Math 273: **Homework #2**
Due on Wednesday, February 20

[**1**] Apply the gradient descent method described in class to the two-dimensional diffusion problem

$$F(u) = \sum_{1\le i,j\le n} \Big[(u_{i+1,j}-u_{i,j})^2 + (u_{i,j+1}-u_{i,j})^2 + \lambda(u_{i,j}-f_{i,j})^2\Big],$$

where $f_{i,j}$ is given for $0 \le i, j \le n+1$, and with boundary conditions $u_{i,j} = f_{i,j}$ if $i = 0$ or $i = n+1$ or $j = 0$ or $j = n+1$ (chosen for simplicity). Here $\lambda > 0$ is a tunning parameter. Choose a function f of your choice (for example an image). If you do not have one, you can create a synthetic image. Test various values of the parameter λ and observe the properties of your implementation. Give your choice of the stopping criterion and also plot the value of the objective function versus steps. Plot the data f, your starting point and your final result, as 2D images.

[**2**] Consider the constrained optimization problem

$$\min_{x\in\mathbb{R}^n} f(x), \quad \text{subject to } Ax = b,$$

where $A \in \mathbb{R}^{m\times n}$ has full row rank, $m \le n$, $b \in \mathbb{R}^m$. Tranform the problem into an unconstrained minimization problem. Hint: use a representation for the Null space of the matrix A.

[**3**] The problem of finding the shortest distance from a point x_0 to the hyperplane $\{x :\ Ax = b\}$ where A has full row rank can be formulated as the quadratic program

$$\min \frac{1}{2}(x-x_0)^T(x-x_0), \quad \text{s.t. } Ax = b.$$

(i) Show that the optimal multiplier is $\lambda^* = (AA^T)^{-1}(b - Ax_0)$, and that the solution is $x_* = x_0 + A^T(AA^T)^{-1}(b - Ax_0)$.

(ii) Show that in the special case where A is a row vector, the shortest distance from x_0 to the solution set of $Ax = b$ is $\frac{|b-Ax_0|}{\|A\|}$.

[**4**] Consider the quadratic program

$$\min_x q(x) = \frac{1}{2}x^TGx + x^Td, \text{ subject to } Ax = b,$$

where $G \in \mathbb{R}^{n\times n}$ is a symmetric matrix, $A \in \mathbb{R}^{m\times n}$. Assume that A has full row rank m.

(a) Express the first order necessary conditions for x_* to be a solution in the form of a linear matrix equation in the unknown $(x_* \ \ \lambda^*)^T$.

(b) Express in (a) x_* by $x + p$, with x some fixed feasible estimate and unknown $p \in Null(A)$. Re-write the matrix equation now in the unknown $(-p \ \ \lambda^*)^T$.

(c) Assume in addition that the reduced-Hessian Z^TGZ is positive definite. Show that the coefficient matrix in (b) is non-singular, thus there is a unique vector pair (x_*, λ^*) satisfying the matrix equation in (a).

Optional problems

• Repeat problem [2] from Homework #1, using now the BFGS formula that updates the inverse.

• Recall the BFGS update formula for the Hessian approximation:

$$B_{k+1} = B_k - \frac{B_k s_k s_k^t B_k}{s_k^t B_k s_k} + \frac{y_k y_k^t}{y_k^t s_k}$$

(where B_k is symmetric and positive definite), and the formula to directly update the inverse of Hessian approximation:

$$H_{k+1} = (I - \rho_k s_k y_k^t) H_k (I - \rho_k y_k s_k^t) + \rho_k s_k s_k^t$$

(where H_k is symmetric and positive definite, as inverse of B_k, and $\rho_k = \frac{1}{y_k^t s_k}$). Show that H_{k+1} is the inverse of B_{k+1}.

Notes

• Sherman-Morrison-Woodbury formula. If A is an $n \times n$ nonsingular matrix, and a, b vectors in $\mathbb{R}^n$, let $\overline{A} = A + ab^t$. Then the following (SMW) formula holds:

$$\text{(SMW)} \qquad \overline{A}^{-1} = A^{-1} - \frac{A^{-1} a b^t A^{-1}}{1 + b^t A^{-1} a}.$$

• If A is a symmetric (or self-adjoint) linear operator on X, then $\text{Im}A^{\perp} = \text{Ker}A$. Let $p_{\text{Im}A}$ be the operator of orthogonal projection onto $\text{Im}A$. For given $y \in X$, there is a unique $x = x(y)$ in $\text{Im}A$ such that $Ax = p_{\text{Im}A} y$. Forthermore, the mapping $y \mapsto x(y)$ is linear. This mapping is called the pseudo-inverse, or generalized inverse of A.

• Let Ω be an open and bounded subset of $\mathbb{R}^d$, with Lipschitz-continuous (or sufficiently smooth) boundary $\partial\Omega$. Let $\vec{n} = (n_1, n_2, ..., n_d)$ be the exterior unit normal to $\partial\Omega$. Recall the following fundamental Green's formula, or integration by parts formula: given two functions u, v (with u, v, and all their 1st order partial derivatives belonging to $L^2(\Omega)$, or $u, v \in H^1(\Omega)$), then

$$\int_{\Omega} u v_{x_i} dx = -\int_{\Omega} u_{x_i} v dx + \int_{\partial\Omega} u v n_i dS.$$