SQP method for equality constrained optimization

Jae Hwa LEE

AORC
Sungkyunkwan University

Jan 31 2018
The 1st AORC Center-wide Monthly Seminar

Equality Constrained Optimization (ECO)

$$
\begin{aligned}
&\text{minimize} \quad f(x) \\
&\text{subject to} \quad c_i(x) = 0, \quad i = 1, ..., m_e
\end{aligned}
$$

- $f(x)$, $c_i(x)$ are smooth functions.
- Lagrangian function

$$L(x, \lambda) = f(x) + \lambda^T c(x),$$

where $\lambda = (\lambda_1, \ldots, \lambda_{m_e})^T$, $c(x) = (c_1(x), \ldots, c_{m_e}(x))^T$

Notations

- $g(x) = \nabla f(x)$, $A(x) = \nabla c(x) = (\nabla c_1(x), \ldots, \nabla c_{m_e}(x))$
- $f_k = f(x_k)$, $c_k = c(x_k)$, $g_k = g(x_k) = \nabla f(x_k)$, and so on.
- Hessian of Lagrangian

$$\nabla_x^2 L(x, \lambda) = W(x, \lambda) = \nabla^2 f(x) + \sum_{i=1}^{m_e} \lambda_i \nabla^2 c_i(x)$$

- Exact penalty function

$$\phi(x, \sigma) = f(x) + \sigma ||c(x)||,$$

where $\sigma > 0$ is a penalty parameter.

Lagrange-Newton method

$$
\begin{aligned}
\nabla_x L(x, \lambda) = g(x) + A(x)\lambda = 0 \\
c(x) = 0
\end{aligned}
$$

$$\Downarrow \quad \text{Newton's method}$$

$$
\begin{bmatrix} W(x_k, \lambda_k) & A_k \\ A_k^T & 0 \end{bmatrix} \begin{bmatrix} d \\ \Delta\lambda \end{bmatrix} = - \begin{bmatrix} g_k + A_k \lambda_k \\ c_k \end{bmatrix}
$$

$$\Updownarrow$$

$$
\begin{bmatrix} W(x_k, \lambda_k) & A_k \\ A_k^T & 0 \end{bmatrix} \begin{bmatrix} d \\ \lambda^+ \end{bmatrix} = - \begin{bmatrix} g_k \\ c_k \end{bmatrix}
$$

Sequential Quadratic Programming (SQP) method

$$B_k \approx W(x_k, \lambda_k) = \nabla^2 f(x_k) + \sum_{i=1}^{m_e} \lambda_i^{(k)} \nabla^2 c_i(x_k)$$

$$\Downarrow \quad \text{SQP method}$$

$$\begin{bmatrix} B_k & A_k \\ A_k^T & 0 \end{bmatrix} \begin{bmatrix} d \\ \lambda^+ \end{bmatrix} = - \begin{bmatrix} g_k \\ c_k \end{bmatrix}$$

$$\Updownarrow$$

$$\begin{aligned} \text{minimize} \quad & g_k^T d + \frac{1}{2} d^T B_k d \\ \text{subject to} \quad & c_k + A_k^T d = 0 \end{aligned}$$

SQP Algorithm

- Step 1 Given x_0, B_0, $\epsilon > 0$, σ_0, $k := 0$.
- Step 2 If $||g_k + A_k\lambda_k|| + ||c_k|| \le \epsilon$, then stop.

  Solve the subproblem (1)

$$\begin{aligned} \text{minimize} \quad & g_k^T d + \frac{1}{2} d^T B_k d \\ \text{subject to} \quad & c_k + A_k^T d = 0 \end{aligned}$$

  to obtain d_k.
- Step 3 Carry out line search to obtain $\alpha_k > 0$, set

$$x_{k+1} = x_k + \alpha_k d_k.$$

- Step 4 Generate B_{k+1}, σ_{k+1}, $k =: k+1$, go to Step 2.

Quasi-Newton update

$$\nabla_x^2 L(x_{k+1}, \lambda_{k+1})(x_{k+1} - x_k) \approx \nabla_x L(x_{k+1}, \lambda_{k+1}) - \nabla_x L(x_k, \lambda_k)$$

- Approximate Hessian

$$B_{k+1} \approx \nabla_x^2 L(x_{k+1}, \lambda_{k+1})$$

- Quasi-Newton equation

$$B_{k+1} s_k = y_k,$$

  where $s_k = x_{k+1} - x_k$, $y_k = \nabla_x L(x_{k+1}, \lambda_{k+1}) - \nabla_x L(x_k, \lambda_k)$
- In general, symmetric and positive definite B_k is required in line search type method.

