

Classical Gram–Schmidt vs Modified Gram–Schmidt

Let $A \in \mathbb{R}^{m \times n}$, with $m \geq n$, and let A have n linearly independent columns a_1, a_2, \dots, a_n . There are many ways to implement the Gram–Schmidt process. Here are two *very* different implementations:

Classical	Modified
for k=1:n,	for k=1:n,
$w = a_k$	$w = a_k$
for j = 1:k-1,	for j=1:k-1,
$r_{jk} = q_j^T w$	$r_{jk} = q_j^T w$
end	$w = w - r_{jk} q_j$
for j = 1:k-1,	end
$w = w - r_{jk} q_j$	$r_{kk} = \ w\ _2$
end	$q_k = w / r_{kk}$
$r_{kk} = \ w\ _2$	end
$q_k = w / r_{kk}$	
end	

Please study the pseudocode above carefully. In exact arithmetic, these two methods generate exactly the same output (exercise: convince yourself of this). The algorithm genotypes are very similar (MGS is CGS with a few “pieces of DNA” removed). In exact arithmetic the phenotypes are identical, while in the presence of rounding errors the phenotypes are dramatically different.

In classical Gram–Schmidt (CGS) we compute the (signed) lengths of the orthogonal projections of $w = a_k$ onto q_1, q_2, \dots, q_{k-1} , and then subtract those projections (with the rounding errors) from w . If $Q_{k-1} = [q_1, q_2, \dots, q_{k-1}]$, then the orthogonal projector onto $\text{ColSp}(Q_{k-1})$ is $P = Q_{k-1}(Q_{k-1}^T Q_{k-1})^{-1} Q_{k-1}^T$. If Q_{k-1} has orthonormal columns, then $P = Q_{k-1} Q_{k-1}^T$:

$$w = (I - Q_{k-1} Q_{k-1}^T) a_k.$$

But because of rounding errors, Q_{k-1} does not have truly orthogonal columns. In modified Gram–Schmidt (MGS) we compute the length of the projection of $w = a_k$ onto q_1 and subtract that projection (and the rounding errors) from w . Next we compute the length of the projection of the *computed* w onto q_2 and subtract that projection (and the rounding errors) from w , and so on, but *always orthogonalizing against the **computed** version of w* . Evaluated from right to left:

$$w = (I - q_{k-1} q_{k-1}^T) \dots (I - q_2 q_2^T) (I - q_1 q_1^T) a_k.$$

If the computed $Q_{k-1}^T Q_{k-1} = I + E$, then this is very nearly the same w that would be computed by

$$w = (I - Q_{k-1} (Q_{k-1}^T Q_{k-1})^{-1} Q_{k-1}^T) a_k$$

where we replace $(Q_{k-1}^T Q_{k-1})^{-1}$ by $I - E$, and is much “more orthogonal” to Q_{k-1} than the CGS w .