

The diagonal element for row $k-1$ has indices $(i-1,j)$ and the $k^{\text{th}}$ column has indices $(i,j)$ for which $a_{k-1,k} = -\alpha_{i-1/2,j}$, indeed verifying $a_{k,k-1} = a_{k-1,k}$. Such opaque index manipulations can readily be avoided by symmetry considerations as stated above: self-adjoint operator expressed through symmetric discretization. The physical argument is even simpler. Diffusivity expresses how readily heat is transferred between two spatial positions of unequal temperature, and there is no reason for this material property to differ in considering the heat flux from point $P$ to point $Q$, $q_{\text{PQ}} = \alpha_{\text{PQ}}(u_P - u_Q)$ from that from point $Q$ to $P$, $q_{\text{QP}} = \alpha_{\text{QP}}(u_Q - u_P)$. Setting $q_{\text{PQ}} = -q_{\text{QP}}$ to account for direction of heat flow leads to $\alpha_{\text{PQ}} = \alpha_{\text{QP}}$, and this material property is reflected in symmetry of $\boldsymbol{A}$. Note that even though the operator $\nabla\cdot(\alpha\,\nabla)$ might be self-adjoint under appropriate boundary conditions, unsymmetric discretization such as one-sided finite differences can lead to a non-symmetric system matrix $\boldsymbol{A}$.

The implications for iterative method convergence can again be surmised from the one-dimensional case with homogeneous boundary conditions $\partial_x(\alpha(x)\,\partial_x u) = f$, $u(0) = u(1) = 0$. The convergence rate for an iterative method depends on the eigenvalues of the matrix $\boldsymbol{A}$ obtained by discretization of the operator $\partial_x(\alpha(x)\,\partial_x)$. The regular Sturm-Liouvile eigenproblem $\partial_x(\alpha(x)\,\partial_x u) = \lambda u$, $u(0) = u(1) = 0$ is known to have a solution, albeit difficult to obtain analytically. Replacing analytical estimates by a numerical experiment taking $\alpha(x) = 1 + cx$, Fig. 1 shows that convergence becomes marginally slower as the diffusivity gradient $c$ increases, though the main difficulty is the $\rho(M) \lessapprox 1$ spectral radius for constant diffusivity.

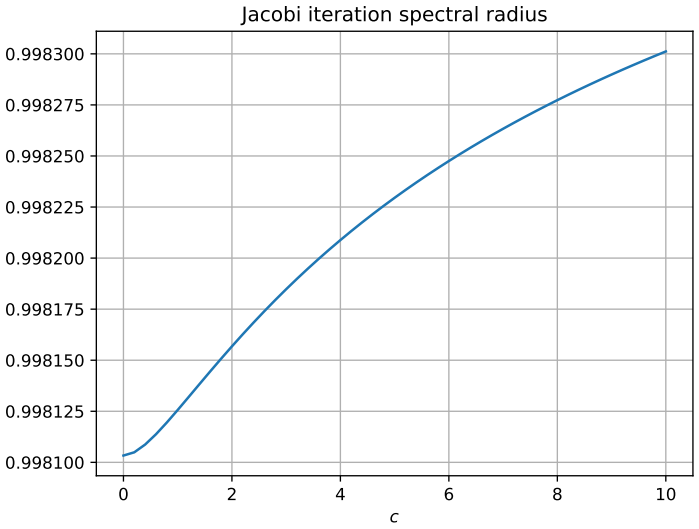


**Figure 1.** Spectral radius of Jacobi iteration $\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{D}^{-1}\boldsymbol{A}$ for $\partial_x(\alpha(x)\,\partial_x u) = f$ with increasing diffusivity gradient $\alpha = 1 + cx$.

## 2. Steepest descent

The heat equation can be obtained as the stationary solution $\delta\Phi = 0$, to an optimization problem for the functional

$$\Phi(u, u') = -\int_\Omega \left[\tfrac{1}{2}\alpha\,(\nabla u)\cdot(\nabla u) + uf\right]\mathrm{d}\boldsymbol{x}, \tag{3}$$

among all functions $u$ that satisfy the boundary condition $u = b$ on $\partial\Omega$. The above can be understood as the generalization of the one-dimensional case

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

The stationarity condition point for $\Phi$ is

$$\delta\Phi = -\int_0^1 (\alpha u'\,\delta u' + f\,\delta u)\,\mathrm{d}x = -\int_0^1 \left(\alpha u'\frac{\mathrm{d}}{\mathrm{d}x}\delta u + f\,\delta u\right)\mathrm{d}x = -[\alpha u'\,\delta u]_{x=0}^{x=1} + \int_0^1 \left(\frac{\mathrm{d}}{\mathrm{d}x}(\alpha u') - f\right)\delta u\,\mathrm{d}x = 0.$$

Since all $u$ must satisfy boundary conditions the perturbations are null at endpoints $\delta u(0) = \delta u(1) = 0$, and stationarity for arbitrary perturbations $\delta u$ implies that

$$\frac{\mathrm{d}}{\mathrm{d}x}(\alpha u') - f = 0,$$

the one-dimensional variable diffusivity heat equation.

