Review: Constrained Optimization and Surrogate Models

Robert Moss — mossr@cs.stanford.edu
Stanford University, Stanford, CA 94305 — *AA222/CS361*

General optimization problem:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{g}(\mathbf{x}) \leq \mathbf{0} \\ & \mathbf{h}(\mathbf{x}) = \mathbf{0} \end{aligned}$$

Equality form:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & \mathbf{c}^\top \mathbf{x} \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{b} \\ & \mathbf{x} \geq \mathbf{0} \end{aligned}$$

KKT conditions:

- **Feasibility**
- **Dual feasibility**
- **Complementary slackness**
- **Stationarity**

(10) ***Constraints.*** Equality constraints $h(\mathbf{x}) = 0$, inequality constraints $g(\mathbf{x}) \leq 0$, *transformations to remove constraints*, *Lagrange multipliers*,[1] *KKT conditions*, *duality*, *primal form*,[2] *dual form*,[3] *min-max inequality*, *duality gap*, *penalty methods*,[4] *augmented Lagrange method*,[5] and *interior point methods*.[6]

(11) ***Linear Constrained Optimization.*** *Linear program*, *general form*, *standard form*, *half-space*, *equality form*, *slack variables*, *simplex algorithm*,[7] and *dual certificates*.[8]

(12) ***Multiobjective Optimization.*** *Pareto optimality*, *dominance*, *criterion space* $\mathcal{Y}$, *Pareto frontier*, *weakly Pareto-optimal*, *utopia point*,[9] *Pareto front generation*, *constraint method*, *lexicographic method*, *weighted sum method*, *goal programming*,[10] *weighted exponential sum*, *weighted min-max method*,[11] *exponential weighted criterion*, *multiobjective population methods*, *subpopulations*, *vector evaluated genetic algorithm*, *nondomination ranking*, *Pareto filter*, *niche techniques*,[12] *preference elicitation*, *paired query selection*, *expert responses*, *prime analytic center*, *polyhedral method*, *design selection*, *decision quality improvement*, *minimax decision*, and *minimax regret*.

(13) ***Sampling Plans.*** *Space-filling*, *full factorial*, *random sampling*, *uniform projection plan*, *Latin-hypercube sampling*, *stratified sampling*, *space-filling metrics*, *discrepancy*, *pairwise distances*, *Morris-Mitchel criterion*, *space-filling subsets*, *greedy local search*, *exchange algorithm*, *quasi-random sequences*,[13] *Monte Carlo integration*, *quasi-Monte Carlo methods*, *additive recurrence*, *Halton sequence*, *Sobol sequence*, and *Sobol sequence*.

(14) ***Surrogate Models.*** *Regression*, *linear model*, *linear regression*, *design matrix* $\mathbf{X}$, *pseudoinverse*,[14] *basis functions*, *polynomial basis functions*, *sinusoidal basis functions*, *radial basis functions*, *regularization term*, *L_2 regularization*, *generalization error*, *training error*, *mean squared error*, *holdout method*, *random sampling*, *k-fold cross validation*, *leave-one-out cross-validation*, *complete cross-validation*, *bootstrap method*,[15] *leave-one-out bootstrap estimate*, and *0.632 bootstrap estimate*.

(15) ***Probabilistic Surrogate Models.*** *Gaussian process*, analytical *conditional* and *marginal distributions*, *mean function* $m(\mathbf{x})$, *covariance function* or *kernel* $k(\mathbf{x}, \mathbf{x}')$, *squared exponential kernel*, *characteristic length-scale* ℓ, *Maérn kernel*, *posterior distribution*, *predicted mean* $\hat{\mu}$,[16] *predicted variance* $\hat{\nu}$,[17] *standard deviation*, *gradient measurements*, *noisy measurements*, and *maximum likelihood estimate*.

(16) ***Surrogate Optimization.*** *Prediction-based exploration*, *error-based exploration*,[18] *lower confidence bound exploration*,[19] *probability of improvement exploration*,[20] *expected improvement exploration*,[21] and *safe optimization*.