BFGS and L-BFGS update

- BFGS update

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

- Limited-memory BFGS update

$$B_k = \sigma_k I + [S_k \quad Y_k] D_k \begin{bmatrix} S_k^T \\ Y_k^T \end{bmatrix}$$

where $\sigma > 0$, D_k is a $2m \times 2m$ matrix and

$$[S_k \quad Y_k] = [s_{k-m}, \ldots, s_{k-1}, y_{k-m}, \ldots, y_{k-1}] \in \mathbb{R}^{n \times 2m},$$

Damped BFGS update

- In ECO case, $s_k^T y_k > 0$ may not hold.
- Damped BFGS updating for SQP

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{\bar{y}_k \bar{y}_k^T}{s_k^T \bar{y}_k}$$

where $\bar{y}_k = \theta_k y_k + (1 - \theta_k) B_k s_k$,

$$\theta_k = \begin{cases} 1, & \text{if} \quad s_k^T y_k \geq 0.2 s_k^T B_k s_k \\ (0.8 s_k^T B_k s_k)/(s_k^T B_k s_k - s_k^T y_k), & \text{otherwise.} \end{cases}$$

Damped L-BFGS update

- Damped limited-memory BFGS method (2)

$$B_k = \sigma_k I + [S_k \quad \bar{Y}_k] D_k \begin{bmatrix} S_k^T \\ \bar{Y}_k^T \end{bmatrix}$$

where D_k is a $2m \times 2m$ matrix and

$$[S_k \quad \bar{Y}_k] = [s_{k-m}, \ldots, s_{k-1}, \bar{y}_{k-m}, \ldots, \bar{y}_{k-1}] \in \mathbb{R}^{n \times 2m},$$

$$s_k = x_{k+1} - x_k, \quad \bar{y}_k = \theta_k y_k + (1 - \theta_k) B_k s_k$$

Global convergence

- This proof is using the idea of Powell's paper (Variable metric methods for constrained optimization 1983).

Theorem

Suppose that iterates are generated by the algorithm. Suppose that the sequences $\{x_k : k = 1, 2, 3, \ldots\}$ *and* $\{x_k + d_k : k = 1, 2, 3, \ldots\}$ *are contained in a closed, bounded, convex region of* $\mathcal{R}^n$ *in which* f *and* c_i *have continuous first derivatives. Suppose that the matrices* $\{B_k : k = 1, 2, 3, \ldots\}$ *and multipliers are bounded and that* σ *satisfies*

$$\sigma \geq \|\lambda_{k+1}\|_\infty + \bar{\sigma}, \quad \forall k,$$

where $\bar{\sigma} > 0$ *is a constant. Then all limit points of the sequence* $\{x_k : k = 1, 2, 3, \ldots\}$ *are KKT points of the original problem (1).*

Applications

- Y. Wang, S. Ma and Q. Ma. Full Space and Subspace Methods for Large Scale Image Restoration, in: Y. F. Wang, A. G. Yagola and C. C. Yang eds., Optimization and Regularization for Computational Inverse Problems and Applications, Beijing/Berlin: Higher Education Press and Springer, 2010.
- Y. Wang and S. Ma, A fast subspace method for image deblurring, Applied Mathematics and Computation, 215 (2009), pp. 2359-2377.
- Liu et al, Limited Memory Block Krylov Subspace Optimization for Computing Dominant Singular Value Decompositions, *SIAM Journal on Scientific Computing*, 35 (2013), pp.A1641A1668.

References

- R. H. Byrd, R. B. Schnabel and G. A. Shultz, Approximate solution of the trust region problem by minimization over two-dimensional subspaces, *Mathematical Programming*, 40 (1988), pp. 247-263.
- J. E. Dennis, Jr. and H. H. W. Mei, Two new unconstrained optimization algorithms which use function and gradient values, *Journal of Optimization Theory and Applications*, 28 (1979), pp. 453-482.
- G.N.Grapiglia et al, A subspace version of the Powell-Yuan trust region algorithm for equality constrained optimization, *Journal of the OR Society of China*, 1(2013), pp.425-451.
- D. C. Liu and J. Nocedal, On the limited memory BFGS method for large scale optimization, *Mathematical Programming*, 45 (1989), pp. 503-528.

- M. J. D. Powell, A new algorithm for unconstrained optimization, in: J. B. Rosen, O. L. Mangasarian and K. Ritter, eds., *Nonlinear Programming*, Academic Press, New York, 1970, pp. 31-65.
- G. A. Shultz, R. B. Schnabel and R. H. Byrd, A family of trust-region-based algorithms for unconstrained minimization with strong global convergence properties, *SIAM Journal on Numerical Analysis*, 22 (1985), pp. 47-67.
- T. Steihaug, The conjugate gradient method and trust regions in large scale optimization, *SIAM Journal on Numerical Analysis*, 20 (1983), pp. 626-637.
- J. Stoer and Y. Yuan, A subspace study on conjugate gradient algorithms, *ZAMM Z. angew. Math. Mech.*, 75 (1995), pp. 69-77.

- Z. -H. Wang, Z. -W. Wen and Y. Yuan, A subspace trust region method for large scale unconstrained optimization, in: Y. Yuan, ed., *Numerical Linear Algebra and Optimization*, Science Press, Beijing/New York, 2004, pp. 265-274.
- Z. -H. Wang and Y. Yuan, A subspace implementation of quasi-Newton trust region methods for unconstrained optimization, *Numerische Mathematik*, 104 (2006), pp. 241-269.
- Y. Yuan, On the truncated conjugate gradient method, *Mathematical Programming Series A*, 87 (2000), pp. 561-573.
- Y. Yuan, Subspace techniques for nonlinear optimization, in: R. Jeltsch, T. Li and I. H. Sloan, eds., *Some Topics in Industrial and Applied Mathematics (Series in Contemporary Applied Mathematics CAM 8)*, Higher Education Press, Beijing, 2007, pp. 206-218.

- Y. Yuan, Subspace methods for large scale nonlinear equations and nonlinear least squares, *Optimization and Engineering*, 10 (2009), pp. 207-218.