How can the above observations guide algorithm construction? The key point is that the discrete problem should also be expressible as an optimization problem for $\Phi\colon \mathbb{R}^m \to \mathbb{R}$

$$\Phi(\boldsymbol{u}) = \frac{1}{2}\boldsymbol{u}^T\boldsymbol{A}\,\boldsymbol{u} - \boldsymbol{u}^T\boldsymbol{c} = \frac{1}{2}\sum_{j=1}^m\sum_{k=1}^m u_j a_{jk} u_k - \sum_{j=1}^m u_j c_j,$$

with $\boldsymbol{A} = [a_{jk}]$. The discrete stationarity condition is $\nabla_{\boldsymbol{u}}\Phi = 0$ leading to

$$\frac{\partial\Phi}{\partial u_l} = \frac{1}{2}\sum_{j=1}^m\sum_{k=1}^m (\delta_{lj} a_{jk} u_k + u_j a_{jk}\,\delta_{lk}) - \sum_{j=1}^m \delta_{lj} c_j.$$

Using the Kronecker delta properties $\delta_{ll}=1,\ \delta_{lj}=0$ for $l\neq j$ gives

$$\frac{\partial\Phi}{\partial u_l}=\frac{1}{2}\sum_{k=1}^{m}a_{lk}u_k+\frac{1}{2}\sum_{j=1}^{m}u_j a_{jl}-c_l,$$

which for symmetric $\boldsymbol{A}$ leads to

$$\frac{\partial\Phi}{\partial u_l}=\sum_{j=1}^{m}a_{lj}u_j-c_l=0\Rightarrow\boldsymbol{A}\boldsymbol{u}=\boldsymbol{c}. \tag{4}$$

Symmetric discretization of the self-adjoint operator $\nabla\cdot(\alpha\nabla u)$ produces a symmetric matrix that is unitarily diagonalizable $\boldsymbol{A}=\boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T$, and, as seen previously, with strictly positive eigenvalues. Hence stationary points of $\Phi(\boldsymbol{u})$ are minima and the solution to $\boldsymbol{A}\boldsymbol{u}=\boldsymbol{c}$ can be sought by minimizing $\Phi(\boldsymbol{u})$.

Equation (4) states that the gradient of $\Phi$ is opposite the direction of the residual $\nabla\Phi=\boldsymbol{A}\boldsymbol{u}-\boldsymbol{c}=-\boldsymbol{r}$. Since this is the direction of fastest increase of $\Phi$, travel in the opposite direction will decrease $\Phi$ leading to an update

$$\boldsymbol{u}_{k+1}=\boldsymbol{u}_k+\beta_k\boldsymbol{r}_k, \tag{5}$$

of the current approximation $\boldsymbol{u}_k$. The correction direction is also referred to as a search direction for the optimization procedure. In the residual correction formulation

$$\boldsymbol{r}_k=\boldsymbol{c}-\boldsymbol{A}\boldsymbol{u}_k,\boldsymbol{e}_k=\boldsymbol{B}\boldsymbol{r}_k,\boldsymbol{u}_{k+1}=\boldsymbol{u}_k+\boldsymbol{e}_k,$$

steepest descent corresponds to the choice $\boldsymbol{B}=\beta_k\boldsymbol{I}$. The remaining question is to determine how far to travel along the $-\nabla\Phi(\boldsymbol{u}_k)=\boldsymbol{r}_k$ search direction. As $\beta$ increases the local gradient direction changes. Steepest descent proceeds along the $\boldsymbol{r}_k$ direction until further decrease is no longer possible, that is when the new gradient direction is orthogonal to the previous one

$$\boldsymbol{r}_k^T\boldsymbol{r}_{k+1}=0\Rightarrow\boldsymbol{r}_k^T(\boldsymbol{c}-\boldsymbol{A}\boldsymbol{u}_{k+1})=\boldsymbol{r}_k^T[\boldsymbol{c}-\boldsymbol{A}\,(\boldsymbol{u}_k+\beta_k\boldsymbol{r}_k)]=\boldsymbol{r}_k^T(\boldsymbol{r}_k-\beta_k\boldsymbol{A}\boldsymbol{r}_k)=0\Rightarrow\beta_k=\frac{\boldsymbol{r}_k^T\boldsymbol{r}_k}{\boldsymbol{r}_k^T\boldsymbol{A}\boldsymbol{r}_k}.$$

The convergence rate is given by the spectral radius of $\boldsymbol{M}=\boldsymbol{I}-\boldsymbol{B}\boldsymbol{A}$ that becomes

$$\boldsymbol{M}=\boldsymbol{I}-\beta_k\boldsymbol{A}=\boldsymbol{I}-\frac{\boldsymbol{r}_k^T\boldsymbol{r}_k}{\boldsymbol{r}_k^T\boldsymbol{A}\boldsymbol{r}_k}\boldsymbol{A}.$$