(17) ***Optimization under Uncertainty.*** *Irreducible uncertainty* vs. *epistemic uncertainty*, *set-based optimization*, *minimax*, *information-gap decision theory*, *probabilistic uncertainty*, *expected value*, *variance*, *sensitive* vs. *robust* design points, *statistical feasibility*, *value at risk* (VaR), and *conditional value at risk* (CVaR).

[1] $\mathcal{L}(\mathbf{x}, \lambda) = f(\mathbf{x}) - \lambda h(\mathbf{x})$

[2] $\underset{\mathbf{x}}{\text{minimize}} \underset{\boldsymbol{\mu} \geq \mathbf{0}, \boldsymbol{\lambda}}{\text{maximize}} \mathcal{L}(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda})$

[3] $\underset{\boldsymbol{\mu} \geq \mathbf{0}, \boldsymbol{\lambda}}{\text{maximize}} \underset{\mathbf{x}}{\text{minimize}} \mathcal{L}(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda})$

[4] $p_{\text{count}}(\mathbf{x}), p_{\text{quadratic}}(\mathbf{x}), p_{\text{mixed}}(\mathbf{x})$

[5] $p_{\text{Lagrange}}(\mathbf{x})$

[6] $p_{\text{inv-barrier}}(\mathbf{x}), p_{\text{log-barrier}}(\mathbf{x})$

[7] *Greedy heuristic*, *Dantzig's rule*, *Bland's rule*.

[8] Verifies optimality.

[9] $y_i^{\text{utopia}} = \underset{\mathbf{x} \in \mathcal{X}}{\text{minimize}} f_i(\mathbf{x})$

[10] $\underset{\mathbf{x} \in \mathcal{X}}{\text{minimize}} \|\mathbf{f}(\mathbf{x}) - \mathbf{y}^{\text{goal}}\|_p$

[11] Also called *weighted Tchebycheff method*.

[12] *Fitness sharing*, *equivalence class sharing*.

[13] Also called *low-discrepancy sequences*.

[14] Moore-Penrose pseudoinverse $\mathbf{X}^+$

[15] Bootstrap method:

$$\begin{aligned} \epsilon_{\text{boot}} &= \frac{1}{b} \sum_{i=1}^{b} \epsilon_{\text{test}}^{(i)} \\ &= \frac{1}{m} \sum_{j=1}^{m} \frac{1}{b} \sum_{i=1}^{b} \left(y^{(j)} - \hat{f}^{(i)}(\mathbf{x}^{(j)}) \right)^2 \end{aligned}$$

[16] $\hat{\mu}(\mathbf{x}) = m(\mathbf{x}) + \boldsymbol{\theta}^\top \mathbf{K}(X, \mathbf{x})$

[17] $\hat{\nu}(\mathbf{x}) = \mathbf{K}(\mathbf{x}, \mathbf{x}) - \mathbf{K}(\mathbf{x}, X)\mathbf{K}(X, X)^{-1}\mathbf{K}(X, \mathbf{x})$

[18] $x^{(m+1)} = \underset{\mathbf{x} \in \mathcal{X}}{\arg\max}\, \hat{\sigma}(\mathbf{x})$

[19] $LB(\mathbf{x}) = \hat{\mu}(\mathbf{x}) - \alpha\hat{\sigma}(\mathbf{x})$

[20] Probability of improvement:

$$\begin{aligned} P(y < y_{\min}) &= \int_{-\infty}^{y_{\min}} \mathcal{N}(y \mid \hat{\mu}, \hat{\sigma}) dy \\ &= \Phi\left(\frac{y_{\min} - \hat{\mu}}{\hat{\sigma}} \right) \end{aligned}$$

[21] Expected improvement:

$$\begin{aligned} \mathbb{E}[I(y)] &= \hat{\sigma} \int_{-\infty}^{y'_{\min}} (y'_{\min} - z)\, \mathcal{N}(z \mid 0, 1) dz \\ &= (y_{\min} - \hat{\mu}) P(y \leq y_{\min}) \\ &\quad + \hat{\sigma}\, \mathcal{N}(y_{\min} \mid \hat{\mu}, \hat{\sigma}^2) \end{aligned}$$