Recall that the one-dimensional, constant diffusivity heat equation had eigenvalues of $\boldsymbol{A}$

$$\nu_l=4\sin^2\left[\frac{l\pi h}{2}\right],l=1,2,\ldots,m.$$

Eigenvalues of gradient descent iteration are therefore

$$\lambda_l=1-\beta_k\,\nu_l,l=1,2,\ldots,m.$$

Since $\beta_k$ is the inverse of a Rayleigh quotient, if the residual is in the direction of eigenvector $l$, $\beta_k=1/\nu_l$ and $\lambda_l=0$ suggesting the possibility of fast convergence. However, the distribution of eigenvalues $\nu_k$ for $\boldsymbol{A}$ is uniformly distributed in the interval [0,4] such that the residual component in other eigendirections is not significantly reduced. The typical behavior of gradient descent is rapid decrease of the residual in the first few iterations followed by slow convergence to the solution. Consider the problem

$$-u_{xx}=\pi^2\sum_{k=1}^{K}k^2\,\sin(k\pi x),u(0)=u(1)=0,$$

with solution

$$u(x)=\sum_{k=1}^{K}\sin(k\pi x).$$

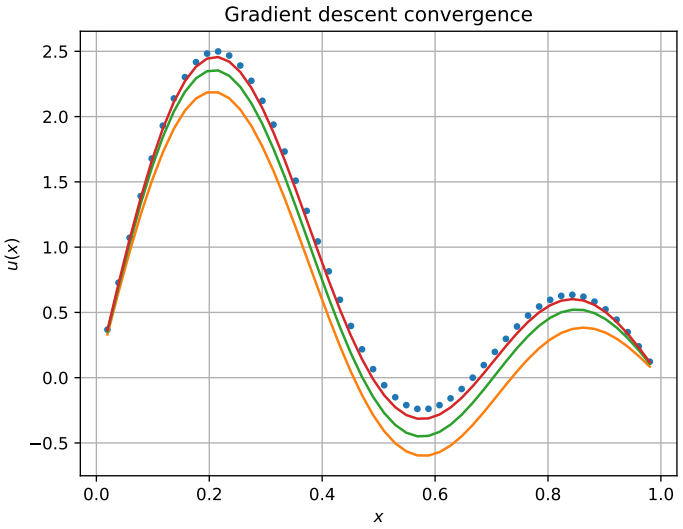


**Figure 2.** Convergence of gradient descent. Blue: exact solution. Orange, green, red: iterates after 4, 40, 400 iterations.

## 3. Conjugate gradient

Steepest descent is characterized by a correction in the direction of the residual (5). Enforcing $\boldsymbol{r}_k^T \boldsymbol{r}_{k+1} = 0$ leads to orthogonality of both succesive residuals and correction directions. A more insightful interpretation of (3) is to recognize the role of the scalar products

$$(f,g) = \tfrac{1}{2}\int_\Omega \alpha\,(\nabla f)\cdot(\nabla g)\mathrm{d}\boldsymbol{x},\ (\boldsymbol{u},\boldsymbol{v}) = \tfrac{1}{2}\boldsymbol{u}^T\boldsymbol{A}\boldsymbol{v},$$

in the continuum, discrete cases respectively. Similarly to how vectors that satisfy $\boldsymbol{u}^T\boldsymbol{v} = 0$ are said to be orthogonal, those that satisfy $\boldsymbol{u}^T\boldsymbol{A}\boldsymbol{v} = 0$ are said to be $\boldsymbol{A}$-conjugate. Gradient descent minimizes the 2-norm of the error $\|\boldsymbol{e}_k\|$ at each iteration. However, the variational formulation suggests that a more appropriate norm is the $\boldsymbol{A}$-norm

$$\|\boldsymbol{e}_k\|_{\boldsymbol{A}} = (\boldsymbol{e}_k^T\boldsymbol{A}\boldsymbol{e}_k)^{1/2}.$$

This leads to a modification of the search directions $\boldsymbol{p}_k$, which are no longer taken in the direction of the residual and orthogonal, but rather $\boldsymbol{A}$-conjugate

$$\boldsymbol{p}_{k+1}^T\boldsymbol{A}\boldsymbol{p}_k = 0.$$

**Algorithm Conjugate gradient**

$\boldsymbol{x}_0 = \boldsymbol{0}, \boldsymbol{r}_0 = \boldsymbol{c}, \boldsymbol{p}_0 = \boldsymbol{r}_0$
for $k = 1$: MaxIter
  $\beta_k = \boldsymbol{r}_{k-1}^T\boldsymbol{r}_{k-1} / (\boldsymbol{p}_{k-1}^T\boldsymbol{A}\boldsymbol{p}_{k-1})$
  $\boldsymbol{x}_k = \boldsymbol{x}_{k-1} + \beta_k\boldsymbol{p}_{k-1}$
  $\boldsymbol{r}_k = \boldsymbol{r}_{k-1} - \beta_k\boldsymbol{A}\boldsymbol{p}_{k-1}$
  $\gamma_k = \boldsymbol{r}_k^T\boldsymbol{r}_k / (\boldsymbol{r}_{k-1}^T\boldsymbol{r}_{k-1})$
  $\boldsymbol{p}_k = \boldsymbol{r}_k + \beta_k\boldsymbol{p}_{k-1}$
end